Gregorio D'Agostino
Antonio Scala   *Editors*

Networks of Networks: The Last Frontier of Complexity

Springer

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems—cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The three major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, and the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations, and the "SpringerBriefs in Complexity" which are concise and topical working reports, case-studies, surveys, essays and lecture notes of relevance to the field. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

Understanding Complex Systems

Founding Editor: Scott Kelso

Future scientific and technological developments in many fields will necessarily depend upon coming to grips with complex systems. Such systems are complex in both their composition—typically many different kinds of components interacting simultaneously and nonlinearly with each other and their environments on multiple levels—and in the rich diversity of behavior of which they are capable.

The Springer Series in Understanding Complex Systems series (UCS) promotes new strategies and paradigms for understanding and realizing applications of complex systems research in a wide variety of fields and endeavors. UCS is explicitly transdisciplinary. It has three main goals: First, to elaborate the concepts, methods and tools of complex systems at all levels of description and in all scientific fields, especially newly emerging areas within the life, social, behavioral, economic, neuro- and cognitive sciences (and derivatives thereof); second, to encourage novel applications of these ideas in various fields of engineering and computation such as robotics, nano-technology and informatics; third, to provide a single forum within which commonalities and differences in the workings of complex systems may be discerned, hence leading to deeper insight and understanding.

UCS will publish monographs, lecture notes and selected edited contributions aimed at communicating new findings to a large multidisciplinary audience.

Gregorio D'Agostino · Antonio Scala
Editors

Networks of Networks: The Last Frontier of Complexity

Editors
Gregorio D'Agostino
ENEA - Italian National Agency for New
  Technology, Energy and Sustainable
  Economic Development
Rome
Italy

Antonio Scala
CNR-ISC Uos "La Sapienza"
Rome
Italy

Sometimes life is complicate, sometimes it is just complex

Preface

Modern life in fully developed countries relies on the coordinated functioning of several infrastructures such as Electric System, Aqueducts, Communication Assets, Fresh food distribution chains, Gas-ducts, Oil Pipelines, Transports, Financial networks, etc. Several of such infrastructures have been regarded as critical since they provide vital services to sustain the modern technological society and its development.

During the last decades, the level of awareness about the importance of protecting our *Critical Infrastructures* (CIs) has been steadily growing. In this respect, US has been the first country to take an official financial commitment by means of the celebrated American Presidential Directive PDD-63 of May 1998 under the Clinton administration. After ten years also the European Community made a similar commitment through the EUDIR Council Directive 2008/114/EC (dated December the 8-th, 2008), that has been afterwards implemented by the EU member states. It has to be noticed that, while the US directive is very broad in its scope, the EU directive is presently limited to the energetic, transport and financial sectors.

The functioning of Critical Infrastructures requires both physical components and human actors. It is therefore important not only to employ reliable components, but also to understand human behaviour at both individual and collective levels. Moreover, each infrastructure resorts to other CIs (typically, but not limited to, energy and ICT) to accomplish its goals: in other words, CIs are *inter-dependent*. Identifying, understanding and analysing critical infrastructure interdependencies is therefore a crucial task to be pursued by the scientific community at both the academic and applied levels [1].

In the development of CIs, the ICT sector has played a crucial role in several respects. ICT pervades any complex activity of modern societies based on communications and represents a fundamental part for the governance of any complex infrastructure. The quality and quantity of information-based services provided to our modern society has been steadily increasing during last 30 years (Web, e-mail, e-commerce, social networking, e-banking, e-health, Web-based entertainment, SCADA systems, etc). In order to improve their performance and to enhance their reliability, the infrastructures have been endowed with increasingly complex connection networks and computerized systems, thus allowing their governance optimization and reducing the humans allocated to that purpose. Nowadays, our

society is on the verge of a new revolution in which the infrastructures are required to become *smart* and to integrate into a *smart* technological environment. Driving the advent of a *smart* society on a painless and secure path represents one of the most difficult challenges for all the technologically advanced countries.

Most of the infrastructures exhibit a network structure. In the last decade, stemming from the availability of large data and based on the statistical physicist perspective of the graph theory, a new paradigm to describe large networks has blossomed: the Network Science [2, 3].

Network Science has revealed a powerful and unifying tool that enables to treat on the same footing widely different networked systems, ranging from biology to sociology to power grids to the Internet and the World Wide Web. In fact, despite their intrinsic differences, all such networks are large systems consisting of simple elementary units (the nodes) interacting via basic mechanisms (represented by the links). Statistical Physics teaches us that this is the case where to expect the occurrence of *emergent behaviours*, i.e., of collective (systemic) effects. In fact, while each component may be perfectly working, the system as a whole can be in a failure state: as an example, think about a big traffic jam, where all the cars, lights, indications, navigators and roads are perfectly functioning and yet everybody is stuck.

Financial networks' analysis represented the forerunner to assess the concept of systemic risk in real infrastructures. Nowadays, several financial institutions consider and employ the global metrics developed by EU network scientists [4] to assess their risk level and robustness consistently with the Basel III Stress Testing [5].

Applying the Network Science paradigm to Inter-dependent Critical Infrastructures has lead to the development of the concept of "Networks of Networks": the *NetONets*. While from the graph-theory point of view a network of networks is just a larger (inhomogeneous) network, in real life infrastructural networks are governed and operated separately and interactions are only allowed at well-defined boundaries. Assessing properties on NetONets instead of that on single networks is like deciding to consider males and females instead of human beings as a single community: depending on the question to answer, either approach may be the most fruitful.

The first applications of the *NetONets* approach to understand critical infrastructures has been related to the propagation of failures in inter-dependent infrastructures modelled as either trees or planar lattices [6, 7]. However, the upheaval of the interest in *NetONets* has followed the publication of a Nature paper on a percolation model of cascade failures in coupled ICT/power networks [8]. Another important step towards real applications has been the analysis of the North America inter-connected electric systems [9] aiming to reduce the global vulnerability of the system.

In general, numerous efforts are nowadays devoted to develop the mathematics of *NetONets*. While most of the current models have a percolative flavour [10–13], some new directions in understanding the dynamics on *NetONets* are being

explored [14–16] resorting to the spectral properties of networks. The European efforts on the subject have recently concentrated in the "MULTIPLEX" project [17] combining top scientists in Complexity and Algorithmic. While the complexity approach allows to concentrate on systemic effects and emergent behavior, other routes have to be considered to perform the analyses of the systems needed for several tasks including management, planning the development, enhancing the security, defining coordinated national and EU/US contingency plans, and assessing the policies at the state and the regional levels. To such an aim, several techniques such as I/O models, federated simulations, agent-based models, time-series analysis are employed. Each of the previous approaches provides a partial perspective of the system behaviour; however to manage and understand the complexity of our society, all of them are required. Our book aims to foster a meta-community able to share and integrate all such perspectives.

This volume is structured along three main sections: part I covers the theoretical approaches, part II provides some applications and part III is devoted to phenomenological modelling. The former taxonomy has been mainly introduced for the sake of presentation. However, due to their inter-disciplinarity, it is difficult to ascribe each contribution to a specific topic only. To improve legibility, each part of the book is endowed with a brief overview of its contents.

We have spent our best efforts to provide the reader with as different contributions as possible; most of the authors have been actively involved in the *NetONets* and related conference series. However, by no means our book can be regarded as exhaustive. Probably, the I/O models [18] represent the most significant lack in our book. Some of the most important topics, such as the systemic risk analysis [19] or time series analysis, would have deserved a more extended treatment. We hope to be able to cover such topics in a nearby future.

Furthermore, there are important topics that are crucial to develop in the nearby future. In particular, the human behaviour, both at the management and at the end user levels, must be accounted for improving the analysis, modelling and simulation of inter-dependent infrastructures. Regarding the complexity approach, it is crucial to build up methodological tools for the statistical analysis of 'small' systems. In fact, while most of the current techniques are aimed to understand the behaviour of the system in the infinite-size limit, almost all infrastructural networks exhibit a relatively small size.

We have tried hard to produce a book that could be regarded as an updated reference for the *NetONets* state-of-the-art. To the same purpose of providing updated information, we have also built a website (netonets.org) wherein to gather and advertise all the initiatives in the field.

One of the main barriers to overcome is the lack of a common language. It is therefore crucial to foster the up-growing *NetONets* community providing a common ground for knowledge sharing. We hope that our efforts will contribute to such a direction.

L'Ace(s)   Gregorio D'Agostino
                                                                              Antonio Scala

References

1. S.M. Rinaldi, J.P. Peerenboom, T.K. Kelly, IEEE Control Systems Magazine 21(6), 11 (2001)
2. D.J. Watts, S.H. Strogatz, Nature 393(6684), 440 (1998)
3. A.L. Barabási, R. Albert, Science 286(5439), 509 (1999)
4. Forecasting financial crisis. URL http://www.focproject.eu/
5. Bank for international settlements. URL http://www.bis.org/bcbs/basel3.htm
6. D. Newman, B. Nkei, B. Carreras, I. Dobson, V. Lynch, P. Gradney, in System Sciences, 2005. HICSS '05. Proceedings of the 38th Annual Hawaii International Conference on (2005), pp. 63c–63c. DOI 10.1109/HICSS.2005.524
7. B.A. Carreras, D.E. Newman, P. Gradney, V.E. Lynch, I. Dobson, in Proceedings of the 40th Annual Hawaii International Conference on System Sciences (IEEE Computer Society, Washington, DC, USA, 2007), HICSS '07, pp. 112–. DOI 10.1109/HICSS.2007.285. URL http://dx.doi.org/10.1109/HICSS.2007.285
8. S. Buldyrev, R. Parshani, G. Paul, H. Stanley, S. Havlin, Nature 464(7291), 1025 (2010)
9. C.D. Brummitt, R.M. D'Souza, E.A. Leicht, Proceedings of the National Academy of Sciences (2012). DOI 10.1073/pnas.1110586109
10. K.M. Lee, J. Kim, W.K. Cho, K.I. Goh, I.M. Kim, New J. Phys. 14, 033027 (2012)
11. E.A. Leicht, R.M. D'Souza, ArXiv e-prints (2009)
12. R. Parshani, S.V. Buldyrev, S. Havlin, Phys. Rev. Lett. 105, 048701 (2010). DOI 10.1103/PhysRevLett.105.048701
13. S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, A.Y. Zyuzin, Phys. Rev. E 78, 056106 (2008). DOI 10.1103/PhysRevE.78.056106. URL http://link.aps.org/doi/10.1103/PhysRevE.78.056106
14. H.Wang, Q. Li, G. D'Agostino, S. Havlin, H.E. Stanley, P. Van Mieghem, Phys. Rev. E (2013)
15. S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C.J. Pérez-Vicente, Y. Moreno, A. Arenas, Phys. Rev. Lett. 110, 028701 (2013).
16. J.M. Hernández, H. Wang, P.V. Mieghem, G. D'Agostino, ArXiv e-prints abs/1304.4731 (2013)
17. Foundational research on multilevel complex networks and systems. URL http://www.multiplexproject.eu/
18. P.D.B. Ronald E. Miller, in Input-Output Analysis, second edition edn. (Cambridge University Press, Cambridge, 2009)
19. W. Kröger, E. Zio, in Vulnerable Systems (Springer London, 2011), pp. 55–64

Contents

Part III   Phenomenological Models

Part I
Theoretical Approaches

This part of the book is devoted to the theoretical approaches to interdependent networks. The state of the art of such a novel and dynamic field is experiencing a continuous growth. Here we have selected, mainly for historical reasons, the contributions stemming from the Statistical Physics approach.

Modelling interdependent networks consists in defining different graphs and the interactions among them. In the multiplex approach, the different layers are modelled by means of different types of links. In the interacting networks approach, the different layers are explicitly modelled as separate networks and the links among them represent the inter-layer interactions.

In Chaps. 1–3 authors rely on 'static' approaches aimed at assessing the robustness and/or the resilience of interdependent systems upon both random failures and targeted attacks. Considering the dynamics of the systems upon continuous stressing leads to the introduction of further effects discussed in Chaps. 4 and 5

Chapter 1
Network of Interdependent Networks: Overview of Theory and Applications

Dror Y. Kenett, Jianxi Gao, Xuqing Huang, Shuai Shao, Irena Vodenska, Sergey V. Buldyrev, Gerald Paul, H. Eugene Stanley and Shlomo Havlin

Abstract Complex networks appear in almost every aspect of science and technology. Previous work in network theory has focused primarily on analyzing single networks that do not interact with other networks, despite the fact that many real-world networks interact with and depend on each other. Very recently an analytical

D. Y. Kenett (✉) · J. Gao · X. Huang · S. Shao · G. Paul · H. E. Stanley
Center for Polymer Studies, Department of Physics, Boston university, Boston, MA 02215, USA
e-mail: drorkenett@gmail.com

X. Huang
e-mail: eqing@bu.edu

S. Shao
e-mail: sshao@bu.edu

G. Paul
e-mail: gerryp@bu.edu

H. E. Stanley
e-mail: hes@bu.edu

J. Gao
Department of Automation, Shanghai Jiao Tong, University, 800 Dongchuan Road, Shanghai 200240, People's Republic of China

Center for Complex Network Research and Department of Physics, Northeastern University, Boston, MA02115, USA
e-mail: jianxi.gao@gmail.com

I. Vodenska
Administrative Sciences Department, Metropolitan College, Boston University, Boston, MA 02215, USA
e-mail: vodenska@bu.edu

S. V. Buldyrev
Department of Physics, Yeshiva University, New York, NY10033, USA
e-mail: buldyrev@verizon.net

S. Havlin
Department of Physics, Bar-Ilan University, Ramat Gan, Israel
e-mail: havlins@gmail.com

framework for studying the percolation properties of interacting networks has been introduced. Here we review the analytical framework and the results for percolation laws for a network of networks (NON) formed by n interdependent random networks. The percolation properties of a network of networks differ greatly from those of single isolated networks. In particular, although networks with broad degree distributions, e.g., scale-free networks, are robust when analyzed as single networks, they become vulnerable in a NON. Moreover, because the constituent networks of a NON are connected by node dependencies, a NON is subject to cascading failure. When there is strong interdependent coupling between networks, the percolation transition is discontinuous (is a first-order transition), unlike the well-known continuous second-order transition in single isolated networks. We also review some possible real-world applications of NON theory.

1.1 Introduction

The interdisciplinary field of network science has attracted great attention in recent years [1–26]. This has taken place because an enormous amount of data regarding social, economic, engineering, and biological systems has become available over the past two decades as a result of the information and communication revolution brought about by the rapid increase in computing power. The investigation and growing understanding of this extraordinary amount of data will enable us to make the infrastructures we use in everyday life more efficient and more robust. The original model of networks, random graph theory, developed in the 1960s by Erdős and Rényi (ER), is based on the assumption that every pair of nodes is randomly connected with the same probability (leading to a Poisson degree distribution). In parallel, lattice networks in which each node has the same number of links have been used in physics to model physical systems. While graph theory was a well-established tool in the mathematics and computer science literature, it could not adequately describe modern, real-world networks. Indeed, the pioneering observation by Barabási in 1999 [2], that many real networks do not follow the ER model but that organizational principles naturally arise in most systems, led to an overwhelming accumulation of supporting data, new models, and novel computational and analytical results, and led to the emergence of a new science: complex networks.

Significant advances in understanding the structure and function of networks, and mathematical models of networks have been achieved in the past few years. These are now widely used to describe a broad range of complex systems, from techno-social systems to interactions amongst proteins. A large number of new measures and methods have been developed to characterize network properties, including measures of node clustering, network modularity, correlation between degrees of neighboring nodes, measures of node importance, and methods for the identification and extraction of community structures. These measures demonstrated that many real networks, and in particular biological networks, contain network motifs—small specific subnetworks—that occur repeatedly and provide information about

functionality [8]. Dynamical processes, such as flow and electrical transport in heterogeneous networks, were shown to be significantly more efficient compared to ER networks [27, 28].

Complex networks are usually non-homogeneous structures that exhibit a power-law form in their degree (number of links per node) distribution. These systems are called scale-free networks. Some examples of real-world scale-free networks include the Internet [3], the WWW [4], social networks representing the relations between individuals, infrastructure networks such as airlines [29, 30], networks in biology, in particular networks of protein-protein interactions [31], gene regulation, and biochemical pathways, and networks in physics, such as polymer networks or the potential energy landscape network. The discovery of scale-free networks has led to a re-evaluation of the basic properties of networks, such as their robustness, which exhibit a character that differs drastically from that of ER networks. For example, while homogeneous ER networks are vulnerable to random failures, heterogeneous scale-free networks are extremely robust [4, 5]. An important property of these infrastructures is their stability, and it is thus important that we understand and quantify their robustness in terms of node and link functionality. Percolation theory was introduced to study network stability and to predict the critical percolation threshold [5]. The robustness of a network is usually (i) characterized by the value of the critical threshold analyzed using percolation theory [32] or (ii) defined as the integrated size of the largest connected cluster during the entire attack process [33]. The percolation approach was also extremely useful in addressing other scenarios, such as efficient attacks or immunization [6, 7, 14, 34, 35], for obtaining optimal path [36] as well as for designing robust networks [33]. Network concepts were also useful in the analysis and understanding of the spread of epidemics [37, 38], and the organizational laws of social interactions, such as friendships [39, 40] or scientific collaborations [41]. Moreira et al. investigated topologically-biased failure in scale-free networks and controlled the robustness or fragility by fine-tuning the topological bias during the failure process [42].

Because current methods deal almost exclusively with individual networks treated as isolated systems, many challenges remain [43]. In most real-world systems an individual network is one component within a much larger complex multi-level network (is part of a network of networks). As technology has advanced, coupling between networks has become increasingly strong. Node failures in one network will cause the failure of dependent nodes in other network, and vice-versa [44]. This recursive process can lead to a cascade of failures throughout the network of networks system. The study of individual particles has enabled physicists to understand the properties of a gas, but in order to understand and describe a liquid or a solid the interactions between the particles also need to be understood. So also in network theory, the study of isolated single networks brings extremely limited results—real-world noninteracting systems are extremely rare in both classical physics and network study. Most real-world network systems continuously interact with other networks, especially since modern technology has accelerated network interdependency.
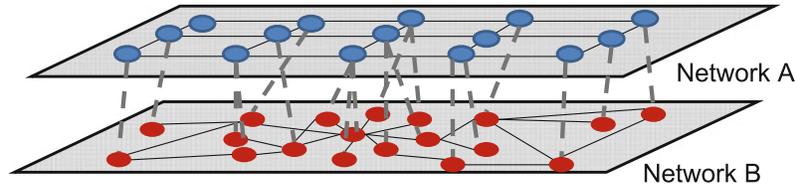
Fig. 1.1 Example of two interdependent networks. Nodes in network B (communications network) are dependent on nodes in network A (power grid) for power; nodes in network A are dependent on network B for control information

To adequately model most real-world systems, understanding the interdependence of networks and the effect of this interdependence on the structural and functional behavior of the coupled system is crucial. Introducing coupling between networks is analogous to the introduction of interactions between particles in statistical physics, which allowed physicists to understand the cooperative behavior of such rich phenomena as phase transitions. Surprisingly, preliminary results on mathematical models [44, 45] show that analyzing complex systems as a network of coupled networks may alter the basic assumptions that network theory has relied on for single networks. Here we will review the main features of the theoretical framework of Network of Networks (NON), and present some real world applications.

1.2 Overview

In order to model interdependent networks, we consider two networks, A and B, in which the functionality of a node in network A is dependent upon the functionality of one or more nodes in network B (see Fig. 1.1), and vice-versa: the functionality of a node in network B is dependent upon the functionality of one or more nodes in network A. The networks can be interconnected in several ways. In the most general case we specify a number of links that arbitrarily connect pairs of nodes across networks A and B. The direction of a link specifies the dependency of the nodes it connects, i.e., link $A_i \rightarrow B_j$ provides a critical resource from node A_i to node B_j. If node A_i stops functioning due to attack or failure, node B_j stops functioning as well but not vice-versa. Analogously, link $B_i \rightarrow A_j$ provides a critical resource from node B_i to node A_j.

To study the robustness of interdependent networks systems, we begin by removing a fraction $1 - p$ of network A nodes and all the A-edges connected to these nodes. As an outcome, all the nodes in network B that are connected to the removed A-nodes by A $\rightarrow$ B links are also removed since they depend on the removed nodes in network A. Their B edges are also removed. Further, the removed B nodes will cause the removal of additional nodes in network A which are connected to the removed B-nodes by B $\rightarrow$ A links. As a result, a cascade of failures that eliminates virtually all nodes in both networks can occur. As nodes and edges are removed, each

network breaks up into connected components (clusters). The clusters in network A (connected by A-edges) and the clusters in network B (connected by B-edges) are different since the networks are each connected differently. If one assumes that small clusters (whose size is below certain threshold) become non-functional, this may invoke a recursive process of failures that we now formally describe.

Our insight based on percolation theory is that when the network is fragmented the nodes belonging to the giant component connecting a finite fraction of the network are still functional, but the nodes that are part of the remaining small clusters become non-functional. Thus in interdependent networks only the giant mutually-connected cluster is of interest. Unlike clusters in regular percolation whose size distribution is a power law with a p-dependent cutoff, at the final stage of the cascading failure process just described only a large number of small mutual clusters and one giant mutual cluster are evident. This is the case because the probability that two nodes that are connected by an A-link and their corresponding two nodes are also connected by a B-link scales as $1/N_B$, where N_B is the number of nodes in network B. So the centrality of the giant mutually-connected cluster emerges naturally and the mutual giant component plays a prominent role in the functioning of interdependent networks. When it exists, the networks preserve their functionality, and when it does not exist, the networks split into fragments so small they cannot function on their own.

We ask three questions: What is the critical $p = p_c$ below which the size of any mutual cluster constitutes an infinitesimal fraction of the network, i.e., no mutual giant component can exist? What is the fraction of nodes $P_\infty(p)$ in the mutual giant component at a given p? How do the cascade failures at each step damage the giant functional component?

Note that the problem of interacting networks is complex and may be strongly affected by variants in the model, in particular by how networks and dependency links are characterized. In the following section we describe several of these model variants.

1.3 Theory of Interdependent Networks

In order to better understand how present-day crucially-important infrastructures interact, Buldyrev et al. [44] recently developed a mathematical framework to study percolation in a system of two coupled interdependent networks subject to cascading failure. Their analytical framework is based on a generating function formalism widely used in studies of single-network percolation and single-network structure [41, 44, 46]. Using the framework to study interdependent networks, we can follow the dynamics of the cascading failures as well as derive analytic solutions for the final steady state. Buldyrev et al. [44] found that interdependent networks were significantly more vulnerable than their noninteracting counterparts. The failure of even a small number of elements within a single network in a system may trigger a catastrophic cascade of events that propagates across the global connectivity. For a

Fig. 1.2 Schematic demonstration of first and second order percolation transitions. In the second order case, the giant component is continuously approaching zero at the percolation threshold $p = p_c$. In the first order case the giant component approaches zero discontinuously. After [47]
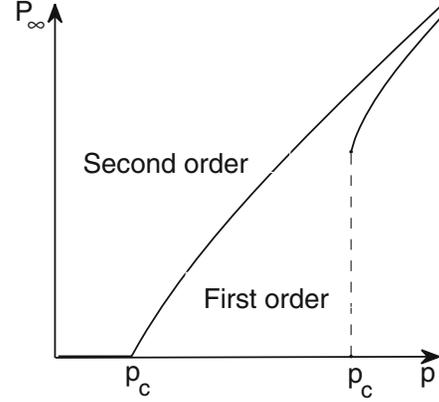
fully coupled case in which each node in one network depends on a functioning node in another network and vice versa, Buldyrev et al. [44] found a first-order discontinuous phase transition, which differs significantly from the second-order continuous phase transition found in single isolated networks (Fig. 1.2). This interesting phenomenon is caused by the presence of two types of links: (i) connectivity links within each network and (ii) dependency links between networks. Parshani et al. [45] showed that, when the dependency coupling between the networks is reduced, at a critical coupling strength the percolation transition becomes second-order.

We now present the theoretical methodology used to investigate networks of interdependent networks (see Ref. [47]), and provide examples from different classes of networks.

1.3.1 Generating Functions for a Single Network

We begin by describing the generating function formalism for a single network that is also useful when studying interdependent networks. Here we assume that all N_i nodes in network i are randomly assigned a degree k from a probability distribution $P_i(k)$, and are randomly connected, the only constraint being that the node with degree k has exactly k links [48]. We define the generating function of the degree distribution

$$G_i(x) \equiv \sum_{k=0}^{\infty} P_i(k)x^k, \tag{1.1}$$

where x is an arbitrary complex variable. The average degree of network i is

$$\langle k \rangle_i = \sum_{k=0}^{\infty} k P_i(k) = \left. \frac{\partial G_i}{\partial x} \right|_{x \to 1} = G_i'(1). \tag{1.2}$$

In the limit of infinitely large networks $N_i \to \infty$, the random connection process can be modeled as a branching process in which an outgoing link of any node has a probability $k P_i(k)/\langle k \rangle_i$ of being connected to a node with degree k, which in turn has $k - 1$ outgoing links. The generating function of this branching process is defined as

$$H_i(x) \equiv \frac{\sum_{k=0}^{\infty} P_i(k)kx^{k-1}}{\langle k \rangle_i} = \frac{G_i'(x)}{G_i'(1)}. \tag{1.3}$$

The probability f_i that a randomly chosen outgoing link does not lead to an infinitely large giant component satisfies a recursive relation $f_i = H_i(f_i)$. Accordingly, the probability that a randomly chosen node does belong to a giant component is given by $g_i = G_i(f_i)$. Once a fraction $1 - p$ of nodes is randomly removed from a network, its generating function remains the same, but must be computed from a new argument $z \equiv px + 1 - p$ [46]. Thus $P_{\infty,i}$, the fraction of nodes that belongs to the giant component, is given by [46],

$$P_{\infty,i} = p g_i(p), \tag{1.4}$$

where

$$g_i(p) = 1 - G_i[p f_i(p) + 1 - p], \tag{1.5}$$

and $f_i(p)$ satisfies

$$f_i(p) = H_i[p f_i(p) + 1 - p]. \tag{1.6}$$

As p decreases, the nontrivial solution $f_i < 1$ of Eq. (1.6) gradually approaches the trivial solution $f_i = 1$. Accordingly, $P_{\infty,i}$—selected as an order parameter of the transition—gradually approaches zero as in the second-order phase transition and becomes zero when two solutions of Eq. (1.6) coincide at $p = p_c$. At this point the straight line corresponding to the right hand side of Eq. (1.6) becomes tangent to the curve corresponding to its left hand side, yielding

$$p_c = 1/H'_i(1). \tag{1.7}$$

For example, for Erdős-Rényi (ER) networks [49–51], characterized by the Poisson degree distribution,

$$G_i(x) = H_i(x) = \exp[\langle k \rangle_i(x - 1)], \tag{1.8}$$

$$g_i(p) = 1 - f_i(p), \tag{1.9}$$

$$f_i(p) = \exp\{p\langle k \rangle_i[f_i(p) - 1]\}, \tag{1.10}$$

and

$$p_c = \frac{1}{\langle k \rangle_i}. \tag{1.11}$$

Finally, using Eqs. (1.4), (1.9), and (1.10), one obtains a direct equation for $P_{\infty,i}$

$$P_{\infty,i} = p[1 - \exp(-\langle k \rangle_i P_{\infty,i})]. \tag{1.12}$$

1.3.2 Two Networks with One-to-One Correspondence of Interdependent Nodes

To initiate an investigation of the multitude of problems associated with interacting networks, Buldyrev et al. [44] restricted themselves to the case of two randomly and independently connected networks with the same number of nodes, specified by their degree distributions $P_A(k)$ and $P_B(k)$. They also assumed every node in the two networks to have one B → A link and one A → B link connecting the same pair of nodes, i.e., the dependencies between networks A and B establish a isomorphism between them that allows us to assume that nodes in A and B coincide (e.g., are at the same corresponding geographic location—if a node in network A fails, the corresponding node in network B also fails, and vice versa). We also assume, however, that the A-edges and B-edges in the two networks are independent.

Unlike the percolation transition in a single network, the mutual percolation transition in this model is a first-order phase transition at which the order parameter (i.e., the fraction of nodes in the mutual giant component) abruptly drops from a finite value at $p_c + \varepsilon$ to zero at $p_c - \varepsilon$. Here ε is a small number that vanishes as the size of network increases $N \to \infty$. In this range of p, a removal of single critical node may lead to a complete collapse of a seemingly robust network. The size of the largest component drops from $N P_\infty$ to a small value, which rarely exceeds 2.

Note that the value of p_c is significantly larger than in single-network percolation. In two interdependent ER networks, for example, $p_c = 2.4554/\langle k \rangle$, while in a single network, $p_c = 1/\langle k \rangle$. For two interdependent scale-free networks with a power-law degree distribution $P_A(k) \sim k^{-\lambda}$, the mutual percolation threshold is $p_c > 0$, even for $2 < \lambda \le 3$, when the percolation threshold in a single network is zero.

Note also that, in this new model, networks with a broader degree distribution are less robust against random attack than networks having a narrower degree distribution but the same average degree. This behavior also differs from that found in single networks. To understand this we note that (i) in interdependent networks, nodes are randomly connected—high degree nodes in one network can connect to low degree nodes in other networks, and (ii) at each time step, failing nodes in one network cause their corresponding nodes (and their edges) in the other network to also fail. Thus although hubs in single networks strongly contribute to network robustness, in interdependent networks they are vulnerable to cascading failure. If a network has a fixed average degree, a broader distribution means more nodes with low degree to balance the high degree nodes. Since the low degree nodes are more easily disconnected the advantage of a broad distribution in single networks becomes a disadvantage in interdependent networks.

The following features have been investigated analytically in Ref. [52], a study that assumes that the degrees of the interdependent nodes exactly coincide, but that both networks are randomly and independently connected by their connectivity links. Reference [52] shows that, for two networks with the same degree distribution $P_A(k)$ of connectivity links and random dependency links, studied in Ref. [44], the fraction of nodes in the giant component is

$$P_\infty = p[1 - G_A(z)]^2, \tag{1.13}$$

where $0 \leq z \leq 1$ is a new variable $z = 1 - p + pf_A$ satisfying equation

$$\frac{[1 - H_A(z)][1 - G_A(z)]}{1 - z} = \frac{1}{p}. \tag{1.14}$$

while in case of coinciding degrees of interdependent nodes Eqs. (1.13) and (1.14) become respectively

$$P_\infty = p[1 - 2G_A(z) + G_A(z^2)] \tag{1.15}$$

and

$$\frac{1 - (1 + z)H_A(z) + zH_A(z^2)}{1 - z} = \frac{1}{p}. \tag{1.16}$$

The left-hand side of Eq. (1.14) always has a single maximum at $0 < z_c < 0$, and the solution abruptly disappear if p becomes less than p_c, the inverse left hand side at z_c. This situation corresponds to the first order transition. In contrast, the left-hand side of Eq. (1.16) has a maximum only if $H'_A(1)$ converges, which corresponds to $\lambda > 3$ when there is a power law tail in the degree distribution. In this case, p_c is the inverse maximum value of the left-hand side of Eq. (1.16), e.g., for ER networks, $p_c = 1.7065/\langle k \rangle$. When $\lambda < 3$, $H'(z)$ diverges for $z \to 1$ and $p_c = 0$, $P_\infty = 0$ as in the case of regular percolation on a single network, for which Eqs. (1.4), (1.5), and (1.6) give

$$P_\infty = p[1 - G_A(z)], \tag{1.17}$$

and

$$\frac{1 - H_A(z)}{1 - z} = \frac{1}{p}. \tag{1.18}$$

Thus for networks with coinciding degrees of the interdependent nodes for $\lambda < 3$, the transition becomes a second-order transition with $p_c = 0$. In the marginal case of $\lambda = 3$, $p_c > 0$, but the transition is second-order.

From Eqs. (1.13)–(1.18) it follows that, if $H'_A(1)$ converges, the networks with coinciding degrees of interdependent nodes are still less robust than single networks, still undergo collapse via a first-order phase transition, but are always more robust than networks with uncorrelated degrees of interdependent nodes. If the average degree is fixed, the robustness of the networks with coinciding degrees of inter-

dependent nodes increases as the degree distribution broadens in the same way as for single networks. Similar observations have been made in numerical studies of interdependent networks with correlated degrees of interdependent nodes [53]. In conclusion, the robustness of interdependent networks increases if the degrees of the interdependent nodes are correlated, i.e., if the hubs are more likely to depend on hubs than on low-degree nodes. For the case of common connectivity links in both networks see Dong et al. [54] and Cellai et al. [55].

1.3.3 Framework of Two Partially Interdependent Networks

A generalization of the percolation theory for two fully interdependent networks was developed by Parshani et al. [45], who studied a more realistic case of a pair of partially-interdependent networks. Here both interacting networks have a certain fraction of completely autonomous nodes whose function does not directly depend on nodes in the other network. They found that when the fraction of autonomous nodes increases above a certain threshold, the collapse of the interdependent networks characterized by a first-order transition observed in Ref. [44] changes, at a critical coupling strength, to a continuous second-order transition as in classical percolation theory [32].

We now describe in more detail the framework developed in [45]. This framework consists of two networks A and B with the number of nodes N_A and N_B, respectively. Within network A, the nodes are randomly connected by A edges with degree distribution $P_A(k)$, and the nodes in network B are randomly connected by B edges with degree distribution $P_B(k)$. In addition, a fraction q_A of network A nodes depends on the nodes in network B and a fraction q_B of network B nodes depends on the nodes in network A. We assume that a node from one network depends on no more than one node from the other network, and if A_i depends on B_j, and B_j depends on A_k, then $k = i$. The latter "no-feedback" condition (see Fig. 1.3) disallows configurations that can collapse without taking into account their internal connectivity [56]. Suppose that the initial removal of nodes from network A is a fraction $1 - p$.

We next present the formalism for the cascade process, step by step (see Fig. 1.4). The remaining fraction of network A nodes after an initial removal of nodes is $\psi'_1 \equiv p$. The initial removal of nodes disconnects some nodes from the giant component. The remaining functional part of network A thus contains a fraction $\psi_1 = \psi'_1 g_A(\psi'_1)$ of the network nodes, where $g_A(\psi'_1)$ is defined by Eqs. (1.5) and (1.6). Since a fraction q_B of nodes from network B depends on nodes from network A, the number of nodes in network B that become nonfunctional is $(1 - \psi_1)q_B = q_B[1 - \psi'_1 g_A(\psi'_1)]$. Accordingly, the remaining fraction of network B nodes is $\phi'_1 = 1 - q_B[1 - \psi'_1 g_A(\psi'_1)]$, and the fraction of nodes in the giant component of network B is $\phi_1 = \phi'_1 g_B(\phi'_1)$.

Following this approach we construct the sequence, ψ'_t and ϕ'_t, of the remaining fraction of nodes at each stage of the cascade of failures. The general form is given by
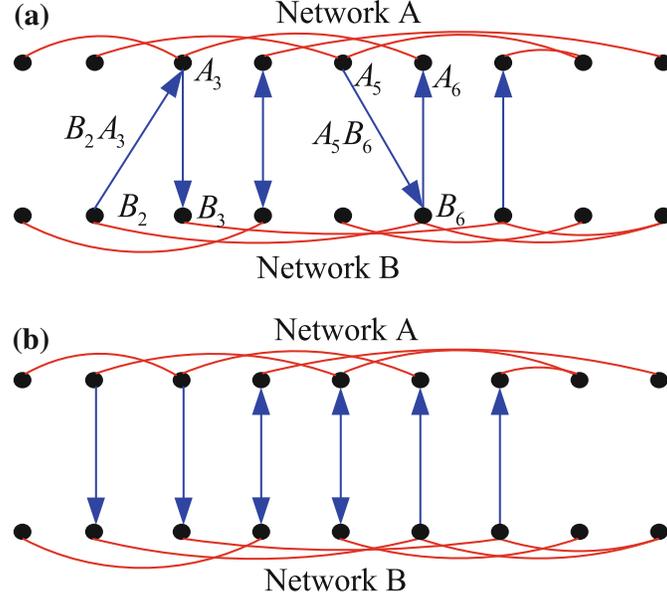
Fig. 1.3 Description of differences between the (**a**) feedback condition and (**b**) no-feedback condition. In the case (**a**), node A_3 depends on node B_2, and node $B_3 \neq B_2$ depends on node A_3, while in case (**b**) this is forbidden. In case (**a**), when $q = 1$ both networks will collapse if one node is removed from one network, which is far from being real. So in our model, we use the no-feedback condition [case (**b**)]. The *blue* links between two networks show the dependency links and the *red* links in each network show the connectivity links which enable each network to function. After [47]

$$
\begin{aligned}
\psi'_1 &\equiv p, \\
\phi'_1 &= 1 - q_B[1 - pg_A(\psi'_1)], \\
\psi'_t &= p[1 - q_A(1 - g_B(\phi'_{t-1}))], \\
\phi'_t &= 1 - q_B[1 - pg_A(\psi'_{t-1})].
\end{aligned}
\tag{1.19}
$$

To determine the state of the system at the end of the cascade process we look at ψ'_τ and ϕ'_τ at the limit of $\tau \to \infty$. This limit must satisfy the equations $\psi'_\tau = \psi'_{\tau+1}$ and $\phi'_\tau = \phi'_{\tau+1}$ since eventually the clusters stop fragmenting and the fractions of randomly removed nodes at step τ and $\tau+1$ are equal. Denoting $\psi'_\tau = x$ and $\phi'_\tau = y$, we arrive at the stationary state to a system of two equations with two unknowns,

$$
\begin{aligned}
x &= p\{1 - q_A[1 - g_B(y)]\}, \\
y &= 1 - q_B[1 - g_A(x)p].
\end{aligned}
\tag{1.20}
$$

The giant components of networks A and B at the end of the cascade of failures are, respectively, $P_{\infty,A} = \psi_\infty = xg_A(x)$ and $P_{\infty,B} = \phi_\infty = yg_B(y)$. The numerical results were obtained by iterating system (1.19), where $g_A(\psi'_t)$ and $g_B(\phi'_t)$ are computed using Eqs. (1.9) and (1.10). Figure 1.5 shows excellent agreement between simulations of cascading failures of two partially interdependent networks with $N = 2 \times 10^5$ nodes and the numerical iterations of system (1.19). In the simu-
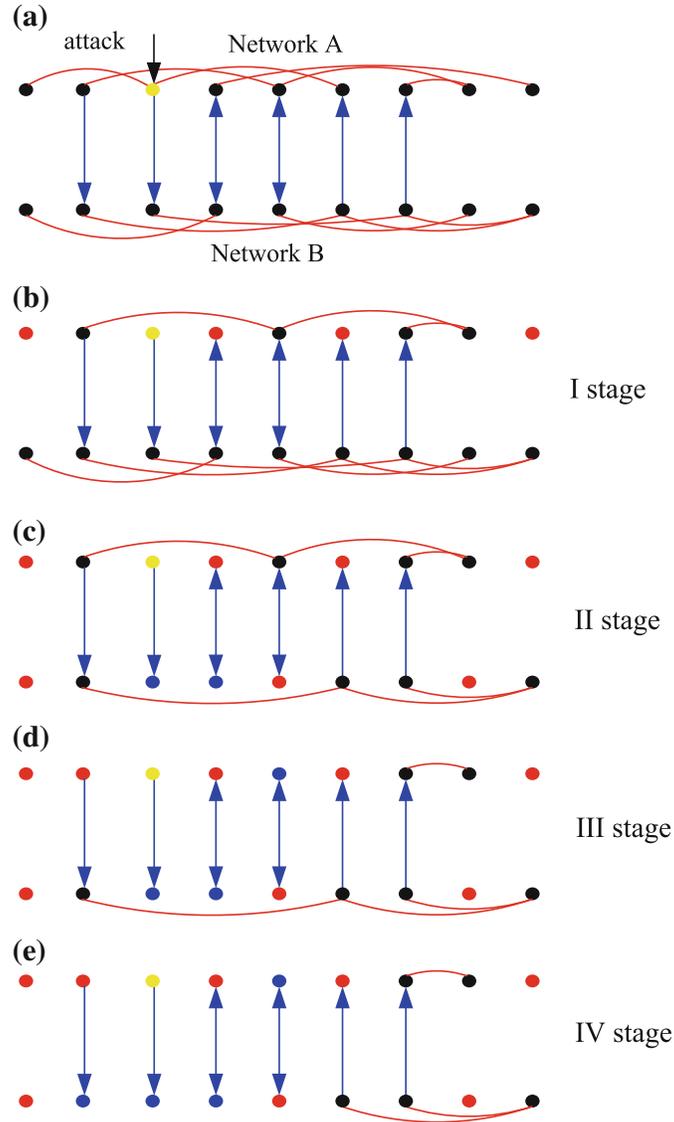
Fig. 1.4 Description of the dynamic process of cascading failures on two partially interdependent networks, which can be generalized to n partially interdependent networks. In this figure, the *black* nodes are the survived nodes, the *yellow* node represents the initially attacked node, the *red* nodes are the nodes removed because they do not belong to the largest cluster, and the *blue* nodes are the nodes removed because they depend on the failed nodes in the other network. In each stage, for one network, we first remove the nodes that depend on the failed nodes in the other network or on the initially attacked nodes. Next we remove the nodes which do not belong to the largest cluster of the network. After [47]

lations, p_c can be determined by the sharp peak in the average number of cascades (iterations), $\langle \tau \rangle$, before the network either stabilizes or collapses [15].

An investigation of Eq. (1.20) can be illustrated graphically by two curves crossing in the (x, y) plane. For sufficiently large q_A and q_B the curves intersect at two points $(0 < x_0, 0 < y_0)$ and $(x_0 < x_1 < 1, y_0 < y_1 < 1)$. Only the second solution (x_1, y_1)
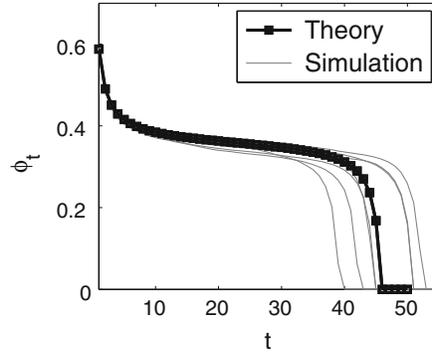
Fig. 1.5 Cascade of failures in two *partially* interdependent ER networks. The giant component ϕ_t for every iteration of the cascading failures is shown for the case of a first order phase transition with the initial parameters $p = 0.8505$, $a = b = 2.5$, $q_A = 0.7$ and $q_B = 0.8$. In the simulations, $N = 2 \times 10^5$ with over 20 realizations. The *gray lines* represent different realizations. The *squares* is the average over all realizations and the *black line* is the theory, Eq. (1.19). After [47]

has any physical meaning. As p decreases, the two solutions become closer to each other, remaining inside the unit square ($0 < x < 1; 0 < y < 1$), and at a certain threshold $p = p_c$ they coincide: $0 < x_0 = x_1 = x_c < 1$, $0 < y_0 = y_1 = y_c < 1$. For $p < p_c$ the curves no longer intersect and only the trivial solution $g_A(x) = g_B(y) = 0$ remains. For sufficiently large q_A and q_B, $P_{\infty,A}$ and $P_{\infty,B}$ as a function of p show a first order phase transition. As q_B decreases, $P_{\infty,A}$ as a function of p shows a second order phase transition. For the graphical representation of Eq. (1.20) and all possible solutions see Fig. 3 in Ref. [45].

In a recent study [33, 57], it was shown that a pair of interdependent networks can be designed to be more robust by choosing the autonomous nodes to be high degree nodes. This choice mitigates the probability of catastrophic cascading failure.

1.3.4 Framework for a Network of Interdependent Networks

In many real systems there are more than two interdependent networks, and diverse infrastructures—water and food supply networks, communications networks, fuel networks, financial transaction networks, or power station networks—can be coupled together [58]. Understanding the way system robustness is affected by such interdependencies is one of the major challenges when designing resilient infrastructures.

Here we review the generalization of the theory of a pair of interdependent networks [44, 45] to a system of n interacting networks [59, 60], which can be graphically represented (see Fig. 1.6) as a network of networks (NON). We review an exact analytical approach for percolation of an NON system composed of n *fully* or *partially* coupled randomly interdependent networks. The approach is based on analyzing the dynamical process of the cascading failures. The results generalize the
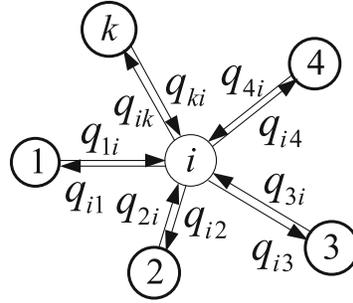
Fig. 1.6 Schematic representation of a network of networks. *Circles* represent interdependent networks, and the *arrows* connect the partially interdependent pairs. For example, a fraction of q_{3i} of nodes in network i depend on the nodes in network 3. The networks which are not connected by the dependency links do not have nodes that directly depend on one another. After [47]

known results for percolation of a single network ($n = 1$) and the $n = 2$ result found in [44, 45], and show that while for $n = 1$ the percolation transition is a second-order transition, for $n > 1$ cascading failures occur and the transition becomes first-order. Our results for n interdependent networks suggest that the classical percolation theory extensively studied in physics and mathematics is a limiting case of $n = 1$ of a general theory of percolation in NON. As we will discuss here, this general theory has many novel features that are not present in classical percolation theory.

In our generalization, each node in the NON is a network itself and each link represents a *fully* or *partially* dependent pair of networks. We assume that each network i ($i = 1, 2, ..., n$) of the NON consists of N_i nodes linked together by connectivity links. Two networks i and j form a partially dependent pair if a certain fraction $q_{ji} > 0$ of nodes of network i directly depends on nodes of network j, i.e., they cannot function if the nodes in network j on which they depend do not function. Dependent pairs are connected by unidirectional dependency links pointing from network j to network i. This convention indicates that nodes in network i get a crucial commodity from nodes in network j, e.g., electric power if network j is a power grid.

We assume that after an attack or failure only a fraction of nodes p_i in each network i will remain. We also assume that only nodes that belong to a giant connected component of each network i will remain functional. This assumption helps explain the cascade of failures: nodes in network i that do not belong to its giant component fail, causing failures of nodes in other networks that depend on the failing nodes of network i. The failure of these nodes causes the direct failure of dependency nodes in other networks, failures of isolated nodes in them, and further failure of nodes in network i and so on. Our goal is to find the fraction of nodes $P_{\infty,i}$ of each network that remain functional at the end of the cascade of failures as a function of all fractions p_i and all fractions q_{ij}. All networks in the NON are randomly connected networks characterized by a degree distribution of links $P_i(k)$, where k is a degree of a node in network i. We further assume that each node a in network i may depend with probability q_{ji} on only one node b in network j with no feed-back condition.

To study different models of cascading failures, we vary the survival time of the dependent nodes after the failure of the nodes in other networks on which they depend, and the survival time of the disconnected nodes. We conclude that the final state of the networks does not depend on these details but can be described by a system of equations somewhat analogous to the Kirchhoff equations for a resistor network. This system of equations has n unknowns x_i. These represent the fraction of nodes that survive in network i after the nodes that fail in the initial attack are removed and the nodes depending on the failed nodes in other networks at the end of cascading failure are also removed, but without taking into account any further node failure due to the internal connectivity of the network. The final giant component of each network is $P_{\infty,i} = x_i g_i(x_i)$, where $g_i(x_i)$ is the fraction of the remaining nodes of network i that belongs to its giant component given by Eq. (1.5).

The unknowns x_i satisfy the system of n equations, [53]

$$x_i = p_i \prod_{j=1}^{K} [q_{ji} y_{ji} g_j(x_j) - q_{ji} + 1], \tag{1.21}$$

where the product is taken over the K networks interlinked with network i by partial dependency links (see Fig. 1.6) and

$$y_{ij} = \frac{x_i}{p_j q_{ji} y_{ji} g_j(x_j) - q_{ji} + 1}, \tag{1.22}$$

is the fraction of nodes in network j that survives after the damage from all the networks connected to network j except network i is taken into account. The damage from network i must be excluded due to the no-feedback condition. In the absence of the no-feedback condition, Eq. (1.21) becomes much simpler since $y_{ji} = x_j$. Equation (1.21) is valid for any case of interdependent NON, while Eq. (1.22) represents the no-feedback condition.

A more the most general case of interdependency links was studied by Shao et al. [56]. They assumed that a node in network i is connected by s supply links to s nodes in network j from which it gets a crucial commodity. If $s = \infty$, the node does not depend on nodes in network j and can function without receiving any supply from them. The generating function of the degree distribution $P^{ij}(s)$ of the supply links $G^{ji}(x) = \sum_{i=0}^{\infty} P^{ji}(s)x^s$ does not include the term $P^{ji}(\infty) = 1 - q_{ji}$, and hence $G_{ji}(1) = q_{ji} \leq 1$. It is also assumed that nodes with $s < \infty$ can function only if they are connected to at least one functional node in network j. In this case, Eq. (1.21) must be changed to

$$x_i = p_i \prod_{j=1}^{K} \{1 - G^{ji}[1 - x_j g_j(x_j)]\}. \tag{1.23}$$

When all dependent nodes have exactly one supply link, $G_{ij}(x) = x q_{ij}$ and Eq. (1.23) becomes

$$x_i = p_i \prod_{j=1}^{K} [1 - q_{ji} + q_{ji} x_j g_j(x_j)], \qquad (1.24)$$

analogous to Eq. (1.21) without the no-feedback condition.

1.3.5 Examples of Classes of Network of Networks

Finally, we present four examples that can be explicitly solved analytically: (i) a tree-like ER NON *fully* dependent, (ii) a tree-like random regular (RR) NON *fully* dependent, (iii) a loop-like ER NON *partially* dependent, and (iv) an RR network of *partially* dependent ER networks. All cases represent different generalizations of percolation theory for a single network.

1.3.5.1 Tree-Like NON of ER Networks

We solve explicitly the case of a tree-like NON (see Fig. 1.7) formed by n ER [49–51] networks with average degrees $k_1, k_2, ...k_i, ..., k_n$, $p_1 = p$, $p_i = 1$ for $i \neq 1$ and $q_{ij} = 1$ (fully interdependent). Using Eqs. (1.21) and (1.22) for x_i and taking into account Eqs. (1.8), (1.9) and (1.10), we find that

$$f_i = \exp\left[-pk_i \prod_{j=1}^{n}(1 - f_j)\right], i = 1, 2, ..., n. \qquad (1.25)$$

These equations can be solved analytically [59]. They have only a trivial solution ($f_i = 1$) if $p < p_c$, where p_c is the mutual percolation threshold. When the n networks have the same average degree k, $k_i = k$ ($i = 1, 2, ..., n$), we obtain from Eq. (1.25) that $f_c \equiv f_i(p_c)$ satisfies

$$f_c = \exp\left[\frac{f_c - 1}{nf_c}\right]. \qquad (1.26)$$

where the solution can be expressed in terms of the Lambert function $W_-(x)$, $f_c = -[nW_-(-\frac{1}{n}e^{-\frac{1}{n}})]^{-1}$, where $W_-(x)$ is the most negative of the two real roots of the Lambert equation $e^{[W(x)]}W(x) = x$ for $x < 0$.

Once f_c is known, we can obtain p_c and the giant component at p_c $P_{\infty,n} \equiv P_\infty$

$$\begin{aligned} p_c &= [nkf_c(1 - f_c)^{(n-1)}]^{-1}, \\ P_\infty(p_c) &= \frac{1-f_c}{nkf_c}. \end{aligned} \qquad (1.27)$$

Chain-like NON   Star-like NON   Tree-like NON
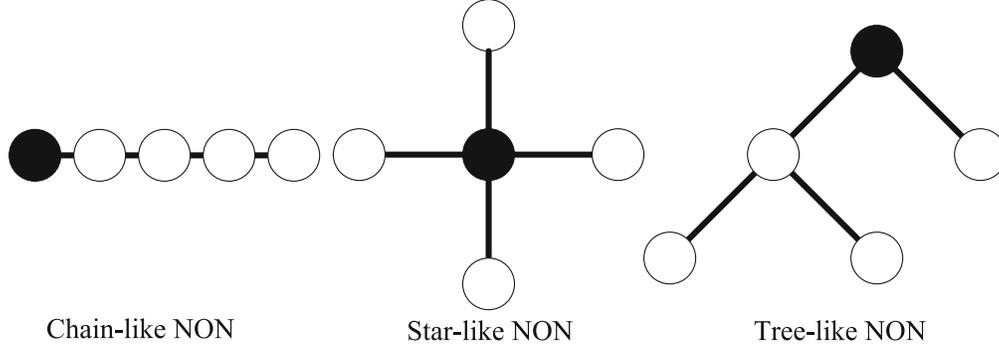
Fig. 1.7 Three types of loopless networks of networks composed of five coupled networks. All have same percolation threshold and same giant component. The *dark* node is the origin network on which failures initially occur. After [47]

Equation (1.27) generalizes known results for $n = 1, 2$. For $n = 1$, we obtain the known result $p_c = 1/k$, Eq. (1.11), of an ER network [49–51] and $P_\infty(p_c) = 0$, which corresponds to a continuous second-order phase transition. Substituting $n = 2$ in Eqs. (1.26) and (1.27) yields the exact results of [44].

From Eqs. (1.21)–(1.22) we obtain an exact expression for the order parameter $P_\infty(p_c)$, the size of the mutual giant component for all p, k, and n values,

$$P_\infty = p[1 - \exp(-kP_\infty)]^n. \tag{1.28}$$

Solutions of Eq. (1.28) are shown in Fig. 1.8a for several values of n. Results are in excellent agreement with simulations. The special case $n = 1$ is the known ER second-order percolation law, Eq. (1.12), for a single network [49–51]. In contrast, for any $n > 1$ the solution of (1.28) yields a first-order percolation transition, i.e., a discontinuity of P_∞ at p_c.

To analyze p_c as a function of n for different k values, we find f_c from Eq. (1.26), substitute it into Eq. (1.27), and obtain p_c. Figure 1.8 shows that the NON becomes more vulnerable with increasing n or decreasing k (p_c increases when n increases or k decreases). Furthermore, when n is fixed and k is smaller than a critical number $k_{\min}(n)$, $p_c \geq 1$, which means that when $k < k_{\min}(n)$ the NON will collapse even if a single node fails. The minimum average degree $k_{\min}$ as a function of the number of networks is

$$k_{\min}(n) = [nf_c(1 - f_c)^{(n-1)}]^{-1}. \tag{1.29}$$

Equations (1.25)–(1.29) are valid for all tree-like structures such as those shown in Fig.1.7. Note that Eq. (1.29) together with Eq. (1.26) yield the value of $k_{\min}(1) = 1$, reproducing the known ER result, that $\langle k \rangle = 1$ is the minimum average degree needed to have a giant component. For $n = 2$, Eq. (1.29) also yields results obtained in [44], i.e., $k_{\min} = 2.4554$.

1.3.5.2 Tree-Like NON of RR Networks

We review the case of a tree-like network of interdependent RR networks [59, 61] in which the degree of each network is assumed to be the same k (Fig. 1.7). By introducing a new variable $r = f^{\frac{1}{k-1}}$ into Eqs. (1.21) and (1.22) and the generating function of RR network [59], the n equations reduce to a single equation

$$r = (r^{k-1} - 1)p(1 - r^k)^{n-1} + 1, \qquad (1.30)$$

which can be solved graphically for any p. The critical case corresponds to the tangential condition leading to critical threshold p_c and P_∞

$$p_c = \frac{r - 1}{(r^{k-1} - 1)(1 - r^k)^{n-1}}, \qquad (1.31)$$

$$P_\infty = p\left(1 - \left\{p^{\frac{1}{n}} P_\infty^{\frac{n-1}{n}}\left[\left(1 - \left(\frac{P_\infty}{p}\right)^{\frac{1}{n}}\right)^{\frac{k-1}{k}} - 1\right] + 1\right\}^k\right)^n. \qquad (1.32)$$

Comparing this with the results of a tree-like ER NON, we find that the robustness of n coupled RR networks of degree k is significantly higher than the n interdependent ER networks of average degree k. Although for an ER NON there exists a critical minimum average degree $k = k_{\min}$ that increases with n below which the system collapses, there is no such analogous $k_{\min}$ for a RR NON system. For any $k > 2$, the RR NON is stable, i.e., $p_c < 1$. In general, this is the case for any network with any degree distribution such that $P_i(0) = P_i(1) = 0$, i.e., for a network without disconnected and singly-connected nodes [61].

1.3.5.3 Loop-Like NON of ER Networks

In the case of a loop-like NON (for dependencies in one direction) of n ER networks, all the links are unidirectional and the no-feedback condition is irrelevant. If the initial attack on each network is the same $1 - p$, $q_{i-1i} = q_{n1} = q$, and $k_i = k$, using Eqs. (1.21) and (1.22) we find that P_∞ satisfies

$$P_\infty = p(1 - e^{-kP_\infty})(q P_\infty - q + 1). \qquad (1.33)$$

Note that when $q = 1$ Eq. (1.33) has only a trivial solution $P_\infty = 0$, but when $q = 0$ it yields the known giant component of a single network, Eq. (1.12), as expected. We present in Fig. 1.8b numerical solutions of Eq. (1.33) for two values of q. Note that when $q = 1$ and the structure is tree-like, Eqs. (1.28) and (1.32) depend on n, but for loop-like NON structures, Eq. (1.33) is independent of n.

1.3.5.4 RR Network of ER Networks

Now we review results [47] for a NON in which each ER network is dependent on exactly m other ER networks. This system represents the case of RR network of ER networks. We assume that the initial attack on each network is $1 - p$, and each partially dependent pair has the same q in both directions with no-feedback condition. The n equations of Eq. (1.21) are exactly the same due to symmetries, and hence p_c and P_∞ can be solved analytically,

$$p_c^{II} = \frac{1}{k(1-q)^m},$$ (1.34)

$$P_\infty = \frac{p}{2^m}(1 - e^{-kP_\infty})[1 - q + \sqrt{(1-q)^2 + 4qP_\infty}]^m.$$ (1.35)

where p_c^{II} denotes the critical threshold for the second order phase transition.

Again, as in the case of the loop-like structure, it is surprising that both the critical threshold and the giant component do not depend on the number of networks n, in contrast to tree-like NON, but only on the coupling q and on both degrees k and m. Numerical solutions of Eq. (1.35) are shown in Fig. 1.8. In the special case of $m = 0$, Eqs. (1.34) and (1.35) coincide with the known results for a single ER network, Eqs. (1.11) and (1.12) separately. It can be shown that when $q < q_c$ we have "weak coupling" represented by a second-order phase transition and when $q_c < q < q_{max}$ we have "strong coupling" and a first-order phase transition. When $q > q_{max}$ the system become unstable due to the "very strong coupling" between the networks. In the last case, removal of a single node in one network may lead to the collapse of the NON.

1.3.6 Resilience of Networks to Targeted Attacks

In real-world scenarios, initial system failures seldom occur randomly and can be the result of targeted attacks on central nodes. Such attacks can also occur in less central nodes in an effort to circumvent central node defences, e.g., heavily-connected Internet hubs tend have more effective firewalls. Targeted attacks on high degree nodes [4, 6, 7, 13, 42] or high betweenness nodes [62] in *single* networks dramatically affect their robustness. To study the targeted attack problem on interdependent networks [13, 63–65] we assign a value $W_\alpha(k_i)$ to each node, which represents the probability that a node i with k_i degree will be initially attacked and become inactive, i.e.,

$$W_\alpha(k_i) = \frac{k_i^\alpha}{\sum_{i=1}^{N} k_i^\alpha}, \quad -\infty < \alpha < +\infty.$$ (1.36)

When $\alpha > 0$, higher-degree nodes are more vulnerable to intentional attack. When $\alpha < 0$, higher-degree nodes are less vulnerable and have a lower probability of failure. The case $\alpha = 0$, $W_0 = \frac{1}{N}$, represents the random removal of nodes [44].
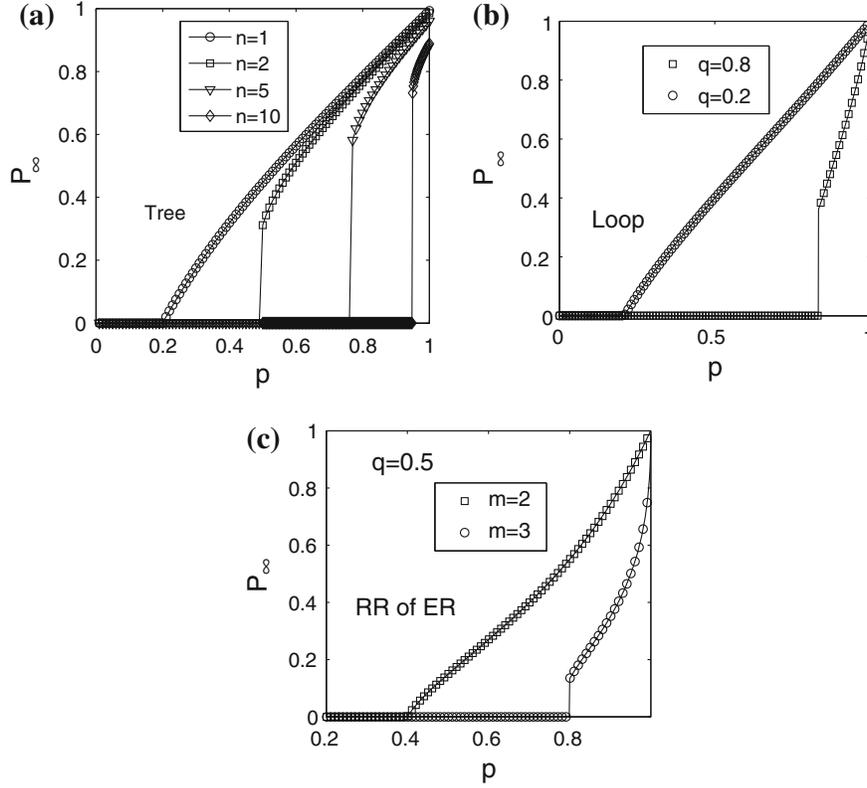
Fig. 1.8 The fraction of nodes in the giant component P_∞ as a function of p for three different examples discussed in Sects. 1.3.5.2–1.3.5.4. (**a**) For a tree-like *fully* ($q = 1$) interdependent NON is shown P_∞ as a function of p for $k = 5$ and several values of n. The results obtained using Eq. (1.28). Note that increasing n from $n = 2$ yields a first order transition. (**b**) For a loop-like NON, P_∞ as a function of p for $k = 6$ and two values of q. The results obtained using Eq. (1.33). Note that increasing q yields a first order transition. (**c**) For an RR network of ER networks, P_∞ as a function of p, for two different values of m when $q = 0.5$. The results are obtained using Eq. (1.35) , and the number of networks, n, can be any number with the condition that any network in the NON connects exactly to m other networks. Note that changing m from 2 to $m > 2$ changes the transition from second order to first order (for $q = 0.5$). Simulation results are in excellent agreement with theory. After [47]

In the interdependent networks model with networks A and B described in Ref. [44], a fraction $1 - p$ of the nodes from one network are removed with a probability $W_\alpha(k_i)$ [Eq. (1.36)]. The cascading failures are then the same as those described in Ref. [44]. To analytically solve the targeted attack problem we must find an equivalent network A', such that the *targeted* attack problem on interdependent networks A and B can be solved as a *random* attack problem on interdependent networks A' and B. We begin by finding the new degree distribution of network A after using Eq. (1.36) to remove a $1 - p$ fraction of nodes but before the links of the remaining nodes that connect to the removed nodes are removed. If $A_p(k)$ is the number of nodes with degree k and $P_p(k)$ the new degree distribution of the remaining fraction p of nodes in network A, then

$$P_p(k) = \frac{A_p(k)}{pN}.$$ (1.37)

When another node is removed, $A_p(k)$ changes as

$$A_{(p-1/N)}(k) = A_p(k) - \frac{P_p(k)k^\alpha}{\langle k(p)^\alpha \rangle},$$ (1.38)

where $\langle k(p)^\alpha \rangle \equiv \sum P_p(k)k^\alpha$. In the limit of $N \to \infty$, Eq. (1.38) can be presented in terms of a derivative of $A_p(k)$ with respect to p,

$$\frac{dA_p(k)}{dp} = N\frac{P_p(k)k^\alpha}{\langle k(p)^\alpha \rangle}.$$ (1.39)

Differentiating Eq. (1.37) with respect to p and using Eq. (1.39), we obtain

$$-p\frac{dP_p(k)}{dp} = P_p(k) - \frac{P_p(k)k^\alpha}{\langle k(p)^\alpha \rangle},$$ (1.40)

which is exact for $N \to \infty$. In order to solve Eq. (1.40), we define a function $G_\alpha(x) \equiv \sum_k P(k)x^{k^\alpha}$, and substitute $f \equiv G_\alpha^{-1}(p)$. We find by direct differentiation that [46]

$$P_p(k) = P(k)\frac{f^{k^\alpha}}{G_\alpha(f)} = \frac{1}{p}P(k)f^{k^\alpha},$$ (1.41)

$$\langle k(p)^\alpha \rangle = \frac{fG_\alpha'(f)}{G_\alpha(f)},$$ (1.42)

satisfy the Eq. (1.40). With this degree distribution, the generating function of the nodes left in network A before removing the links to the removed nodes is

$$G_{Ab}(x) \equiv \sum_k P_p(k)x^k = \frac{1}{p}\sum_k P(k)f^{k^\alpha}x^k.$$ (1.43)

Because network A is randomly connected, the probability of a link emanating from a remaining node is equal to the ratio of the number of links emanating from the remaining nodes to the total number of links emanating from all the nodes of the original network,

$$\tilde{p} \equiv \frac{pN\langle k(p) \rangle}{N\langle k \rangle} = \frac{\sum_k P(k)kf^{k^\alpha}}{\sum_k P(k)k},$$ (1.44)

where $\langle k \rangle$ is the average degree of the original network A, and $\langle k(p) \rangle$ is the average degree of remaining nodes before the links that are disconnected are removed. Removing the links that connect to the deleted nodes of a randomly connected network is equivalent to randomly removing a $(1 - \tilde{p})$ fraction of links of the remaining
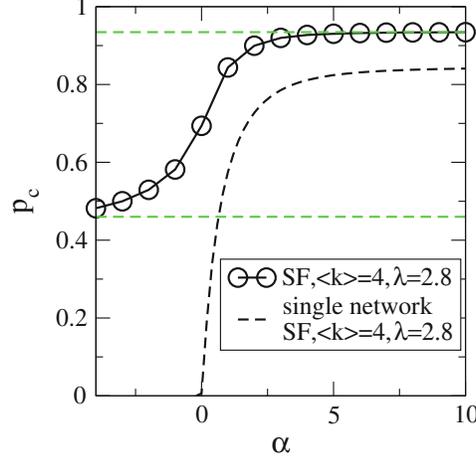
Fig. 1.9 Dependence of p_c on α for SF single and interdependent networks with average degree $\langle k \rangle = 4$ for targeted attacks described in Sect. 3.5. The *lower* cut-off of the degree is $m = 2$. The horizontal lines represent the *upper* and *lower* limits of p_c. The *black dashed line* represents p_c for single SF network. After [63]

nodes. It is known [46] that the generating function of the remaining nodes after random removal of $(1 - \tilde{p})$ fraction of links is equal to the original distribution of the network with a new argument $z = 1 - \tilde{p} + x\tilde{p}$. Thus the generating function of the new degree distribution of the nodes left in network A after their links to the removed nodes are also removed is

$$G_{Ac}(x) \equiv G_{Ab}(1 - \tilde{p} + \tilde{p}x). \tag{1.45}$$

The only difference in the cascading process under *targeted* attack from the case under *random* attack is in the first stage when network A is attacked. If we find a network A' with generating function $\tilde{G}_{A0}(x)$ such that after a random attack with a $(1 - p)$ fraction of nodes removed the generating function of nodes left in A' is the same as $G_{Ac}(x)$, then the targeted attack problem on interdependent networks A and B can be solved as a random attack problem on interdependent networks A' and B. We find $\tilde{G}_{A0}(x)$ by solving the equation $\tilde{G}_{A0}(1 - p + px) = G_{Ac}(x)$ and from, Eq. (1.45),

$$\tilde{G}_{A0}(x) = G_{Ab}(1 + \frac{\tilde{p}}{p}(x - 1)). \tag{1.46}$$

This formalism allows us to map the problem of cascading node failure in interdependent networks caused by an initial *targeted* attack to the problem of *random* attack. We note that the evolution of equations only depends on the generating function of network A, and not on any information concerning how the two networks interact with each other. Thus this approach can be applied to the study of other general interdependent network models.

Finally we analyze the specific class of scale-free (SF) networks. Figure 1.9 shows the critical thresholds p_c of SF networks. Note that p_c in interdependent SF networks

is nonzero for the entire range of α because failure of the least-connected nodes in one network may lead to failure of well-connected nodes in a second network, making interdependent networks significantly more difficult to protect than a single network. A significant role in the vulnerability to random attacks is also played by network assortativity [66].

1.3.7 Interdependent Clustered Networks

Clustering quantifies the propensity of two neighbors of the same vertex to also be neighbors of each other, forming triangle-shaped configurations in the network [1, 10, 67]. Unlike random networks in which there is little or no clustering, real-world networks exhibit significant clustering. Recent studies have shown that, for single isolated networks, both bond percolation and site percolation have percolation and epidemic thresholds that are higher than those in unclustered networks [68–73]. Here we review a mathematical framework for understanding how the robustness of interdependent networks is affected by clustering within the network components. We extend the percolation method developed by Newman [68] for single clustered networks to coupled clustered networks. Huang et al. [65] found that interdependent networks that exhibit significant clustering are more vulnerable to random node failure than networks with low significant clustering. They studied two networks, A and B, each having the same number of nodes N. The N nodes in A and B have bidirectional dependency links to each other, establishing a one-to-one correspondence. Thus the functioning of a node in network A depends on the functioning of the corresponding node in network B and vice versa. Each network is defined by a joint degree distribution P_{st} (generating function $G_0(x, y) = \sum_{s,t=0}^{\infty} P_{st} x^s y^t$) that specifies the fraction of nodes connected to s single edges and t triangles [68]. The conventional degree of each node is thus $k = s + 2t$. The clustering coefficient c is

$$c = \frac{\sum_{st} t P_{st}}{\sum_k k(k-1)P(k)/2}. \tag{1.47}$$

1.3.7.1 Percolation on Interdependent Clustered Networks

To study how clustering within interdependent networks affects a system's robustness, we apply the interdependent networks framework [44]. In interdependent networks A and B, a fraction $(1 - p)$ of nodes is first removed from network A. Then the size of the giant components of networks A and B in each cascading failure step is defined to be $p_1, p_2, ..., p_n$, which are calculated iteratively

$$\begin{aligned} p_n &= \mu_{n-1} g_A(\mu_{n-1}), \text{ n is odd,} \\ p_n &= \mu_n g_B(\mu_n), \text{ n is even,} \end{aligned} \tag{1.48}$$

where $\mu_0 = p$ and μ_n are intermediate variables that satisfy

$$\begin{aligned} \mu_n &= pg_A(\mu_{n-1}), \text{ n is odd,} \\ \mu_n &= pg_B(\mu_{n-1}), \text{ n is even.} \end{aligned} \tag{1.49}$$

As interdependent networks A and B form a stable mutually-connected giant component, $n \to \infty$ and $\mu_n = \mu_{n-2}$, the fraction of nodes left in the giant component is p_∞. This system satisfies

$$\begin{aligned} x &= pg_A(y), \\ y &= pg_B(x), \end{aligned} \tag{1.50}$$

where the two unknown variables x and y can be used to calculate $p_\infty = xg_B(x) = yg_A(y)$. Eliminating y from these equations, we obtain a single equation

$$x = pg_A[pg_B(x)]. \tag{1.51}$$

The critical case ($p = p_c$) emerges when both sides of this equation have equal derivatives,

$$1 = p^2 \frac{dg_A}{dx}[pg_B(x)]\frac{dg_B}{dx}(x)|_{x=x_c, p=p_c}, \tag{1.52}$$

which, together with Eq. (1.51), yields the solution for p_c and the critical size of the giant mutually-connected component, $p_\infty(p_c) = x_c g_B(x_c)$.

Consider for example the case in which networks A and B have Poisson degree distributions P_{st}^A and P_{st}^B for both s and t:

$$\begin{aligned} P_{st}^A &= e^{-\mu_A - \nu_A} \frac{\mu_A^s \nu_A^t}{s!t!}, \\ P_{st}^B &= e^{-\mu_B - \nu_B} \frac{\mu_B^s \nu_B^t}{s!t!}. \end{aligned} \tag{1.53}$$

Using techniques in Ref. [68] it is possible to show that in this case $x = p(1 - u_A)$, $y = p(1 - u_B)$, where

$$\begin{aligned} u_A &= v_A = e^{[\mu_A y + 2y(1-y)\mu_A](u_A-1) + \nu_A p^2(v_A^2-1)}, \\ u_B &= v_B = e^{[\mu_B x + 2x(1-x)\mu_B](u_B-1) + \nu_B p^2(v_B^2-1)}. \end{aligned} \tag{1.54}$$

If the two networks have the same clustering, $\mu \equiv \mu_A = \mu_B$ and $\nu \equiv \nu_A = \nu_B$, p_∞ is then

$$p_\infty = p(1 - e^{\nu p_\infty^2 - (\mu+2\nu)p_\infty})^2. \tag{1.55}$$

Here μ and ν are the average number of single links and triangles per node respectively.

The giant component, p_∞, for interdependent clustered networks can thus be obtained by solving Eq. (1.55). Note that when $\nu = 0$ we obtain from Eq. (1.55) the
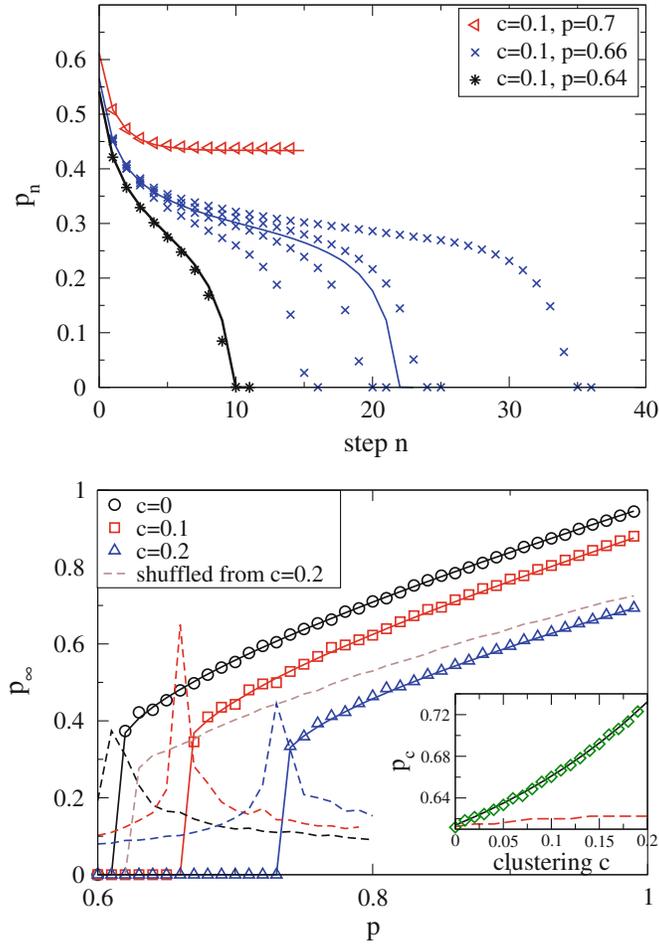
Fig. 1.10 Behavior of interdependent networks with different clustering coefficients. **a** Size of mutually connected giant component as a function of cascading failure steps n. Results are for $c = 0.1$, $p = 0.64$ (*below p_c*), $p = 0.66$ (at p_c) and $p = 0.7$ (*above p_c*). Lines represent theory (Eqs. (1.48) and (1.49)) and dots are from simulations. Note that at p_c there are large fluctuations. **b** Size of giant component, p_∞, in interdependent networks with both networks having clustering via Poisson degree distributions of Eq. (1.53) and average degree $\langle k \rangle = \mu_A + 2\nu_A = 4$, as a function of p. *Dashed lines* are number of interactions (NOI) before cascading failure stops obtained by simulation [74]. *Inset*: Green line is the critical threshold p_c in interdependent networks as function of clustering coefficient c. *Red* dashed line represents critical threshold of shuffled interdependent networks which originally has clustering coefficient c. The shuffled networks have zero clustering and degree-degree correlation, but has the same degree distribution as the original clustered networks. Symbols and *dashed lines* represent simulation, *solid curves* represent theoretical results. After [65]

result obtained in Ref. [44] for random interdependent ER networks. Figure 1.10, using numerical simulation, compares the size of the giant component after n stages of cascading failure with the theoretical prediction of Eq. (1.48). When $p = 0.7$ and $p = 0.64$, which are not near the critical threshold ($p_c = 0.6609$), the agreement with simulation is perfect. Below and near the critical threshold, the simulation initially agrees with the theoretical prediction but then deviates for large n due to the random

fluctuations of structure in different realizations [44]. By solving Eq. (1.55), we have p_∞ as a function of p in Fig. 1.10 for a given average degree and several values of clustering coefficients. The figure shows that the interdependent networks with higher clustering become less robust than the networks with low clustering and the same average degree k, i.e., p_c is a monotonically increasing function of c (see inset of Fig. 1.10).

1.4 Application to Infrastructure

In interacting networks, the failure of nodes in one network generally leads to the failure of dependent nodes in other networks, which in turn may cause further damage to the first network, leading to cascading failures and catastrophic consequences. It is known, for example, that blackouts in various countries have been the result of cascading failures between interdependent systems such as communication and power grid systems [75] (Fig. 1.11). Furthermore, different kinds of critical infrastructures are also coupled together, e.g., systems of water and food supply, communications, fuel, financial transactions, and power generation and transmission (Fig. 1.11). Modern technology has produced infrastructures that are becoming increasingly interdependent, and understanding how robustness is affected by these interdependencies is one of the major challenges faced when designing resilient infrastructures [56, 58, 75, 76].

Blackouts are a demonstration of the important role played by the dependencies between networks. For example, the 28 September 2003 blackout in Italy resulted in a widespread failure of the railway network, healthcare systems, and financial services and, in addition, severely influenced communication networks. The partial failure of the communication system in turn further impaired the power grid management system, thus producing a negative feedback on the power grid. This example emphasizes how interdependence can significantly magnify the damage in an interacting network system [44, 45, 58, 75].

Thus understanding the coupling and interdependencies of networks will enable us to design and implement future infrastructures that are more efficient and robust.

1.5 Application to Finance and Economics

Financial and economic networks are neither static nor independent of one another. As global economic convergence progresses, countries increasingly depend on each other through such links as trade relations, foreign direct investments, and flow of funds in international capital markets. Economic systems such as real estate markets, bank borrowing and lending operations, and foreign exchange trading are interconnected and constantly affect each other. As economic entities and financial markets become increasingly interconnected, a shock in a financial network can provoke

How interdependent are infrastructures?

Peerenboom, Fisher, and Whitfield, 2001
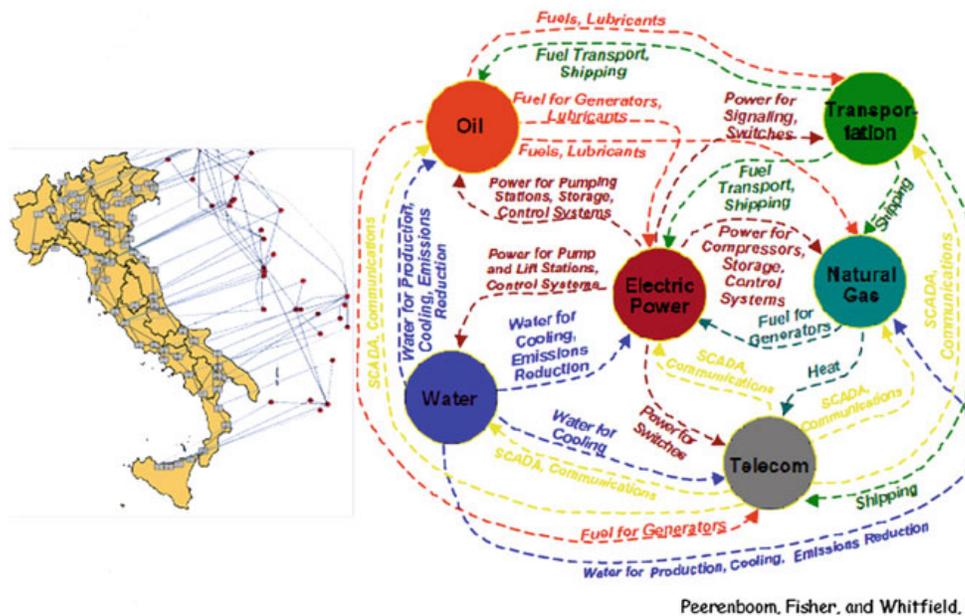
Fig. 1.11 *Left*: Power grid and Internet dependence in Italy. Analysis of this system can explain the cascade failure that led to the 2003 blackout. *Right*: Inter-dependence of fundamental infrastructures. A further example is a recent event in Cyprus (July 2011), where an explosion caused a failure of the electrical power lines, which in turn caused the countries water supply to shut down, due to the strong coupling between these two networks

significant cascading failures throughout the global economic system. Based on the success of complex networks in modeling interconnected systems, applying complex network theory to study economic systems has been given much attention [77–84].

The strong connectivity in financial and economic networks allows catastrophic cascading node failure to occur whenever the system experiences a shock, especially if the shocked nodes are hubs or are highly central in the network [7, 63, 76, 85, 86]. To thus minimize systemic risk, financial and economic networks should be designed to be robust to external shocks.

In the wake of the recent global financial crisis, increased attention has been given to the study of the dynamics of economic systems and to systemic risk in particular. The widespread impact of the current EU sovereign debt crisis and the 2008 world financial crisis show that, as economic systems become increasingly interconnected, local exogenous or endogenous shocks can provoke global cascading system failure that is difficult to reverse and that can cripple the system for a prolonged period of time. Thus policy makers are compelled to create and implement safety measures that prevent cascading system failures or that soften their systemic impact.

To study the systemic risk to financial institutions, we analyze a coupled (bipartite) bank-asset network in which a link between a bank and a bank asset exists when the bank has the asset on its balance sheet. Recently, Huang et al. [87] presented a

model that focuses on real estate assets to examine banking network dependencies on real estate markets. The model captures the effect of the 2008 real estate market failure on the US banking network. Between 2000 and 2007, 27 banks failed in the US, but between 2008 and early 2013 the number rose to over 470. The model proposes a cascading failure algorithm to describe the risk propagation process during crises. This methodology was empirically tested with balance sheet data from US commercial banks for the year 2007, and model predictions are compared with the actual failed banks in the US after 2007 as reported by the Federal Deposit Insurance Corporation (FDIC). The model identifies a significant portion of the actual failed banks, and the results suggest that this methodology could be useful for systemic risk stress testing for financial systems. The model also indicates that commercial rather than residential real estate markets were the major culprits for the failure of over 350 US commercial banks during the period 2008–2011.

There are two main channels of risk contagion in the banking system, (i) direct interbank liability linkages between financial institutions and (ii) contagion via changes in bank asset values. The former, which has been given extensive empirical and theoretical study [88–92], focuses on the dynamics of loss propagation via the complex network of direct counterpart exposures following an initial default. The latter, based on bank financial statements and financial ratio analysis, has received scant attention. A financial shock that contributes to the bankruptcy of a bank in a complex network will cause the bank to sell its assets. If the financial market's ability to absorb these sales is less than perfect, the market prices of the assets that the bankrupted bank sells will decrease. Other banks that own similar assets could also fail because of loss in asset value and increased inability to meet liability obligations. This imposes further downward pressure on asset values and contributes to further asset devaluation in the market. Damage in the banking network thus continues to spread, and the result is a cascading of risk propagation throughout the system [93, 94].

Using this coupled bank-asset network model, we can test the influence of each particular asset or group of assets on the overall financial system. If the value of agricultural assets drop by 20 %, we can determine which banks are vulnerable to failure and offer policy suggestions, e.g., requiring mandatory reduction in exposure to agricultural loans or closely monitoring the exposed bank, to prevent such failure.

The model shows that sharp transitions can occur in the coupled bank-asset system and that the network can switch between two distinct regions, stable and unstable, which means that the banking system can either survive and be healthy or collapse. Because it is important that policy makers keep the world economic system in the stable region, we suggest that our model for systemic risk propagation might also be applicable to other complex financial systems, e.g., to model how sovereign debt value deterioration affects the global banking system or how the depreciation or appreciation of certain currencies impact the world economy.

1.5.1 Cascading Failures in the US Banking System

During the recent financial crisis, 371 US commercial banks failed between 1 January 2008 and 1 July 2011. The Failed Bank List from the Federal Deposit Insurance Corporation (FBL-FDIC) records the names of failed banks and the dates of their failure. We use this list as an experimental benchmark for our model. The dataset used as input to the model is the US Commercial Banks Balance Sheet Data (CBBSD) from Wharton Research Data Services, which contains the amount of assets in each category that the US commercial banks have on their balance sheets.

To build a sound bank-asset coupled system network and systemic risk cascading failure model, it is important to study the properties of the failed banks and compare them with the properties of the banks that survive. Thus the asset portfolios of commercial banks containing asset categories such as commercial loans or residential mortgages are carefully examined. The banks are modeled according to how they construct their asset portfolios (see the upper panel of Fig. 1.12). For each bank, the CBBSD contains 13 different non-overlapping asset categories, e.g., bank i owns amounts $B_{i,0}$, $B_{i,1}$, ..., $B_{i,12}$ of each asset, respectively. The total asset value B_i and total liability value L_i of a bank i are obtained from CBBSD dataset. The weight of each asset m in the overall asset portfolio of a bank i is then defined as $w_{i,m} \equiv B_{i,m}/B_i$. From the perspective of the asset categories, we define the *total market value* of an asset m as $A_m \equiv \sum_i B_{i,m}$. Thus the market share of bank i in asset m is $s_{i,m} \equiv B_{i,m}/A_m$.

Studying the properties of failed banks between 2008 and 2011 reveals that, for certain assets, asset weight distributions for all banks differ from the asset weight distributions for failed banks. Failed banks cluster in a region heavily weighted with construction and development loans and loans secured by nonfarm nonresidential properties while having fewer agricultural loans in their asset portfolios than the banks that survived. These results confirm the nature of the most recent financial crisis of 2008–2011 in which bank failures were largely caused by real estate-based loans, including loans for construction and land development and loans secured by nonfarm nonresidential properties [95]. In this kind of financial crisis, banks with greater agricultural loan assets are more financially robust [96]. Failed banks also tend to have lower equity-to-asset ratios, i.e., higher leverage ratios than the banks that survived during the financial crisis of 2008–2011 [97].

A financial crisis usually starts with the bursting of an economic or financial bubble. For example, with the bursting of the dot-com bubble, the technology-heavy NASDAQ Composite index lost 66 % of its value, plunging from 5048 in 10 March 2000 to 1720 in 2 April 2001. In our current model, the shock in the bank-asset coupled system originated with the real estate bubble burst. The two categories of real estate assets most relevant to the failure of commercial banks during the 2008–2011 financial crisis were construction and land development loans and loans secured by nonfarm and non-residential properties. Although it is widely believed that the financial crisis was caused by residential real estate assets, the coupled bank-asset network model does not find evidence that loans secured by 1–4 family
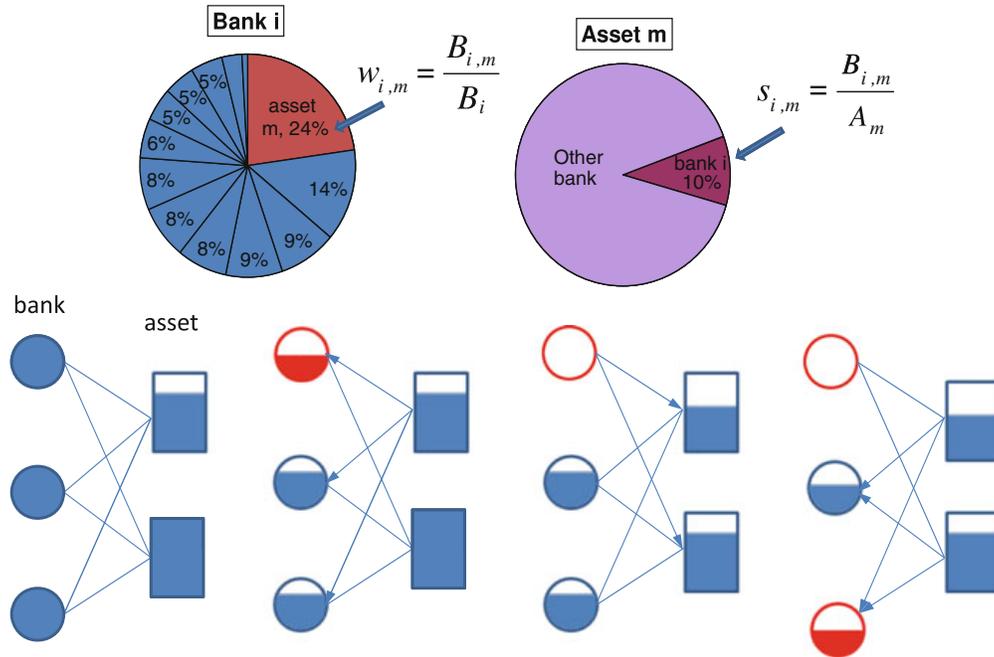
Fig. 1.12 Bank-asset coupled network model with banks as one node type and assets as the other node type. Link between a bank and an asset exists if the bank has the asset on its balance sheet. *Upper* panel: illustration of bank-node and asset-node. $B_{i,m}$ is the amount of asset m that bank i owns. Thus, a bank i with total asset value B_i has $w_{i,m}$ fraction of its total asset value in asset m. $s_{i,m}$ is the fraction of asset m that the bank holds out. *Lower* panel: illustration of the cascading failure process. The *rectangles* represent the assets and the *circles* represent the banks. From *left* to *right*, initially, an asset suffers loss in value which causes all the related banks' total assets to shrink. When a bank's remaining asset value is below certain threshold (e.g., the bank's total liability), the bank fails. Failure of the bank elicits disposal of bank assets which further affects the market value of the assets. This adversely affects other banks that hold this asset and the total value of their assets may drop *below* the threshold which may result in further bank failures. This cascading failure process propagates back and forth between banks and assets until no more banks fail. After [87]

residential properties were responsible for the commercial bank failures. This result is consistent with the conclusion of Ref. [95]: that the cause of commercial bank failure between 2008 and 2011 were commercial real estate-based loans rather than residential mortgages. For more details regarding the coupled bank-asset model see Ref. [87].

1.6 Summary and Outlook

In summary, this paper presents the recently-introduced mathematical framework of a Network of Networks (NON). In interacting networks, when a node in one network fails it usually causes dependent nodes in other networks to fail which, in turn, may cause further damage in the first network and result in a cascade of

failures with catastrophic consequences. Our analytical framework enables us to follow the dynamic process of the cascading failures step-by-step and to derive steady state solutions. Interdependent networks appear in all aspects of life, nature, and technology. Examples include (i) transportation systems such as railway networks, airline networks, and other transportation systems [53, 98]; (ii) the human body as studied by physiology, including such examples of interdependent NON systems as the cardiovascular system, the respiratory system, the brain neuron system, and the nervous system [99]); (iii) protein function as studied by biology, treating protein interaction—the many proteins involved in numerous functions—as a system of interacting networks; (iv) the interdependent networks of banks, insurance companies, and business firms as studied by economics; (v) species interactions and the robustness of interaction networks to species loss as studied by ecology, in which it is is essential to understand the effects of species decline and extinction [100]; and (vi) the topology of statistical relationships between distinct climatologically variables across the world as studied by climatology [101].

Thus far only a few real-world interdependent systems have been thoroughly analyzed [53, 98]. We expect our work to provide insights leading further analysis of real data on interdependent networks. The benchmark models presented here can be used to study the structural, functional, and robustness properties of interdependent networks. Because in real-world NONs individual networks are not randomly connected and their interdependent nodes are not selected at random, it is crucial that we understand the many types of correlation that exist in real-world systems and that we further develop the theoretical tools to take them into account. Further studies of interdependent networks should focus on (i) an analysis of real data from many different interdependent systems and (ii) the development of mathematical tools for studying real-world interdependent systems. Many real networks are embedded in space, and the spatial constraints strongly affect their properties [20, 102, 103]. There is a need to understand how these spatial constraints influence the robustness properties of interdependent networks [98]. Other properties that influence the robustness of single networks, such as the dynamic nature of the configuration in which links or nodes appear and disappear and the directed nature of some links, as well as problems associated with degree-degree correlations and clustering, should be also addressed in future studies of coupled network systems. An additional critical issue is the improvement of the robustness of interdependent infrastructures. Our studies thus far shown that there are three methods of achieving this goal (i) by increasing the fraction of autonomous nodes [45], (ii) by designing dependency links such that they connect the nodes with similar degrees [44, 53], and (iii) by protecting the high-degree nodes against attack [33]. Achieving this goal will provide greater safety and stability in today's socio-techno world.

Networks dominate every aspect of present-day living. The world has become a global village that is steadily shrinking as the ways that human beings interact and connect multiply. Understanding these connections in terms of interdependent networks of networks will enable us to better design, organize, and maintain the future of our socio-techno-economic world.

References

1. D.J. Watts, S.H. Strogatz, Nature **393**(6684), 440 (1998)
2. A.L. Barabási, R. Albert, Science **286**(5439), 509 (1999)
3. M. Faloutsos, P. Faloutsos, C. Faloutsos, in *ACM SIGCOMM Computer Communication Review*, vol. 29 (ACM, 1999), vol. 29, pp. 251–262
4. R. Albert, H. Jeong, A.L. Barabási, Nature **406**(6794), 378 (2000)
5. R. Cohen, K. Erez, D. Ben-Avraham, S. Havlin, Physical Review Letters **85**(21), 4626 (2000)
6. D.S. Callaway, M.E. Newman, S.H. Strogatz, D.J. Watts, Physical Review Letters **85**(25), 5468 (2000)
7. R. Cohen, K. Erez, D. Ben-Avraham, S. Havlin, Physical Review Letters **86**(16), 3682 (2001)
8. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Science Signaling **298**(5594), 824 (2002)
9. D.J. Watts, Proceedings of the National Academy of Sciences **99**(9), 5766 (2002)
10. M.E.J. Newman, SIAM review **45**(2), 167 (2003)
11. A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani, Proceedings of the National Academy of Sciences of the United States of America **101**(11), 3747 (2004)
12. M.E.J. Newman, M. Girvan, Physical Review E **69**(2), 026113 (2004)
13. L.K. Gallos, R. Cohen, P. Argyrakis, A. Bunde, S. Havlin, Physical Review Letters **94**(18), 188701 (2005)
14. V. Latora, M. Marchiori, Physical Review E **71**(1), 015103 (2005)
15. C. Song, S. Havlin, H.A. Makse, Nature **433**(7024), 392 (2005)
16. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, Physics Reports **424**(4), 175 (2006)
17. M.E.J. Newman, A.L. Barabasi, D.J. Watts, *The structure and dynamics of networks* (Princeton University Press, 2011)
18. B.J. West, P. Grigolini, *Complex webs: anticipating the improbable* (Cambridge University Press, 2010)
19. G. Bonanno, G. Caldarelli, F. Lillo, R.N. Mantegna, Physical Review E **68**(4), 046130 (2003)
20. D. Li, K. Kosmidis, A. Bunde, S. Havlin, Nature Physics **7**(6), 481 (2011)
21. D.Y. Kenett, M. Tumminello, A. Madi, G. Gur-Gershgoren, R. Mantegna, E. Ben-Jacob, PloS one **5**(12), e15032 (2010)
22. D.Y. Kenett, T. Preis, G. Gur-Gershgoren, E. Ben-Jacob, International Journal of Bifurcation and Chaos **22**(7), 1250181 (2012)
23. Y.N. Kenett, D.Y. Kenett, E. Ben-Jacob, M. Faust, PloS one **6**(8), e23912 (2011)
24. A. Madi, D. Kenett, S. Bransburg-Zabary, Y. Merbl, F. Quintana, S. Boccaletti, A. Tauber, I. Cohen, E. Ben-Jacob, Chaos**21**(1), 016109 (2011)
25. S. Bransburg-Zabary, D.Y. Kenett, G. Dar, A. Madi, Y. Merbl, F.J. Quintana, A.I. Tauber, I.R. Cohen, E. Ben-Jacob, Physical Biology **10**(2), 025003 (2013)
26. A. Majdandzic, B. Podobnik, S.V. Buldyrev, D.Y. Kenett, S. Havlin, H.E. Stanley, Spontaneous recovery in dynamical networks, Nature Physics, doi:10.1038/nphys2819 (2013)
27. E. López, S.V. Buldyrev, S. Havlin, H.E. Stanley, Physical Review Letters **94**(24), 248701 (2005)
28. M. Boguná, D. Krioukov, Physical Review Letters **102**(5), 058701 (2009)

29. V. Colizza, A. Barrat, M. Barthelemy, A. Vespignani, et al., Proc. Natl. Acad. Sci. USA 103 (2005)
30. Z. Wu, L.A. Braunstein, V. Colizza, R. Cohen, S. Havlin, H.E. Stanley, Physical Review E **74**(5), 056104 (2006)
31. R. Albert, A.L. Barabási, Reviews of modern physics **74**(1), 47 (2002)
32. A. Bunde, S. Havlin, *Fractals and disordered systems* (Springer, Berlin Heidelberg, 1991)
33. C.M. Schneider, A.A. Moreira, J.S. Andrade, S. Havlin, H.J. Herrmann, Proceedings of the National Academy of Sciences **108**(10), 3838 (2011)
34. Y. Chen, G. Paul, S. Havlin, F. Liljeros, H.E. Stanley, Physical Review Letters **101**(5), 058701 (2008)
35. R. Cohen, S. Havlin, D. Ben-Avraham, Physical Review Letters **91**(24), 247901 (2003)
36. L.A. Braunstein, S.V. Buldyrev, R. Cohen, S. Havlin, H.E. Stanley, Physical Review Letters **91**(16), 168701 (2003)
37. R. Pastor-Satorras, A. Vespignani, Physical review letters **86**(14), 3200 (2001)
38. D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J.J. Ramasco, A. Vespignani, Proceedings of the National Academy of Sciences **106**(51), 21484 (2009)
39. G. Palla, I. Derényi, I. Farkas, T. Vicsek, Nature **435**(7043), 814 (2005)
40. G. Kossinets, D.J. Watts, Science **311**(5757), 88 (2006)
41. M.E.J. Newman, Proceedings of the National Academy of Sciences **98**(2), 404 (2001)
42. A.A. Moreira, J.S. Andrade Jr, H.J. Herrmann, J.O. Indekeu, Physical Review Letters **102**(1), 018701 (2009)
43. S. Havlin, D.Y. Kenett, E. Ben-Jacob, A. Bunde, R. Cohen, H. Hermann, J. Kantelhardt, J. Kertész, S. Kirkpatrick, J. Kurths, et al., European Physical Journal-Special Topics **214**(1), 273 (2012)
44. S. Buldyrev, R. Parshani, G. Paul, H. Stanley, S. Havlin, Nature **464**(7291), 1025 (2010)
45. R. Parshani, S.V. Buldyrev, S. Havlin, Physical Review Letters **105**(4), 048701 (2010)
46. J. Shao, S.V. Buldyrev, L.A. Braunstein, S. Havlin, H.E. Stanley, Physical Review E **8**0(3), 036105 (2009)
47. J. Gao, S.V. Buldyrev, H.E. Stanley, S. Havlin, Nature Physics **8**(1), 40 (2011)
48. M. Molloy, B. Reed, Combinatorics probability and computing **7**(3), 295 (1998)
49. P. Erdős, A. Rényi, Publ. Math. Debrecen **6**, 290 (1959)
50. P. Erdős, A. Rényi, Publ. Math. Inst. Hungar. Acad. Sci **5**, 17 (1960)
51. B. Bollobás, *Graph theory*, vol. 62 (North Holland, 1982)
52. S.V. Buldyrev, N.W. Shere, G.A. Cwilich, Physical Review E **83**(1), 016112 (2011)
53. R. Parshani, C. Rozenblat, D. Ietri, C. Ducruet, S. Havlin, EPL (Europhysics Letters) 92(6), 68002 (2010)
54. D. Gaogao, J. Gao, R. Du, L. Tian, H.E. Stanley, S. Havlin, Robustness of network of networks under targeted attack, Physical Review **87**(5), 052804 (2013)
55. D. Cellai, E. López, J. Zhou, J.P. Gleeson, G. Bianconi, Percolation in multiplex networks with overlap, Physical Review **88**(5), 052811 (2013)
56. J. Shao, S.V. Buldyrev, S. Havlin, H.E. Stanley, Physical Review E **83**(3), 036116 (2011)
57. C.M. Schneider, N.A. Araujo, S. Havlin, H.J. Herrmann, arXiv:1106.3234 [cond-mat.stat-mech] (2011)
58. S.M. Rinaldi, J.P. Peerenboom, T.K. Kelly, Control Systems, IEEE **21**(6), 11 (2001)
59. J. Gao, S.V. Buldyrev, S. Havlin, H.E. Stanley, Physical Review Letters **107**(19), 195701 (2011)
60. Gao et al, Percolation of a general network of networks, PRE (in press, 2014)
61. J. Gao, S. Buldyrev, S. Havlin, H. Stanley, Physical Review E **85**(6), 066134 (2012)
62. P. Holme, B.J. Kim, C.N. Yoon, S.K. Han, Physical Review E **65**(5), 056109 (2002)
63. X. Huang, J. Gao, S.V. Buldyrev, S. Havlin, H.E. Stanley, Physical Review E **83**(6), 065101 (2011)
64. T. Tanizawa, S. Havlin, H.E. Stanley, Physical Review E **85**(4), 046109 (2012)
65. X. Huang, S. Shao, H. Wang, S.V. Buldyrev, H.E. Stanley, S. Havlin, EPL (Europhysics Letters) **101**(1), 18002 (2013)

66. D. Zhou, H.E. Stanley, G. D'Agostino, A. Scala, Phys. Rev. E **86**, 066103 (2012)
67. M.A. Serrano, M. Boguna, Physical Review E **74**(5), 056114 (2006)
68. M.E.J. Newman, Physical Review Letters **103**(5), 058701 (2009)
69. J.C. Miller, Physical Review E **80**(2), 020901 (2009)
70. J.P. Gleeson, S. Melnik, Physical Review E **80**(4), 046121 (2009)
71. J.P. Gleeson, S. Melnik, A. Hackett, Physical Review E **81**(6), 066114 (2010)
72. C. Molina, L. Stone, Journal of Theoretical Biology (2012)
73. B. Karrer, M.E.J. Newman, Physical Review E **82**(6), 066118 (2010)
74. R. Parshani, S.V. Buldyrev, S. Havlin, Proceedings of the National Academy of Sciences **108**(3), 1007 (2011)
75. V. Rosato, L. Issacharoff, F. Tiriticco, S. Meloni, S. Porcellinis, R. Setola, International Journal of Critical Infrastructures **4**(1), 63 (2008)
76. A. Vespignani, Nature **464**(7291), 984 (2010)
77. R.M. May, S.A. Levin, G. Sugihara, Nature **451**(7181), 893 (2008)
78. A. Garas, P. Argyrakis, C. Rozenblat, M. Tomassini, S. Havlin, New journal of Physics **12**(11), 113043 (2010)
79. N. Johnson, T. Lux, Nature **469**(7330), 302 (2011)
80. A.G. Haldane, R.M. May, Nature **469**(7330), 351 (2011)
81. F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, D.R. White, science **325**(5939), 422 (2009)
82. S. Battiston, M. Puliga, R. Kaushik, P. Tasca, G. Caldarelli, Scientific Reports **2** (2012)
83. D.Y. Kenett, M. Raddant, T. Lux, E. Ben-Jacob, PloS one **7**(2), e31144 (2012)
84. D.Y. Kenett, M. Raddant, L. Zatlavi, T. Lux, E. Ben-Jacob, International Journal of Modern Physics Conference Series **16**(1), 13 (2012)
85. A.E. Motter, Y.C. Lai, Physical Review E **66**(6), 065102 (2002)
86. A.G. Smart, L.A. Amaral, J.M. Ottino, Proceedings of the National Academy of Sciences **105**(36), 13223 (2008)
87. X. Huang, I. Vodenska, S. Havlin, H.E. Stanley, Scientific reports **3** (2013)
88. S. Wells, Financial Stability Review **13**, 175 (2002)
89. C.H. Furfine, Journal of Money, Credit and Banking pp. 111–128 (2003)
90. C. Upper, A. Worms, European Economic Review **48**(4), 827 (2004)
91. H. Elsinger, A. Lehar, M. Summer, Management science **52**(9), 1301 (2006)
92. E. Nier, J. Yang, T. Yorulmazer, A. Alentorn, Journal of Economic Dynamics and Control **31**(6), 2033 (2007)
93. R. Cifuentes, G. Ferrucci, H.S. Shin, Journal of the European Economic Association **3**(2–3), 556 (2005)
94. I. Tsatskis, Available at SSRN 2062174 (2012)
95. R.A. Cole, L.J. White, Journal of Financial Services Research **42**(1–2), 5 (2012)
96. G.S. Corner, Central Banker (Fall) (2011)
97. Y. Gopalan, Central Banker (Spring) (2010)
98. C.G. Gu, S.R. Zou, X.L. Xu, Y.Q. Qu, Y.M. Jiang, H.K. Liu, T. Zhou, et al., Physical Review E **84**(2), 026101 (2011)
99. A. Bashan, R.P. Bartsch, J.W. Kantelhardt, S. Havlin, P.C. Ivanov, Nature communications **3**, 702 (2012)
100. M.J. Pocock, D.M. Evans, J. Memmott, Science **335**(6071), 973 (2012)
101. J.F. Donges, H.C. Schultz, N. Marwan, Y. Zou, J. Kurths, The European Physical Journal B **84**(4), 635 (2011)
102. M. Barthélemy, Physics Reports **499**(1), 1 (2011)
103. W. Li, A. Bashan, S.V. Buldyrev, H.E. Stanley, S. Havlin, Physical Review Letters **108**(22), 228702 (2012)

Chapter 2
Avalanches in Multiplex and Interdependent Networks

G. J. Baxter, S. N. Dorogovtsev, A. V. Goltsev and J. F. F. Mendes

Abstract Many real-world complex systems are represented not by single networks but rather by sets of interdependent networks. In these specific networks, vertices in each network mutually depend on vertices in other networks. In the simplest representative case, interdependent networks are equivalent to the so-called multiplex networks containing vertices of one sort but several kinds of edges. Connectivity properties of these networks and their robustness against damage differ sharply from ordinary networks. Connected components in ordinary networks are naturally generalized to viable clusters in multiplex networks whose vertices are connected by paths passing over each individual sort of their edges. We examine the robustness of the giant viable cluster to random damage. We show that random damage to these systems can lead to the avalanche collapse of the viable cluster, and that this collapse is a hybrid phase transition combining a discontinuity and the critical singularity. For this transition we identify latent critical clusters associated with the avalanches triggered by a removal of single vertices. Divergence of their mean size signals the approach to the hybrid phase transition from one side, while there are no critical precursors on the other side. We find that this discontinuous transition occurs in scale-free multiplex networks whenever the mean degree of at least one of the interdependent networks does not diverge.

2.1 Introduction

The network representation of complex systems is successfully exploited in various sciences [1]. Numerous real-world systems, however, cannot be represented by a single network. Instead, they consist of several interacting networks. In simple sit-

G. J. Baxter (✉) · S. N. Dorogovtsev · A. V. Goltsev · J. F. F. Mendes
Departamento de Física, I3N, Universidade de Aveiro, Campus Universitário de Santiago,
3810-193, Aveiro, Portugal
e-mail: gjbaxter@ua.pt
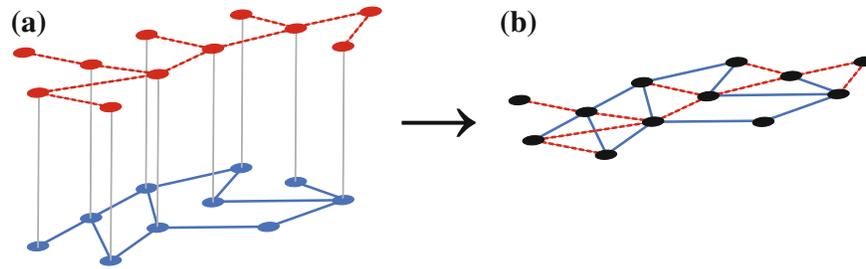
Fig. 2.1 a Two interdependent networks. A vertex in one network has a mutual dependence, represented by *grey vertical lines*, on zero or one vertex in the other network. **b** This can be reduced to a multiplex network by merging the mutually dependent vertices, and representing the edges of each network by different kinds or *colours* of edges

uations, these interactions can be represented by interlinks connecting vertices in different networks [2, 3]. When these interconnections and edges in all these networks are identical, then it is possible to describe the structural organization of this set of networks and the statistics of its connected components similarly to ordinary networks [4]. Here we consider significantly more interesting systems in which vertices in each network mutually depend on vertices in other networks in the sense that the removal (or, generally, change of the state) of a vertex in one network immediately leads to the removal (or change of the state) of its neighbour in another network. These interdependent networks describe numerous complex systems, both natural [5], and man-made [6, 7]. Importantly, the interdependencies can make a system more fragile: damage to one element can lead to avalanches of failures throughout the system [8, 9]. Recent theoretical investigations of interdependent networks consisting of two [10] or more [11] subnetworks have shown that small initial failures can cascade back and forth through the networks, leading, at some critical point, to the collapse of the whole system in a discontinuous phase transition.

In the original formulation of the problem [10] the researchers focused on the final result of the removal of a finite $1 - p$ fraction of vertices from one of the interdependent networks. This removal leads to a complicated infinite (for infinite networks) cascade in back-and-forth damage propagation. Below a critical point p_c, this cascade of failures eliminates the interdependent networks completely, while above the transition, the cascade sweeps out a finite fraction of the networks. Son et al. [12] showed the original approach of studying two interdependent networks can be simplified, if one uses the equivalence of a wide class of interdependent networks to a multiplex network problem. They proposed a simple mapping from the model used in [10] in which a vertex in one network has a mutual dependence on no more than one vertex in the other network, to a multiplex network with one kind of vertex but two kinds of edges. The mapping is achieved by simply merging the mutually dependent vertices from the two networks. Figure 2.1 explains this mapping. In graph theory, the multiplex networks are also called graphs with coloured edges.

As we will see, the phase transition in this system is discontinuous, and hybrid in nature, in contrast to ordinary percolation that occurs as a continuous phase tran-

sition. The difference between hybrid and continuous phase transitions is that the hybrid transition has a discontinuity like a first-order transition, but exhibits critical behavior near the transition, like a second-order transition. Moreover, the hybrid transition is asymmetric: critical correlations appear on only one side of the critical point, whereas they appear on both sides of a continuous phase transition. Another intriguing phenomenon appearing at the critical point of the hybrid transition is scale-invariant avalanches that are absent in a continuous phase transition. Each avalanche is triggered by removal of a single vertex and results in the elimination of multiple vertices. To highlight this principal difference from continuous phase transitions, let us compare with, for example, the continuous percolation phase transition. This is a second order phase transition in an equilibrium system. Percolation can be represented as the removal of uniformly randomly chosen vertices. Removal of a vertex can only split a cluster (connected component) into smaller clusters but, it cannot trigger an avalanche.

In this chapter we describe these discontinuous phase transitions. Our aim is to expand and deepen the understanding of the nature of the phase transition and the avalanche collapse in interdependent and multiplex networks. This understanding has been lacking until recently. We investigate the damage caused by the removal of a single node chosen at random from an infinite network. The removal of a single vertex causes an avalanche of damage (so named to distinguish it from the cascades of failures mentioned above, which are caused by the sudden removal of a finite fraction of the vertices in the network). Our method allows the identification of individual avalanches and the study of their structure.

Why is the problem of the avalanches triggered by the removal of a single vertex principally important and attractive for researchers? The reason is that the statistics of these individual avalanches reveals the critical divergence at the phase transition point. To understand a phase transition, it is not sufficient to obtain an equation showing the emergence of a non-zero order parameter. For continuous and hybrid transitions, one should also find the divergence of susceptibility associated with this transition, and also describe critical correlations. It is avalanches that are responsible for critical correlations. The mean size of the individual avalanches triggered by a randomly removed vertex plays a role of susceptibility and diverges at the critical point manifesting the hybrid transition. The second reason, with a practical perspective, is that knowledge of the organization of individual avalanches enables one to control them and increase robustness of the system.

In the remainder of this chapter, then, we will generally consider multiplex networks, but it should be noted that the results are identical to those for two interdependent networks as defined above, and may be qualitatively extended to interdependent networks in general. The results presented in this chapter are based on results obtained in our paper [13].

This chapter is organised as follows. In Sect. 2.2 we define the multiplex network model, and give an algorithm for identifying the viable clusters. In Sect. 2.3 we derive basic equations for the size of the giant viable cluster, and show how the location and scaling of the transition may be obtained. In Sect. 2.4 we analyse the structure and statistics of the avalanches associated with the transition. These results are extended
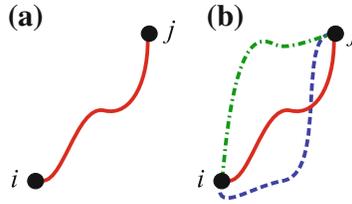
Fig. 2.2 **a** In an ordinary network, two vertices i and j belong to the same cluster if there is a path connecting them. **b** In a multiplex network, vertices i and j belong to the same viable cluster if there is a path connecting them for every kind of edge, following only edges of that kind. In the example shown, there are $m = 3$ kinds of edges. Vertices i and j are said to be 3-connected

to the special case of scale-free networks in Sect. 2.5. Results are summarised in Sect. 2.6.

2.2 Viable Clusters and Algorithm

In ordinary networks, two vertices are connected if there is a path between the vertices. Based on this notion, one introduces clusters of connected vertices and studies emergence of the giant connected component of a graph. In multiplex network, this notion of connection between vertices must be modified. We consider a set of vertices connected by m different types of edges. The connections are essential to the function of each site, so that a vertex is only *viable* if it maintains connections of every type to other viable vertices. A *viable cluster* is defined as follows: For every kind of edge, and for any two vertices i and j within a viable cluster, there must be a path from i to j following only edges of that kind. In other words, in multiplex network with m types of edges, two vertices are m-connected if for every type of edges there is a path between these vertices. Based on this definition, a viable cluster is then a cluster of m-connected vertices. Figure 2.2 explains the viable clusters. In a large system, we wish to find when there is a giant cluster of viable vertices. From this definition of viable clusters, it follows that any giant viable cluster is a subgraph of the giant connected component of each of the m networks formed by considering only a single type of edge in the multiplex network. The absence of, at least, any one giant connected component means the absence of the giant viable cluster. Note that viable clusters are simple generalization of clusters of connected vertices in ordinary networks with a single type of edges. The important difference is that in a multiplex networks we demand that vertices in a viable cluster must be connected by every type of edges (m-connected). It is this additional condition that leads to discontinuous emergence of the giant viable cluster as a result of a hybrid phase transition in contrast to a continuous phase transition in ordinary percolation.

The viable clusters of any size may be identified by an iterative pruning algorithm, based on the principles of percolation. Here we give such an algorithm for identifying viable clusters that may be implemented, for example, in a computer program for investigations of the resilience of real-world multiplex networks.
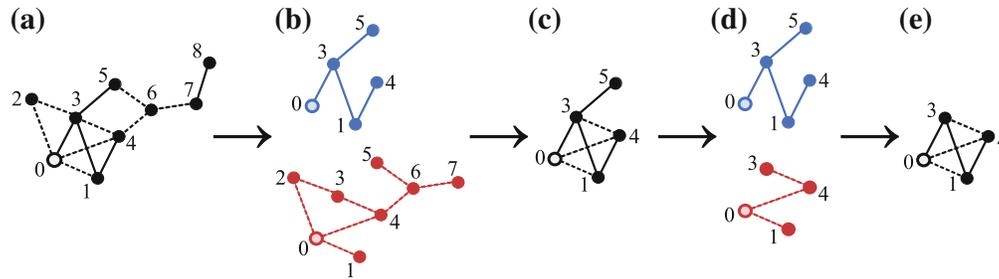
Fig. 2.3 An example demonstrating the algorithm for identifying a viable cluster in a small network with two kinds of edges. **a** In the original network, in step (i) we select vertex 0 as the test vertex. **b** In step (ii) we identify the clusters of vertices connected to 0 by each kind of edge. **c** Step (iii): the intersection of these two clusters forms becomes the new candidate set for the viable cluster to which 0 belongs. **d** We repeat steps (ii) using only vertices from the candidate set shown in **c**. Repeating step (iii), we find the overlap between the two clusters from **d**, shown in **e**. Further repetition of steps (ii) and (iii) does not change this cluster, meaning that the cluster consisting of vertices 0, 1, 3 and 4 is a viable cluster
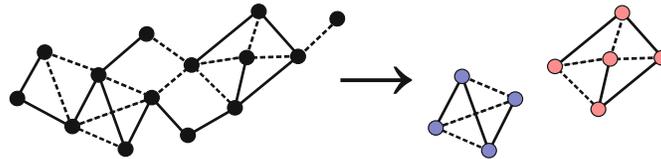
Fig. 2.4 A small network with two kinds of edges (*left*). Applying the algorithm described in the text, non-viable vertices are removed, leaving two viable clusters (*right*)

Consider a multiplex network, with vertices $i = 1, 2, \ldots, N$ connected by m kinds of edges labeled $s = a, b, \ldots$. Viable clusters in any multiplex network may be identified by the following algorithm.

(i) Choose a test vertex i at random from the network.
(ii) For each kind of edge s, compile a list of vertices that can be reached from i by following only edges of type s.
(iii) The intersection of these m lists forms a new candidate set for the viable cluster containing i.
(iv) Repeat steps (ii) and (iii) but traversing only the current candidate set. When the candidate set no longer changes, it is either a viable cluster, or contains only vertex i.
(v) To find further viable clusters, remove the viable cluster of i from the network (cutting any edges) and repeat steps (i)–(iv) on the remaining network beginning from a new test vertex.

Repeated application of this procedure will identify every viable cluster in the network. A simple example of the use of the algorithm to identify a small viable cluster is given in Fig. 2.3. The results of applying the algorithm to a graph containing two finite viable clusters is illustrated in Fig. 2.4.
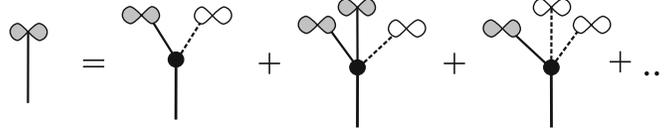
Fig. 2.5 Diagrammatic representation of Eq. (2.1) in a system of two interdependent networks *a* and *b*. The probability X_a, represented by a *shaded infinity symbol* can be written recursively as a sum of second-neighbor probabilities. *Open infinity symbols* represent the equivalent probability X_b for network *b*, which obeys a similar recursive equation. The filled *circle* represents the probability *p* that the vertex remains in the network

2.3 Hybrid Transition in Multiplex Networks

In this section we will study collapse of giant viable cluster in multiplex networks damaged by random removal of vertices. We will use the fraction *p* of vertices remaining undamaged as a control variable, however other control variables such as mean degree could also be used. As we will show below, in uncorrelated random networks the giant viable cluster collapses at a critical undamaged fraction p_c in a discontinuous hybrid transition, similar to that seen in the *k*-core or bootstrap percolation [14, 15].

Hybrid transitions, like those which occurs in the collapse of multiplex and interdependent networks, and associated avalanches, also occur in a wide variety of other systems. For example, a jump in activity in neural networks [16], population collapse in biological systems [17, 18], jamming and rigidity transitions and glassy dynamics [19, 20], and magnetic systems [21].

Let us construct the basic equations which allow us to analyse the hybrid transition. Consider the case of sparse uncorrelated networks, which are locally tree-like in the infinite size limit $N \rightarrow \infty$. In order to find the giant viable cluster, we take advantage of the locally tree-like property of the network, and define X_s, with the index $s \in \{a, b, \ldots\}$, to be the probability that, on following an arbitrarily chosen edge of type *s*, we encounter the root of an infinite sub-tree formed solely from type *s* edges, whose vertices are also each connected to at least one infinite subtree of every other type. We call this a type *s* infinite subtree. This is illustrated in Fig. 2.5, which shows the probability X_a as the sum of second-level probabilities in terms of X_a and X_b. The vector $\{X_a, X_b, \ldots\}$ plays the role of the order parameter. Writing this graphical representation in equation form, using the joint degree distribution $P(q_a, q_b, \ldots)$, we arrive at the self consistency equations

$$X_s = p \sum_{q_a, q_b, \ldots} \frac{q_s}{\langle q_s \rangle} P(q_a, q_b, \ldots) \big[1 - (1 - X_s)^{q_s - 1}\big] \prod_{l \neq s} \big[1 - (1 - X_l)^{q_l}\big]$$

$$\equiv \Psi_s(X_a, X_b, \ldots). \tag{2.1}$$

The multiplier *p* in Eq. (2.1) is the probability that the vertex remains in the network. The term $(q_s/\langle q_s \rangle) P(q_a, q_b, \ldots)$ gives the probability that on following an arbitrary
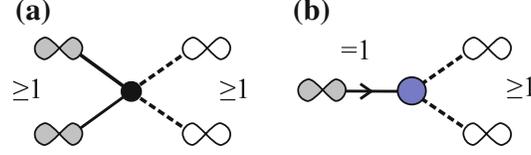
Fig. 2.6 Viable and critical viable vertices for two interdependent networks. **a** A vertex is in the giant viable cluster if it has connections of both kinds to giant viable subtrees, represented by *infinity symbols*, which occur with probabilities X_a (*shaded*) or X_b (*open*)—see text. **b** A critical viable vertex of type a has exactly one connection to a giant sub-tree of type a

edge of type s, we find a vertex with degrees $q_a, q_b, \ldots$, while $[1 - (1 - X_a)^{q_a}]$ is the probability that this vertex has at least one edge of type $a \neq s$ leading to the root of an infinite sub-tree of type a edges. This becomes $[1 - (1 - X_s)^{q_s - 1}]$ when $a = s$. The argument leading to Eq. (2.1) is similar to that used in [12]. Later it will be useful to write the right-hand side of this equation as $\Psi_s(X_a, X_b, \ldots)$.

A vertex is then in the giant viable cluster if it has at least one edge of every type s leading to an infinite type s sub-tree (probability X_s), as shown in Fig. 2.6a.

$$S = p \sum_{q_a, q_b, \ldots} P(q_a, q_b, \ldots) \prod_{s=a, b, \ldots} \left[1 - (1 - X_s)^{q_s} \right], \tag{2.2}$$

which is equal to the relative size of the giant viable cluster of the damaged network.

A hybrid transition appears at the point where $\Psi_s(X_a, X_b, \ldots)$ first meets X_s at a non-zero value, for all s. This occurs when

$$\det[\mathbf{J} - \mathbf{I}] = 0 \tag{2.3}$$

where $\mathbf{I}$ is the unit matrix and $\mathbf{J}$ is the Jacobian matrix $J_{ab} = \partial \Psi_b / \partial X_a$. The critical point p_c is found by solving Eqs. (2.1) and (2.3) together. To find the scaling near the critical point, we expand Eq. (2.1) about the critical value $X_s^{(c)}$. We find that

$$X_s - X_s^{(c)} \propto (p - p_c)^{1/2}. \tag{2.4}$$

This square-root scaling is the typical behaviour of the order parameter near a hybrid transition. In the next section we will show that this results from avalanches which diverge in size near the transition. The scaling of the size of the giant viable cluster, S, immediately follows

$$S - S_c \propto (p - p_c)^{1/2}. \tag{2.5}$$

A similar result is found for other control parameters, for example, mean degrees of the vertices.
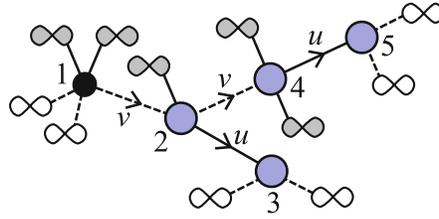
Fig. 2.7 A critical cluster. Removal of any of the shown viable vertices will result in the removal of all downstream critical viable vertices. Vertices 2–5 are critical vertices. Removal of the vertex labeled 1 will result in all of the shown vertices being removed (becoming non-viable). Removal of vertex 2 results in the removal of vertices 3, 4, and 5 as well, while removal of vertex 4 results only in vertex 5 also being removed. As before, *infinity symbols* represent connections to infinite viable subtrees. Other connections to non-viable vertices or finite viable clusters are not shown

2.4 Structure of Avalanches

Having established the behaviour of the order parameter, X_s, and the location of the hybrid transition, we now turn to examining avalanches, in order to understand the nature of the transition more completely. We focus on the case of two types of edges. Consider a viable vertex that has exactly one edge of type a leading to a type a infinite subtree, and at least one edge of type b leading to a type b infinite subtree. We call this a critical vertex of type a. It is illustrated in Fig. 2.6b. Critical vertices of type a will drop out of the viable cluster if they lose their single link to a type a infinite subtree. A vertex may have outgoing edges of this kind, so that removal of this vertex from the giant viable cluster also requires the removal of the critical vertices which depend on it. This is the way that damage propagates in the system. The removal of a single vertex can result in an avalanche of removals of critical vertices from the giant viable cluster. To represent this process visually, we draw a diagram of viable vertices and the edges between them. We mark the special critical edges, that critical viable vertices depend on, with an arrow leading to the critical vertex. An avalanche can only transmit in the direction of the arrows. For example, in Fig. 2.7, removal of the vertex labeled 1 removes the essential edge of the critical vertex 2 which thus becomes non-viable. Removal of vertex 2 causes the removal of further critical vertices 3 and 4, and the removal of 4 then requires the removal of 5. Thus critical vertices form critical clusters. At the head of each critical cluster is a 'keystone vertex' (e.g. vertex 1 in the figure) whose removal would result in the removal of the entire cluster. Graphically, upon removal of a vertex, we remove all vertices found by following the arrowed edges, which constitutes an avalanche. Note that an avalanche is a branching process. Removing a vertex may lead to avalanches along several edges emanating from the vertex (for example, in Fig. 2.7, removing vertex 2 leads to avalanches along two edges).As we approach the critical point from above, the avalanches increase in size. The mean size of avalanches triggered by a randomly removed vertex finally diverges in size at the critical point, which is the cause of the discontinuity in the size of the giant viable cluster, which collapses to zero. These avalanches are thus an inherent part of a hybrid transition.

Fig. 2.8 Symbols used in the diagrams to represent key probabilities. *Solid lines* represent edges of type a, *dashed lines* represent edges of type b

$$1 - X_a - R_a \qquad R_a \qquad X_a$$

$$1 - X_b - R_b \qquad R_b \qquad X_b$$
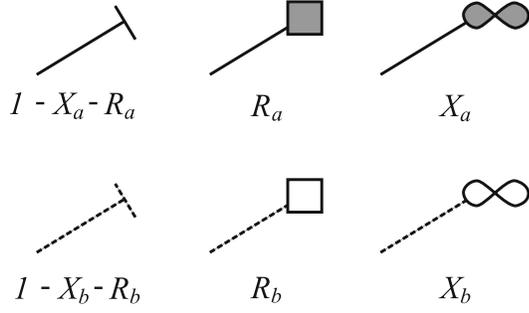
Fig. 2.9 a The probability R_a can be defined in terms of the second-level connections of the vertex found upon following an edge of type a. Note that possible connections to 'dead ends'—vertices not in the viable cluster (probability $1-X_a-R_a$ or $1-X_b-R_b$) are not shown. **b** The equivalent graphical equation for the probability R_b

(a)

(b)
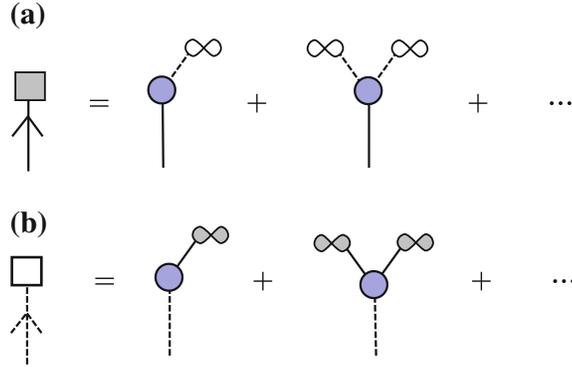
We can use a generating function approach, similar to that developed by Newman [4] to calculate the sizes and structure of avalanches. There are three possibilities when following an arbitrarily chosen edge of a given type: (i) with probability X_s we encounter a type s infinite subtree (ii) with probability R_s we encounter a vertex which has a connection to an infinite subtree of the opposite type, but none of the same type. Such a vertex is part of the giant viable cluster if the parent vertex was; or (iii) with probability $1 - X_s - R_s$, we encounter a vertex which has no connections to infinite subtrees of either kind. These probabilities are represented graphically in Fig. 2.8. We will use these symbols in subsequent diagrams.

The probability R_a obeys

$$R_a = \sum_{q_a} \sum_{q_b} \frac{q_a}{\langle q_a \rangle} P(q_a, q_b)(1-X_a)^{q_a-1} \left[1-(1-X_b)^{q_b}\right] \tag{2.6}$$

and similarly for R_b. This equation is represented graphically in Fig. 2.9.

The generating function for the size of an avalanche triggered by removing an arbitrary type a edge which does not lead to an infinite type a subtree can be found by considering the terms represented in Fig. 2.10. The first term represents the probability, upon following an edge of type a (solid lines) of reaching a "dead end", that is, a vertex with no connection to a type b subtree (and hence is not a viable vertex). In other words, a critical cluster of size 0. The second term represents a critical cluster of size 1: the vertex encountered has a connection to the type b infinite subtree (infinity
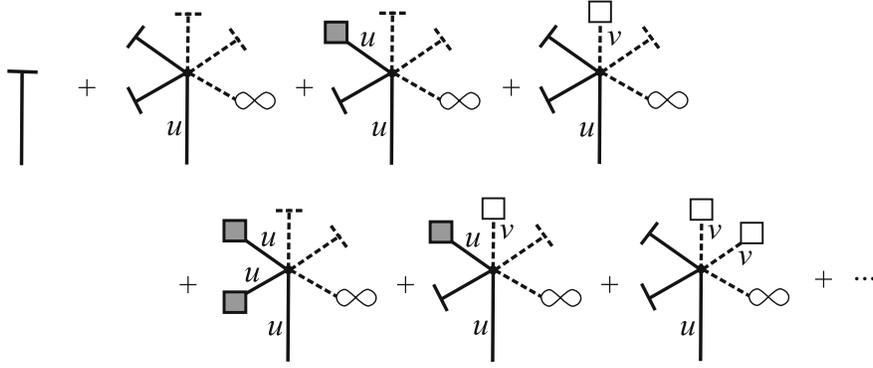
Fig. 2.10 Representation of the generating function $H_a(x, y)$ (*right-hand side* of Eq. 2.8) for the size of a critical cluster encountered upon following an edge of type a

symbol), but no further connections to viable vertices. Subsequent terms represent recursive probabilities that the vertex encountered has 1 (third and fourth terms), 2 (fifth, sixth, seventh terms) or more connections to further potential critical clusters. The variable u (for type a edges) or v (type b) are assigned to each such edge. The equation for this generating function can be written in terms of functions $F_a(x, y)$ and $F_b(x, y)$ which we define as follows:

$$F_a(x, y) \equiv \sum_{q_a} \sum_{q_b} \frac{q_a}{\langle q_a \rangle} P(q_a, q_b) x^{q_a-1} \sum_{r=1}^{q_b} \binom{q_b}{r} X_b^r y^{q_b-r} \tag{2.7}$$

and similarly for $F_b(x, y)$, by exchanging all subscripts a and b. While the function $F_a(x, y)$ does not necessarily represent a physical quantity or probability, we can see that it incorporates the probability of encountering a vertex with at least one child edge of type b leading to a giant viable subtree (probability X_b) upon following an edge of type a. All other outgoing edges then contribute a factor x (for type a edges) or y (type b).

In terms of these functions, we can write the generating function for the number of critical vertices encountered upon following an arbitrary edge of type a (that is, the size of the resulting avalanche if this edge is removed) as

$$H_a(u, v) = 1 - X_a - R_a + u F_a[H_a(u, v), H_b(u, v)] \tag{2.8}$$

and similarly for $H_b(u, v)$, the corresponding generating function for the size of the avalanche caused by removing a type b edge:

$$H_b(u, v) = 1 - X_b - R_b + v F_b[H_a(u, v), H_b(u, v)]. \tag{2.9}$$

These recursive equations can be understood by noting that $H_a(0, v)$ $= 1 - X_a - R_a$ is the probability that an arbitrarily chosen edge leads to a vertex outside the viable cluster. Here u and v are auxiliary variables. Following through a

critical cluster, a factor u appears for each arrowed edge of type a, and v for each arrowed edge of type b. For example, the critical cluster illustrated in Fig. 2.7 contributes a factor $u^2 v^2$. The mean number of critical vertices reached upon following an edge of type a, i.e. the mean size of the resulting avalanche if this edge is removed, is given by $\partial_u H_a(1, 1) + \partial_v H_a(1, 1)$, where ∂_u signifies the partial derivative with respect to u.

Unbounded avalanches emerge at the point where $\partial_u H_a(1, 1)$ [or $\partial_v H_b(1, 1)$] diverges. Taking derivatives of Eq. (2.8),

$$\partial_u H_a(u, v) = F_a[H_a, H_b] + u \left\{ \partial_u H_a \partial_x F_a[H_a, H_b] + \partial_u H_b \partial_y F_a[H_a, H_b] \right\} \tag{2.10}$$

$$\partial_v H_a(u, v) = u \left\{ \partial_v H_a \partial_x F_a[H_a, H_b] + \partial_v H_b \partial_y F_a[H_a, H_b] \right\} \tag{2.11}$$

with similar equations for $\partial_u H_b(u, v)$ and $\partial_v H_b(u, v)$. Some rearranging gives

$$\partial_u H_a(1, 1) = \frac{R_a + \partial_u H_b(1, 1)\partial_y F_a(1 - X_a, 1 - X_b)}{1 - \partial_x F_a(1 - X_a, 1 - X_b)} \tag{2.12}$$

and

$$\partial_v H_a(1, 1) = \frac{\partial_u H_a(1, 1)\partial_x F_b(1 - X_a, 1 - X_b)}{1 - \partial_y F_b(1 - X_a, 1 - X_b)} \tag{2.13}$$

where we have used that $H_a(1, 1) = 1 - X_a$ and $F_a(1 - X_a, 1 - X_b) = R_a$.

From Eqs. (2.1) and (2.7),

$$\partial_x F_a(1 - X_a, 1 - X_b) = \frac{\partial}{\partial X_a} \Psi_a(X_a, X_b) \tag{2.14}$$

$$\partial_y F_1(1 - X_a, 1 - X_b) = \frac{\langle q_a \rangle}{\langle q_b \rangle} \frac{\partial}{\partial X_a} \Psi_b(X_a, X_b), \tag{2.15}$$

and similarly for $\partial_x F_{1b}$ and $\partial_y F_{1b}$, which when substituted into (2.12) and (2.13) gives

$$\partial_u H_a(1, 1) = \frac{R_a[1 - \frac{\partial}{\partial X_b} \Psi_b(X_a, X_b)]}{\det[\mathbf{J} - \mathbf{I}]} . \tag{2.16}$$

We see that the denominator exactly matches the left-hand side of Eq. (2.3), meaning that the mean size of avalanches triggered by random removal of vertices diverges exactly at the point of the hybrid transition.

The mean size of the avalanche triggered by the removal of a randomly chosen vertex can be related to the susceptibility of the giant viable cluster to random damage, similar to the susceptibility for ordinary percolation. In the latter case, the susceptibility is defined as the mean size of the cluster to which a randomly chosen vertex belongs [22]. Due to the similarity of Eq. (2.4) to the k-core version [23], we can

expect that, at the critical point $p = p_c$, the size distribution of avalanches triggered by randomly removed vertices obeys a power law $p(s) \propto s^{-\sigma}$ with $\sigma = 3/2$.

2.5 Avalanches in Scale-Free Networks

In ordinary and k-core percolation, networks with degree distributions that are asymptotically power laws $P(q) \sim q^{-\gamma}$ may exhibit qualitatively different transitions from those described above, especially when $\gamma < 3$. To investigate such effects in the giant viable cluster, we consider two uncorrelated scale-free networks, so $P(q_a, q_b) = P_a(q_a)P_b(q_b)$, having powerlaw degree distributions with fixed minimum degree $q_0 = 1$ (then $\langle q \rangle \approx (\gamma - 1)q_0/(\gamma - 2)$), so that

$$P_s(q_s) = \zeta(\gamma_s)q^{-\gamma_s} \tag{2.17}$$

where s takes the values a or b, and $\zeta(\gamma)$ is the Riemann zeta function. As before, we apply random damage to the system as a whole as a control parameter, so that vertices survive with probability p.

First consider the case that at least one of the degree distribution exponents γ_a and γ_b is greater than three. The giant viable cluster is necessarily a subgraph of the overlap between the giant-components of each graph. We know from ordinary percolation that for $\gamma > 3$, the giant component appears at a finite value of p [24]. It follows that the giant viable cluster, also, cannot appear from $p = 0$; there must be a finite threshold p_c, (with a hybrid transition). This is true even if one of the networks has $\gamma_s < 3$.

The more interesting case is when $\gamma_a, \gamma_b < 3$, when the percolation threshold is zero for each network when considered separately. Let us write $\gamma_a = 2 + \delta_a$ and $\gamma_b = 2 + \delta_b$, and examine the behavior for small δ_a and δ_b. We proceed by assuming that in this situation, for p near p_c, Eq. (2.1) have a solution with small $X_a, X_b \ll 1$. Writing only leading orders of X_a and X_b, and δ_a and δ_b, we find that

$$\Psi_a(X_a, X_b) = p\frac{\pi^2}{6\,\delta_b}X_a^{\delta_a}\left(X_b - X_b^{1+\delta_b}\right) \tag{2.18}$$

and similarly for $\Psi_b(X_a, X_b)$. The location of the critical point is found from Eq. (2.3) which becomes

$$\delta_a + \delta_b = p\frac{\pi^2}{6}X_a^{\delta_a}X_b^{\delta_b}\left(\frac{X_a}{X_b} + \frac{X_b}{X_a}\right). \tag{2.19}$$

Substituting Eq. (2.18) into (2.1) and solving with Eq. (2.19), we find X_s and S at p_c. We find in general that the hybrid transition persists for $\delta_a, \delta_b \neq 0$, that is $p_c > 0$, but that the height of the discontinuity $X_s^{(c)}$ at the hybrid transition becomes extremely

small for small δ small. In experiments or simulations, this could be misinterpreted as evidence of a continuous phase transition.

To illustrate the results in this case, we describe two representative examples. First, we fix δ_b at some small value, and examine the limit $\delta_a \to 0$, so that $\delta_a \ll \delta_b$. That is, $\gamma_a \to 2$ while $\gamma_b > 2$. We find that the location, p_c, of the transition tends to a finite value as $\delta_a \to 0$, proportional to the larger δ_b,

$$p_c = \sqrt{\frac{1-2f}{f(1-f)^2}\frac{\delta_b}{\zeta(2)}} \approx 1.19\delta_b, \tag{2.20}$$

where $f \approx 0.236$. The values of X_a and X_b become very small at the critical point, $X_b = f^{1/\delta_b}$ and $X_a \approx 1.5X_b$, meaning the size of the giant viable cluster at the critical point is exponentially small

$$S_c = \left(\frac{1-2f}{f}\right)^{3/2} f^{2/\delta_b} = Ae^{-B/\delta_b} \tag{2.21}$$

where $A \approx 3.36$ and $B \approx 2.89$. We see that a hybrid transition occurs, albeit with an extremely small discontinuity, at a non-zero threshold p_c as long as at least one of δ_a and δ_b is not equal to zero.

To examine the case that both δ_a and δ_b tend to zero, we consider the symmetric case $\delta_a = \delta_b \equiv \delta$. Then $X_a = X_b \equiv X$.

Equation (2.1) become a single equation,

$$\Psi(X) \approx p\frac{\zeta(2)}{\delta}\left(X^{1+\delta} - X^{1+2\delta}\right). \tag{2.22}$$

The discontinuity is found by requiring $\Psi'(X) = 1$ [from Eq. (2.3)] which condition becomes

$$\Psi'(X) \approx p\,\zeta(2)\left[(1+\delta)X^\delta - (1+2\delta)X^{2\delta}\right] = 1. \tag{2.23}$$

Solving these two equations, we find that $X_c = (1/2)^{1/\delta}$ and

$$p_c = \frac{24}{\pi^2}\delta \tag{2.24}$$

$$S_c = 4\left(\frac{1}{2}\right)^{2/\delta}. \tag{2.25}$$

The location of the hybrid transition tends to $p = 0$ as $\delta \to 0$, and the size of the 'jump' becomes very small even for nonzero δ, but vanishes completely as $\delta \to 0$. In Fig. 2.11 we plot the size of the giant viable cluster in this symmetric case for three values of γ. For values not close to two, the transition looks similar to that observed in, say, Erdős–Rényi graphs. As γ approaches 2, however, we see that the height
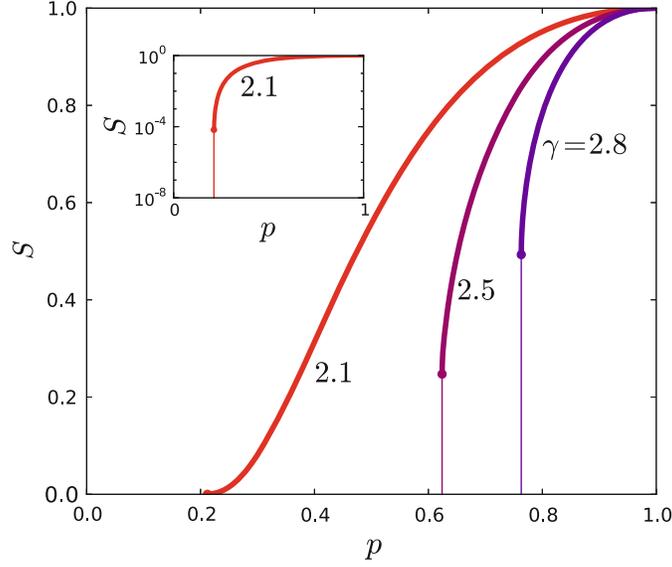
Fig. 2.11 Size of the giant viable cluster S as a function of the fraction p of vertices remaining undamaged for two symmetric powerlaw distributed networks with, from *right* to *left*, $\gamma = 2.8$, 2.5, and 2.1. The height of the jump becomes very small as γ approaches 2, but is not zero, as seen in the *inset*, which is S versus p on a logarithmic vertical scale for $\gamma = 2.1$

of the discontinuity becomes extremely small. Nevertheless, the square-root scaling and non-zero critical point are retained.

We can also examine the behaviour of X and S above the transition ($p > p_c$). Expanding $\Psi(X)$ about X_c we find that

$$\frac{X - X_c}{X_c} = \frac{12}{\pi^2 \delta \, p_c} \left(\frac{p - p_c}{p_c} \right)^{1/2} \tag{2.26}$$

which holds so long as $p - p_c \ll \delta^3$. That is, the scaling of the order parameter X, and hence the size of the giant viable cluster, S, is square-root in a narrow region of width $\mathcal{O}(\delta^3)$ above the hybrid transition. This region disappears as $\delta \to 0$.

2.6 Conclusions

In conclusion, we have studied the robustness of multiplex networks, which are networks with two or more different kinds of edges. There is a direct mapping between such multiplex networks and interdependent networks, in which vertices in one network depend on at most one vertex in another network. We found that the giant viable cluster of a multiplex network with two or more kinds of edges collapses with a discontinuous hybrid transition. The collapse occurs through avalanches which diverge in size when the transition is approached from above. We described critical

clusters associated with these avalanches. The avalanches are responsible for both the critical scaling and the discontinuity observed in the size of the giant viable cluster. Remarkably, these specific clusters and avalanches in our problem turned out to be organized in a novel way, different from those in the k-core [15, 23] and bootstrap percolation [14] problems.

In contrast to ordinary networks, where two vertices are connected if there is a path between them, in multiplex network with m types of edges, two vertices are m-connected if for every kind of edge there is a path from one to another vertex. Based on this notion, we introduced *viable clusters* as clusters of m-connected vertices in multiplex network. This new notion of connectivity between vertices leads to the emergence in a multiplex network of a giant viable cluster in a hybrid phase transition in contrast to a continuous phase transition in ordinary percolation.

Surprisingly, when the degree distributions are asymptotically power-law $P(q) \propto q^{-\gamma}$ the critical point p_c (taking the undamaged fraction of vertices p as the control parameter) remains at a finite value even when the exponents γ of the degree distributions are below three, remaining finite until both exponents reach two, in agreement with an argument given in [10]. This is in stark contrast to ordinary percolation in complex networks, in which the threshold falls to zero as soon as γ reaches three [25, 26]. We show, further, that the nature of the transition does not change. Although the height of the discontinuity becomes extremely small near $\gamma = 2$, it remains finite near this limit (see Fig. 2.11). The critical clusters may have important practical applications, helping to identify vulnerabilities to targeted attack, as well as informing efforts to guard against such attack.

References

1. Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.: Critical phenomena in complex networks. Rev. Mod. Phys. **80**, 1275–1335 (2008)
2. Dorogovtsev, S.N., Mendes, J., Samukhin, A., Zyuzin, A.: Organization of modular networks. Phys. Rev. E **78**, 056106 (2008)
3. Leicht, E.A., D'Souza, R.M.: Percolation on interacting networks (2009). arXiv:0908.0894
4. Newman, M.E.J., Strogatz, S.H., Watts, D.J.: Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E **64**, 026118 (2001)
5. Pocock, M.J.O., Evans, D.M., Memmott, J.: The robustness and restoration of a network of ecological networks. Science **335**, 973–976 (2012)
6. Kurant, M., Thiran, P.: Layered complex networks. Phys. Rev. Lett. **96**, 138701 (2006)
7. Rinaldi, S.M., Peerenboom, J.P., Kelly, T.K.: Identifying, understanding, and analyzing critical infrastructure interdependencies. IEEE Control Syst. Mag. **21**, 11–25 (2001)
8. Dueñas, L., Cragin, J.I., Goodno, B.J.: Seismic response of critical interdependent networks. Earthq. Eng. Struct. Dyn. **36**, 285–306 (2007)
9. Poljanšek, K., Bono, F., Gutiérrez, E.: Seismic risk assessment of interdependent critical infrastructure systems: The case of european gas and electricity networks. Earthq. Eng. Struct. Dyn. **41**, 61–79 (2012)

10. Buldyrev, S.V., Parshani, R., Paul, G., Stanley, H.E., Havlin, S.:Catastrophic casdade of failures in interdependent networks. Nature **464**, 08932 (2010)
11. Gao, J., Buldyrev, S.V., Havlin, S., Stanley, H.E.: Robustness of a network of networks. Phys. Rev. Lett. **107**, 195701 (2011)
12. Son, S.W., Bizhani, G., Christensen, C., Grassberger, P., Paczuski, M.: Percolation theory on interdependent networks based on epidemic spreading. EPL **97**, 16006 (2012)
13. Baxter, G.J., Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.:Avalanche collapse of interdependent networks.Phys. Rev. Lett. **109**, 248701 (2012)
14. Baxter, G.J., Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.:Bootstrap percolation on complex networks. Phys. Rev. E **82**, 011103 (2010)
15. Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.: k-core organisation of complex networks. Phys. Rev. Lett. **96**, 040601 (2006)
16. Eckmann, J.P., Feinerman, O., Gruendlinger, L., Moses, E., Soriano, J., Tlusty, T.: The physics of living neural networks. Phys. Rep. **449**, 54–76 (2007)
17. Dai, L., Vorselen, D., Korolev, K.S., Gore, J.: Generic indicators for loss of resilience before a tipping point leading to population collapse. Science **336**, 1175–7 (2012)
18. Klimek, P., Thurner, S., Hanel, R.: Pruning the tree of life: k-core percolation as selection mechanism. J. Theoret. Biol. **256**, 142–146 (2009)
19. Sellitto, M., Biroli, G., Toninell, C.: Facilitated spin models on bethe lattice: Bootstrap percolation, mode-coupling transition and glassy dynamics. Europhys. Lett. **69**(4), 496–502 (2005)
20. Toninelli, C., Biroli, G., Fisher, D.S.: Jamming percolation and glass transitions in lattice models. Phys. Rev. Lett. **96**(3), 035702 (2006)
21. Sabhapandit, S., Dhar, D., Shukla, P.: Hysteresis in the random-field Ising model and bootstrap percolation. Phys. Rev. Lett. **88**, 197202 (2002)
22. Stauffer, D., Aharony, A.: Introduction to percolation theory, 2nd edn. Taylor and Francis, London (1992)
23. Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.: k-core architecture and k-core percolation on complex networks. Physica D **224**, 7–19 (2006)
24. Cohen, R., ben Avraham, D., Havlin, S.: Percolation criticalexponents in scale-free networks. Phys. Rev. E **66**, 036113 (2002)
25. Albert, R., Jeong, H., Barabási, A.L.: Error and attacktolerance in complex networks. Nature **401**,378–382 (2000)
26. Callaway, D.S., Newman, M.E.J., Strogatz, S.H., Watts, D.J.: Networkrobustness and fragility: Percolation on random graphs. Phys. Rev. Lett. **85**, 5468–5471 (2000)

Chapter 3
Multiplex Networks

Kyu-Min Lee, Jung Yeol Kim, Sangchul Lee and K.-I. Goh

Abstract Typical complex system operates through multiple types of interactions between its constituents. The collective function of these multiple interactions, or multiple network layers, is often non-additive, resulting in nontrivial effects on the network structure and dynamics. To better model such situations, the concept of multiplex network, the network with explicit multiple types of links, has recently been applied. In this contribution, we survey recent studies on this subject, focused on the notion of correlated multiplexity. Empirical multiplex network analysis as well as analytical results on the random graph models of correlated multiplex networks are presented, followed by a brief summary of dynamical processes on multiplex networks. It is illustrated that a multiplex complex system can indeed exhibit structural and dynamical properties that cannot be represented by its individual layer's properties alone, establishing the network multiplexity as an essential ingredient in the new physics of "network of networks."

3.1 Introduction

In the last decade, network science has successfully established itself as a unified framework for studying complex systems [1, 2]. Along with its impressive success, the framework has continuously been evolving. One of the most current evolution of complex network theory is the study of multiplex networks, the networks with more than one type of links [3]. Indeed, most studies until quite recently have focused on isolated, single networks, ignoring the existence of multiple types of interactions. In most, if not all, real-world complex systems, however, nodes in the system can engage in more than one type of interactions, and such multiple interactions can make a non-additive effect on network structure and the dynamics on it. For example, as illustrated

K.-M. Lee · J. Y. Kim · S. Lee · K.-I. Goh (✉)
Department of Physics, Korea University, Seoul 136-713, Korea
e-mail: kgoh@korea.ac.kr

Fig. 3.1 A cartoon of multiplex social network as a triplex network consisting of friendship, family, and work-related acquaintanceship layers
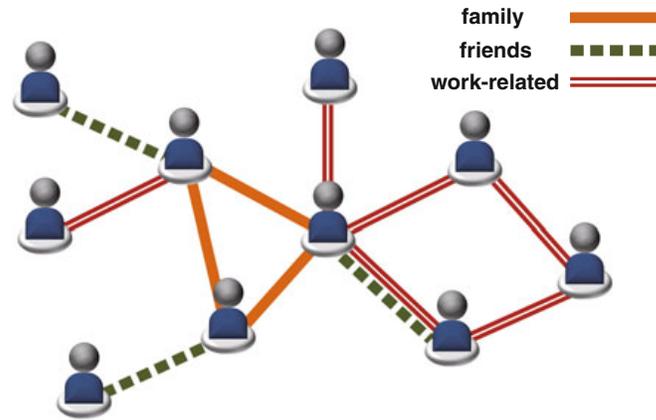
in Fig. 3.1, people in a society interact via their friendship, family relationship, and/or more formal work-related acquaintanceship, *etc.*, which are collectively responsible for complex emergent social phenomena [4, 5]. Countries in the global economic system also interact via various international relations ranging from commodity trade to political alliance [6]. Even proteins in a cell participate in multiple layers of interactions and regulations, from transcriptional regulations and metabolic synthesis to signaling [7]. Obviously, in dealing with such problems the multiplex network representation would be a more appropriate description than the single network, or simplex, one.

In this contribution, we will survey recent works on the topic of multiplex networks. We begin with an analysis of real-world multiplex coauthorship network data to introduce the notion of correlated multiplexity in Sect. 3.2. Then the random graph model of correlated multiplex network is introduced in Sect. 3.3. In Sect. 3.4, analytical formalism based on the joint degree distribution for analyzing the structural properties of multiplex random graph models is developed. The cases of duplex random graphs and duplex scale-free networks are studied in detail in Sects. 3.5 and 3.6, respectively. Topics of network robustness and network dynamics are briefly discussed in Sects. 3.7 and 3.8, respectively. Finally, we will conclude our contribution with a summary and outlook.

3.2 Correlated Multiplexity

In most previous studies of coupled networks—in context of layered, interacting, interdependent networks [8–10]—network layers were coupled randomly. In real-world complex systems, however, nonrandom structure in network multiplexity can be prominent. For example, a person with many links in the friendship layer is likely to also have many links in another social network layer, being a friendly person. We termed the correlated multiplexity [3] to refer such a nonrandom pattern of network

multiplexity. Examples of correlated multiplexity are widespread. Some of examples reported in the literature are:

- Social networks: online-game network [11], coauthorship network [12].
- Organizational networks [13].
- Transportation networks [8, 14, 15].
- Cellular network: Interaction network and perturbation network [16].
- Economic networks: Trade networks in different industrial sectors [17].

The most frequent pattern of correlated multiplexity is the positively correlated multiplexity, such that a node with large degree in one layer likely has more links in the other layer as well. For example, in the online game social network data [11], it was shown that different positive social relations such as friendship and trade are highly correlated as well as overlap.

In Fig. 3.2, we present our own analysis of a multiplex coauthorship network [12]. The network consists of a set of researchers who are connected with one another by three types of collaboration links, first being due to publications in the field of fractal surface growth (denoted KPZ, representing Kardar-Parisi-Zhang equation), second in the field of self-organized criticality (denoted SOC, representing Self-Organized Criticality), and third in the field of complex network theory (denoted CNR, representing Complex Network Research), resulting in a triplex network (for more details on the data collection, see [12]). Despite the separation of timescales of three research topics, degree distributions of the three network layers, and that of the superposed network, are indistinguishable (Fig. 3.2a, inset). Within the individual layer, analysis of degree distributions of restricted set of nodes that participate in more than one layers reveals that there indeed exists a positively correlated multiplexity pattern: the more layer a node participates to, the more likely would they have larger degrees (Fig. 3.2a). The analysis of joint degree distributions (Fig. 3.2c,d) confirms this finding. There is a systematic enrichment of joint degree distribution near the diagonal of the plots, revealing strong correlation between degrees of a node in two network layers. In addition, it was found that a pair of nodes which are closer in one layer tend to be also closer in another layer (Fig. 3.2b). This result extends the classical concept of multiplexity that accounts only for direct link overlap [4] and demonstrates the effect of network multiplexity at all scales.

3.3 Random Graph Model of Correlated Multiplexity

For a systematic mathematical understanding of correlated multiplexity, one needs a graph model. There exist a few random graph models with multiple link-types (or colored edges) [3, 18, 19]. Here we present a way to build correlated multiplex networks, following [3].

Given two network layers with equal number of nodes, we define three particular couplings: (i) uncorrelated, (ii) maximally-positive (MP), and (iii) maximally-negative (MN) correlated couplings (Fig. 3.3). In the uncorrelated coupling, we
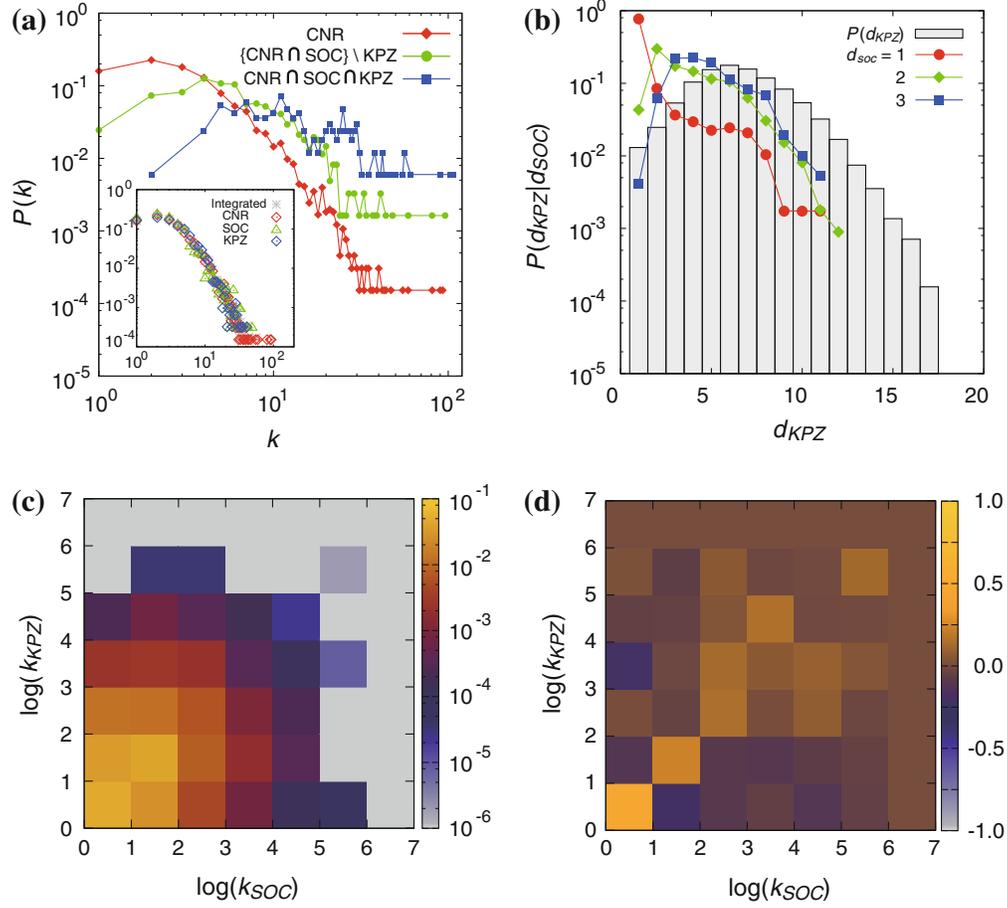
Fig. 3.2 Patterns of correlated multiplexity in multiplex coauthorship network. **a** Degree distribution of nodes participating in a single (*diamond*), double (*circle*), and triple layers (*square*). **b** Conditional distance distribution $P(d_{KPZ}|d_{SOC})$ in KPZ-layer of pairs of nodes of distance d_{SOC} in SOC-layer. **c** Joint degree distribution $P(k_{SOC}, k_{KPZ})$, and **d** Significance plot based on Z-score with respect to randomly coupled counterpart. Z-score is obtained as $Z = (P_{real} - \langle P_{random} \rangle)/\sigma_{P_{random}}$, where the average and standard deviation for P_{random} are evaluated over 10^4 independent randomizations

couple the two layers randomly, that is, we use a random matching between a node in one layer to a node in the other layer. In the MP correlated coupling, a node's degrees in different layers are maximally correlated in their degree order; the node that is hub in one layer is also the hub in the other layer, and the node that has the smallest degree in one layer also has the smallest degree in other layer. Likewise, in the MN correlated coupling, a node's degrees in different layers are maximally anti-correlated in their degree order.

These three particular couplings are useful in their mathematical simplicity and tractability, thus highlighting the effect of correlated multiplexity. Yet in real-world multiplex systems the correlated multiplexity would hardly be maximal. The cases of partially correlated multiplexity can be constructed by maximally correlating a
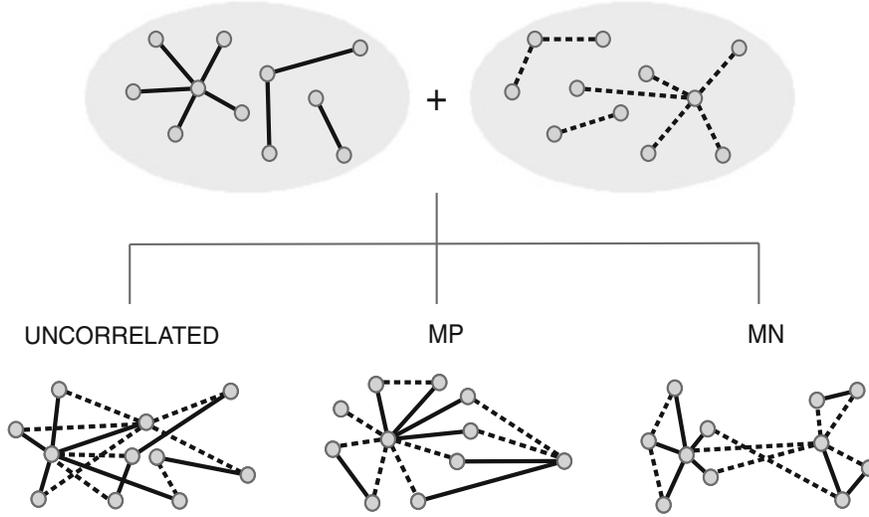
Fig. 3.3 Schematic illustration of constructing the correlated multiplex networks discussed in the text. MP (MN) stands for maximally-positive (maximally-negative) correlated multiplexity

fraction q of nodes in the network while randomly coupling the rest fraction $1 - q$. Using this method one can interpolate between MP, through uncorrelated, and MN couplings, modulating the strength of correlated multiplexity.

3.4 Analytical Formalisms

3.4.1 Degree Distributions

The information of degree distribution of a multiplex network with ℓ layers (ℓ-plex network) can be encoded in the joint degree distribution $P(\{k_\alpha\}) \equiv P(k_1, k_2, \cdots, k_\ell)$. (Throughout this work, we will use Greek subscript to denote the layer index). The degree distribution within a layer α, denoted as $\pi_\alpha(k_\alpha)$, can be obtained as the marginal distribution, $\pi_\alpha(k_\alpha) = \sum_{\{k_{\beta \neq \alpha}\}} P(k_1, k_2, \cdots, k_\ell)$. The total degree of a node in the multiplex network is given by $k = \sum_\alpha k_\alpha$, which can differ from the number of distinct connected nodes when there are link overlaps between network layers. Such link overlaps can be neglected for large, sparse random graphs, but can be significant in real-world multiplex networks as in multiplex social network data [11, 12]. One can obtain the total degree distribution $P(k)$ from the joint degree distribution as $P(k) = \sum_{\{k_\mu\}} P(\{k_\mu\}) \delta_{k, \sum_\nu k_\nu}$, where δ denotes Kronecker delta symbol.

3.4.2 Emergence of the Giant Component

Having established a way to construct the total degree distribution $P(k)$, it is tempting to use it to calculate the connected components properties via standard generating function technique [20]. It turns out that, however, this simplified procedure works only when the degree distributions of all layers are identical, as we will see shortly.

Now we develop a theory which exploits the full joint degree distribution $P(\{k_\alpha\})$, applicable when every layer is uncorrelated and locally tree-like, as in random graph models. Let us define u_α to be probability that a node reached by a randomly chosen link in layer α does not belong to the giant component (which is connected via *any* types of links). Following a similar reasoning as the standard generating function technique, one can construct the self-consistency equations for u_α's as

$$u_\alpha = \sum_{\{k_\mu\}} \frac{k_\alpha P(\{k_\mu\}) \prod_\nu u_\nu^{k_\nu}}{z_\alpha} \frac{}{u_\alpha} \qquad (\alpha = 1, \cdots, \ell) , \qquad (3.1)$$

where z_α is the mean degree of layer α. Then the probability that a randomly chosen node belongs to the giant component (that is, the giant component size), denoted S, can be obtained as

$$S = 1 - \sum_{\{k_\mu\}} P(\{k_\mu\}) \prod_\nu u_\nu^{k_\nu} , \qquad (3.2)$$

with u_ν's being the solution of Eq. (3.1). Therefore, the giant component exists (that is, $S > 0$) if Eq. (3.1) has a nontrivial solution other than $(u_1, \cdots, u_\ell) = (1, \cdots, 1)$. This condition can be extracted from the Jacobian of Eq. (3.1), which reads in the case of duplex network

$$\frac{1}{4} \left[\left(\frac{\kappa_1}{z_1} + \frac{\kappa_2}{z_2} \right) + \sqrt{\left(\frac{\kappa_1}{z_1} - \frac{\kappa_2}{z_2} \right)^2 + \frac{4\kappa_{12}^2}{z_1 z_2}} \right] > 1 , \qquad (3.3)$$

where $\kappa_1 = \langle k_1^2 \rangle$, $\kappa_2 = \langle k_2^2 \rangle$, and $\kappa_{12} = \langle k_1 k_2 \rangle$ are second-order moments of joint degree distribution.

When the degree distributions of all layers are identical, one has the solution of Eq. (3.1) satisfying $u_1 = u_2 = \cdots = u_\ell$, which reduces Eqs. (3.1–3.3) to those of standard generating function technique [20]. For example, Eq. (3.3) reduces to the well-known Molloy-Reed criterion for the total degree distribution, $\langle k^2 \rangle - 2\langle k \rangle > 0$, with $k = k_1 + k_2$ [21]. This shows that in such a case, one can use the reduced total degree distribution $P(k)$ to study the component structure, but in general Eqs. (3.1–3.3) should be used to have the correct results. Note that similar generating function-type techniques for clustered [22], multi-type [23], and interdependent networks [24] have also been developed recently, which slightly differ from the current formalism.

3.4.3 Degree-Degree Correlations

The fact that one cannot use the reduced total degree distribution $P(k)$ for component structure of correlated multiplex network suggests that the superposed network possesses degree correlations even when uncorrelated random networks are coupled. To show this explicitly, let us consider the assortativity coefficient r defined as [25]

$$r = \frac{\langle kk' \rangle_l - \langle k \rangle_l^2}{\langle k^2 \rangle_l - \langle k \rangle_l^2}, \tag{3.4}$$

where k and k' are the *total* degrees of nodes at two ends of an edge and $\langle \cdots \rangle_l$ denotes the average over all edges in the superposed network. Nonzero value of r dictates the presence of degree-degree correlations between connected nodes. Following the steps developed in [22], one can show that the numerator of Eq. (3.4) can be expressed, after some manipulations, as

$$\sum_{k,k'} kk' Q(k,k') - \left(\sum_{k,k'} kQ(k,k') \right)^2 = \sum_{\mu} c_\mu X_\mu^2 - \left(\sum_{\mu} c_\mu X_\mu \right)^2$$

$$= \frac{1}{2} \sum_{\mu,\nu} c_\mu c_\nu (X_\mu - X_\nu)^2 \geq 0, \tag{3.5}$$

where $Q(k,k')$ denotes the probability that a randomly chosen link (of any kind) connects two nodes with total degree k and k' at each end, c_α is the fraction of links of type α, such that $\sum_\alpha c_\alpha = 1$, and X_α is the expected total degree of a node that is reached by following a randomly chosen link of type α, which is related to the joint degree distribution as

$$X_\alpha = \sum_k k \sum_{\{k_\mu\}} k_\alpha P(\{k_\mu\}) \delta(k - \sum_\nu k_\nu)/z_\alpha . \tag{3.6}$$

Therefore, a multiplex network can become assortative ($r > 0$), even when uncorrelated layers are coupled, unless the degree distributions of all layers are identical, so that all X_α's are equal. (Another exception is the uncorrelated multiplex ER graphs, see Sect. 3.5.1.) It also allows one to calculate the assortativity coefficient r, once the joint degree distribution is given.

3.5 Duplex ER Graphs

To illustrate basic effects of multiplex couplings, in this section we apply the formalism to duplex Erdős-Rényi (ER) graphs [26] in which two ER graph layers are multiplex coupled, summarizing the results reported in [3].

3.5.1 Uncorrelated Duplex ER Graphs

In the absence of correlation between network layers, the joint degree distribution factorizes, $P_{uncorr}(k_1, k_2) = \pi_1(k_1)\pi_2(k_2)$. The total degree distribution is then given by the convolution of $\pi_\alpha(k_\alpha)$, $P_{uncorr}(k) = \sum_{k_1=0}^{k} \pi_1(k_1)\pi_2(k - k_1)$. It is easy to see that the resulting superposed network is nothing but an ER graph with the total mean degree $z_1 + z_2$, so that

$$P_{uncorr}(k) = \frac{e^{-z}z^k}{k!} \tag{3.7}$$

with $z = z_1 + z_2$. The connectivity and component properties follow the conventional behaviors [20, 26].

3.5.2 Duplex ER Networks with Equal Link Densities

The case of duplex ER networks with layers of equal link densities is particularly simple, as one can use standard generating function technique with the total degree distribution. Furthermore it is amenable for a number of explicit exact results.

MP coupling.—In this case, degrees of a node in the two layers would become almost equal in the thermodynamic limit (more precisely, relative dispersion of the two degrees would decay with N and vanish as $N \to \infty$), so that the total degree distribution of the duplex network can be approximated as

$$P_{MP}(k) = \begin{cases} e^{-z_1}z_1^{k/2}/(k/2)! & (k \text{ even}), \\ 0 & (k \text{ odd}), \end{cases} \tag{3.8}$$

where z_1 is the mean degree of the layer 1. Therefore, the Molloy-Reed criterion is fulfilled for all nonzero z_1, as $\langle k^2 \rangle - 2\langle k \rangle = 4(z_1 + z_1^2) - 2(2z_1) = 4z_1^2 > 0$ for $z_1 \neq 0$, which can also follow from the condition Eq. (3.3). This means that surprisingly the giant component exists for any nonzero link density, that is, the critical single-layer mean degree z_c above which the giant component exists vanishes,

$$z_c^{MP} = 0 \ . \tag{3.9}$$

One can further obtain the giant component size S and the average size of finite components $\langle s \rangle$ from the standard generating function technique [20], which are given explicitly by:

$$S = 1 - P(0) = 1 - e^{-z_1} \ , \tag{3.10}$$

and

$$\langle s \rangle = 1 \ . \tag{3.11}$$

This shows that the giant component grows linearly in the vicinity of z_c^{MP}, and that only the isolated nodes are outside the giant component and all the linked nodes form a single giant component. All these predictions are fully supported by numerical simulations (Fig. 3.4).

MN coupling.—In this case, there exist distinct regimes of z_1, three of which among them are of relevance for the giant component properties (in $N \to \infty$ limit).

(i) $0 \le z_1 \le \ln 2$.

In this regime, more than half of nodes are of degree zero in each layer so every linked node in one layer is coupled with a degree-0 node in the other layer under MN coupling. After some inspection one obtains the total degree distribution $P(k)$ as

$$P_{MN}(k) = \begin{cases} 2\pi(0) - 1 & (k = 0), \\ 2\pi(k) & (k \ge 1). \end{cases} \tag{3.12}$$

In this regime there is no giant component.

(ii) $\ln 2 \le z_1 \le z^*$.

Following similar steps, $P(k)$ in this regime is obtained as

$$P_{MN}(k) = \begin{cases} 0 & (k = 0), \\ 2[2\pi(0) + \pi(1) - 1] & (k = 1), \\ 2\pi(2) - 2\pi(0) + 1 & (k = 2), \\ 2\pi(k) & (k \ge 3). \end{cases} \tag{3.13}$$

In this regime, $\langle k^2 \rangle - 2\langle k \rangle = 2(z_1^2 - z_1 - 2e^{-z_1} + 1)$, which becomes positive for $z_1 > z_c^{MN}$ where

$$z_c^{MN} = 0.838587497... \tag{3.14}$$

Therefore the giant component emerges at a much higher link density. Being delayed in its birth, however, the giant component grows more abruptly once formed (Fig. 3.4c). This regime is terminated at $z_1 = z^*$, determined by the condition $2\pi(0) + \pi(1) = 1$, from which we have $z^* = 1.14619322...$

(iii) $z_1 \ge z^*$.
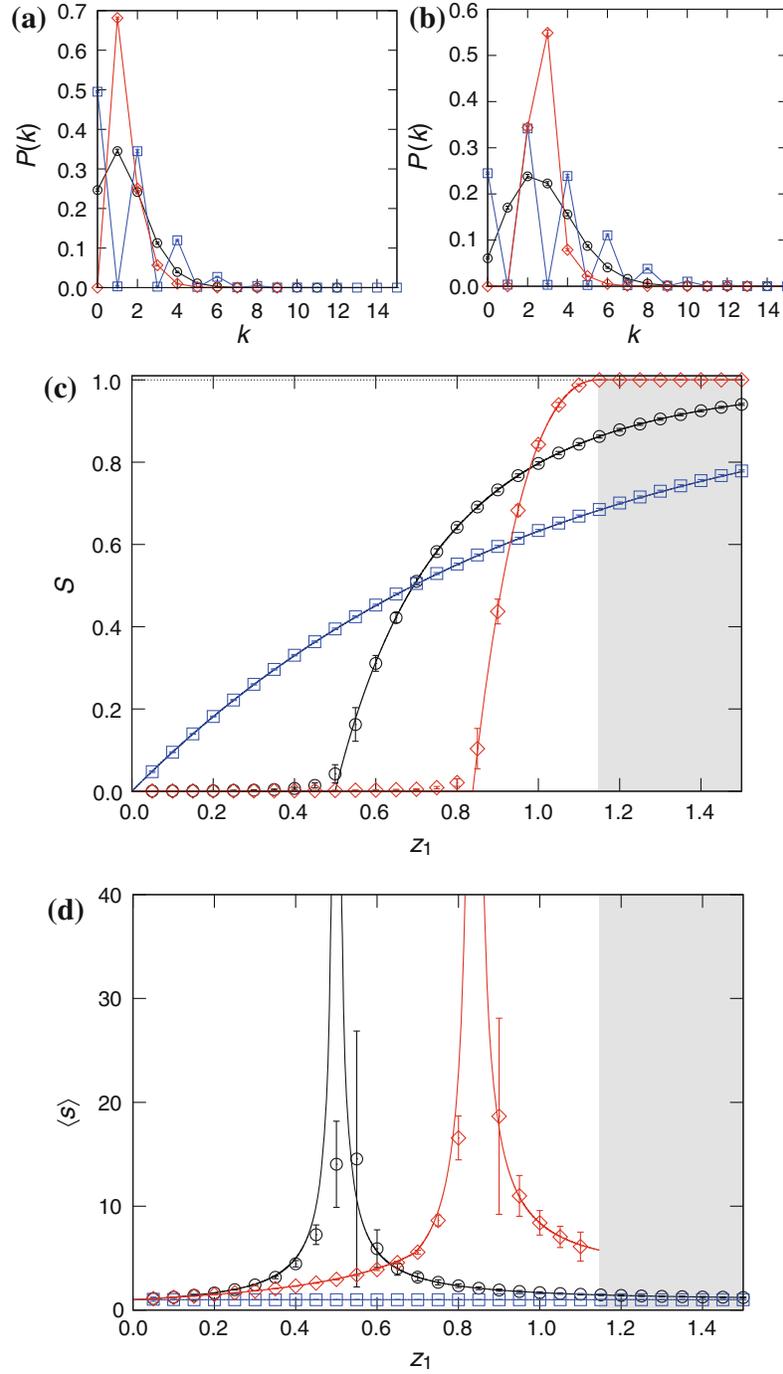
Fig. 3.4 **a, b** Total degree distribution $P(k)$ of duplex ER graphs with $z_1 = z_2 = 0.7$ (**a**) and $z_1 = z_2 = 1.4$ (**b**). Different symbols denotes MP (*square*), uncorrelated (*circle*), and MN (*diamond*) couplings. **c, d** The giant component size S (**c**) and the average size of finite components $\langle s \rangle$ (**d**) as a function of z_1 of duplex ER graphs with $z_1 = z_2$. Same symbols as (a, b) are used. Gray shade denotes the region in which $S = 1$ for the MN case ($z_1 > z^*$). Lines represent the theoretical curves and symbols the numerical simulation results. Errorbars denote standard deviations. Adapted from [3]
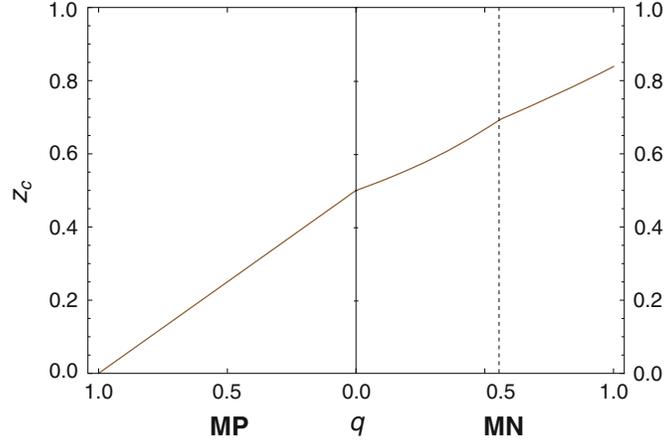
Fig. 3.5 Plot of Eq. (3.15a) for the critical mean degree z_c as a function of q, the fraction of correlated multipex nodes. The cases $q = 1$ denote maximally correlated multiplexity, $0 < q < 1$ partially correlated multiplexity, and $q = 0$ uncorrelated multiplextiy. The vertical dotted line is drawn at $q = 2 - 1/\ln 2$ across which z_c takes different formulae in Eq. (3.15b). Adapted from [3]

In this regime we have $P(0) = P(1) = 0$ and thereby $S = 1$. This means that the entire network becomes connected into a single component at this finite link density z^*, which can never be achieved for ordinary ER networks.

All these theoretical results are confirmed numerically (Fig. 3.4). Meanwhile, it is noteworthy that despite these abnormal behaviors and apparently more rapid growth of S near z_c, the critical behavior in the MN case is found to be consistent with that of standard mean-field [3].

Imperfect correlated multiplexity.—So far we have seen that maximally correlated or anti-correlated multiplexity crucially affects the onset of emergence of giant component in multiplex ER networks. For a partially correlated duplex ER network (with equal link densities) in which a fraction q of nodes are maximally correlated coupled while the rest fraction $1 - q$ are randomly coupled, the total degree distribution can be obtained as $P_{partial}(k) = qP_{maximal}(k) + (1 - q)P_{uncorr}(k)$, where *maximal* is either *MP* or *MN*. Using Eqs. (3.8, 3.12, 3.13) and following similar steps as in the previous section we obtain the critical link density as a function of q as

$$z_c = (1 - q)/2 \tag{3.15a}$$

for positively correlated case and

$$z_c = \begin{cases} 1/(2 - q) & (q < 2 - 1/\ln 2), \\ z_1(q) & (q > 2 - 1/\ln 2) \end{cases} \tag{3.15b}$$

for negatively correlated case, where $z_1(q)$ is the solution of $(2-q)z_1^2 - z_1 - 2qe^{-z_1} + q = 0$. This result shows that z_c depends continuously on q (Fig. 3.5), illustrating that the effect of correlated multiplexity is present for general q.
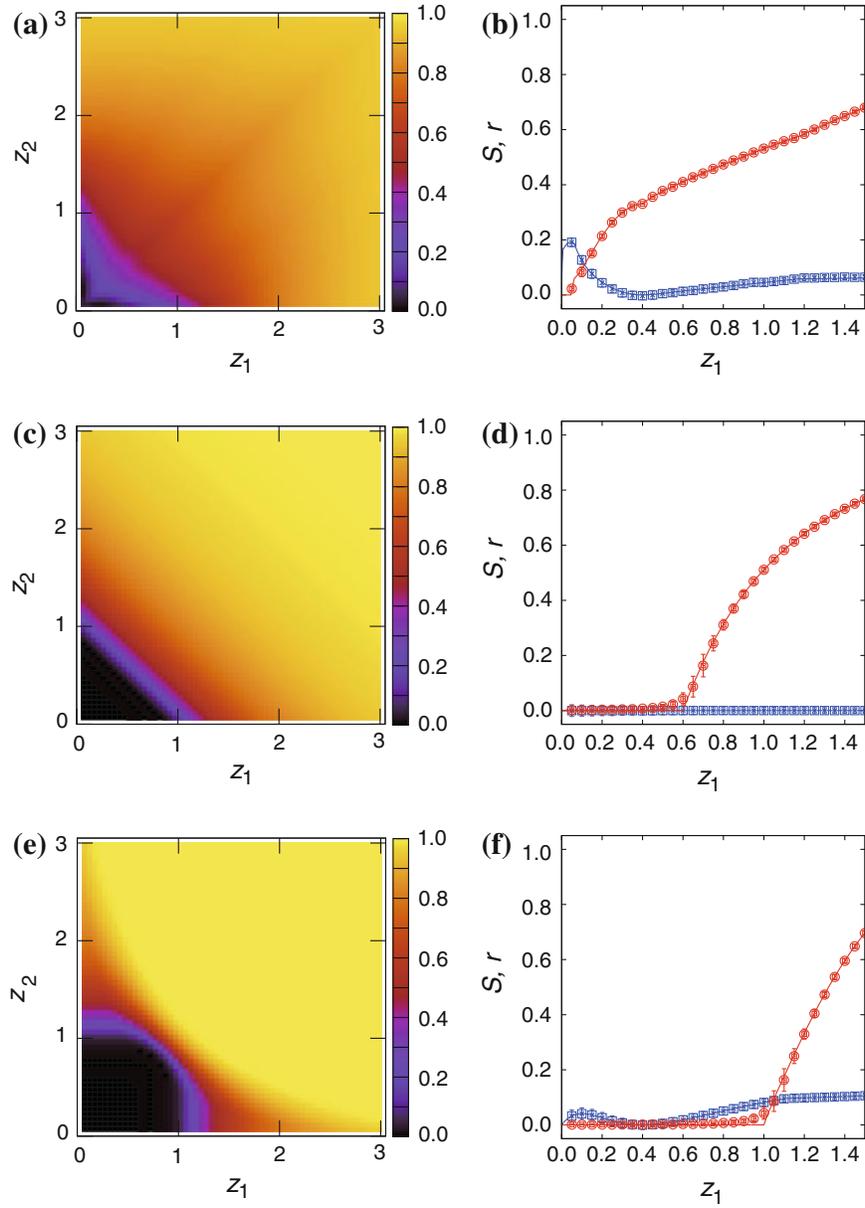
Fig. 3.6 **a**, **c**, **e** Numerical simulation results of the size of giant component of duplex ER networks of size $N = 10^4$ with **a** MP, **c** uncorrelated, and **e** MN couplings. **b**, **d**, **f** The giant component size S (*red*) is plotted for $z_2 = 0.4$, along with the assortativity coefficient r (*blue*) for the MP (**b**), uncorrelated (**d**), and MN (**f**) cases. Theoretical predictions based on the joint degree distribution in Sect. 3.4 are shown in lines, demonstrating excellent agreements with simulations. Errorbars denote standard deviations from 10^4 independent runs. Adapted partly from [3]

3.5.3 Duplex ER Networks with General Link Densities

In this section we consider general duplex ER networks with $z_1 \neq z_2$. Qualitative picture of behavior of giant component size is similar to the equal link density case: the giant component emerges at lower link densities for the MP case but grows more slowly than the uncorrelated case, whereas it emerges at higher link densities for the MN case but grows more abruptly and connects all the nodes in the network at finite link density (Fig. 3.6).

It should be emphasized, however, that one should use the formalism in Sect. 3.4, which fully exploits the joint degree distribution, in order to obtain correct theoretical results for $z_1 \neq z_2$ (Fig. 3.4b, d, f). Indeed, the assortativity coefficient r calculated both analytically by Eqs. (3.4–3.6) and numerically shows that it is assortative in MP and MN cases, except for $z_1 = z_2$. This clearly shows that the correlated multiplexity can not only modulate the total degree distribution $P(k)$ of the superposed network but also introduce higher-order correlations in its network structure.

3.6 Duplex SF Networks

Now we consider a duplex scale-free (SF) network, in which two SF networks constructed by the static model [27] are multiplex-coupled. The static model network is constructed as follows. Each node i ($i = 1, \cdots, N$) is assigned a weight $w_i = i^{-a}$, where a is a constant greater than 1. By successively connecting two nodes each chosen with probability proportional to its weight until desired number of links are made, one obtains a network with asymptotic power-law degree distribution $\pi(k) \sim k^{-\gamma}$, with γ (called the degree exponent) given by $\gamma = 1 + 1/a$ [27]. Thus one can tune both the degree exponent and the mean degree of the network.

An important property of SF networks is the vanishing percolation threshold for $\gamma \leq 3$ [28], fundamentally different from the case with $\gamma > 3$. The case of $\gamma = 2.5$ is examined first (Fig. 3.7a). In this case the giant component exists for any $z_1 > 0$ even in the single layer, so $z_c = 0$ in all three cases. For small z_1, MP has the largest giant component size as in the ER case. Peculiar behavior is observed for the MN coupling, in which the giant component size increases slowly until it makes a jump around $z_{jump} \approx 1.05$, almost doubling its size. This unusual behavior is rooted in the fact that with MN coupling each layer's hub supports giant component of its own and the two giant components are totally disjoint until the link density reaches the threshold z_{jump}. Beyond this threshold, the two equally-large giant components cannot but overlap and merge, thereby making a jump. This picture is supported by the observations that sizes of the largest and second largest component are almost equal, and the position of jump coincides with the point at which all nodes in the network acquire at least one link (Fig. 3.7a, inset). The case of $\gamma = 5.0$ is examined next (Fig. 3.7b), yielding overall similar qualitative behaviors as the duplex ER networks, without any discontinuous jump.
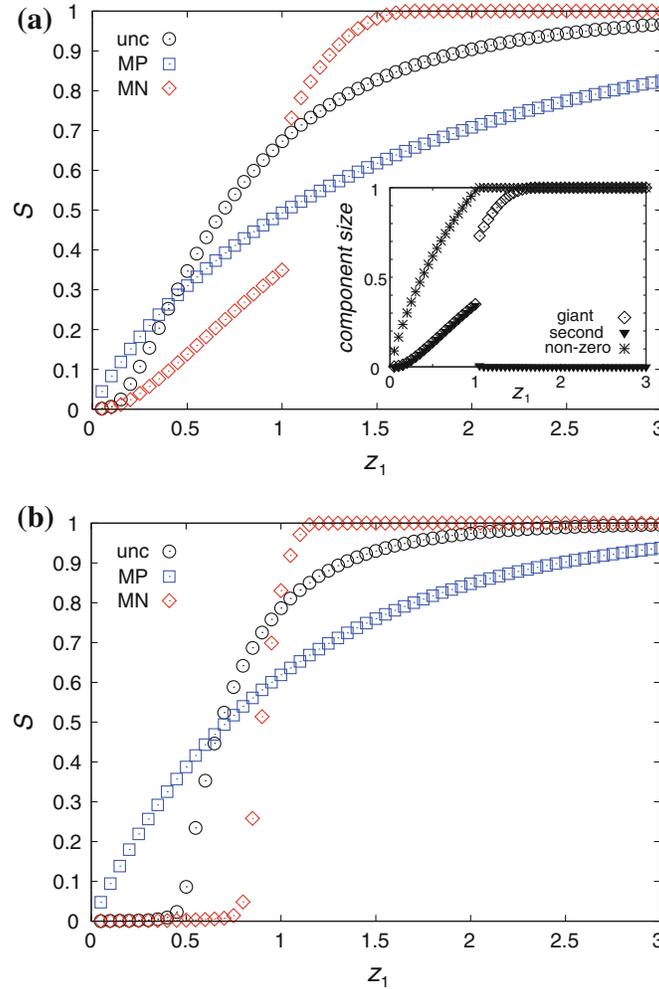
Fig. 3.7 Giant component size of duplex SF networks of equal link densities with $\gamma = 2.5$ (**a**) and $\gamma = 5.0$ (**b**). Symbols stand for uncorrelated (○), MP (□), and MN (◇) couplings. (Inset) Size of largest (◇) and second largest (▽) components, together with the fraction of nonzero-degree nodes (∗), for MN coupling

3.6.1 Betweenness and Load

Betweenness centrality [29] or load [27] is a widely-used centrality measure which characterizes the potential burden or traffic over a node in a network due to simple shortest path-based transport protocols. It has been shown that the load distribution of SF network also follows a power law, with the exponent ≈ 2.2 for non-tree SF networks with $2 < \gamma \leq 3$ [27]. Here we examine how the betweenness and its distribution are affected by the multiplex coupling of SF networks. From the scaling perspective, neither the degree exponent nor the power-law exponent for betweenness distribution is found to be affected by the multiplex coupling (Fig. 3.8a, b). Looking at the individual node level, it is found that the betweenness changes most when the
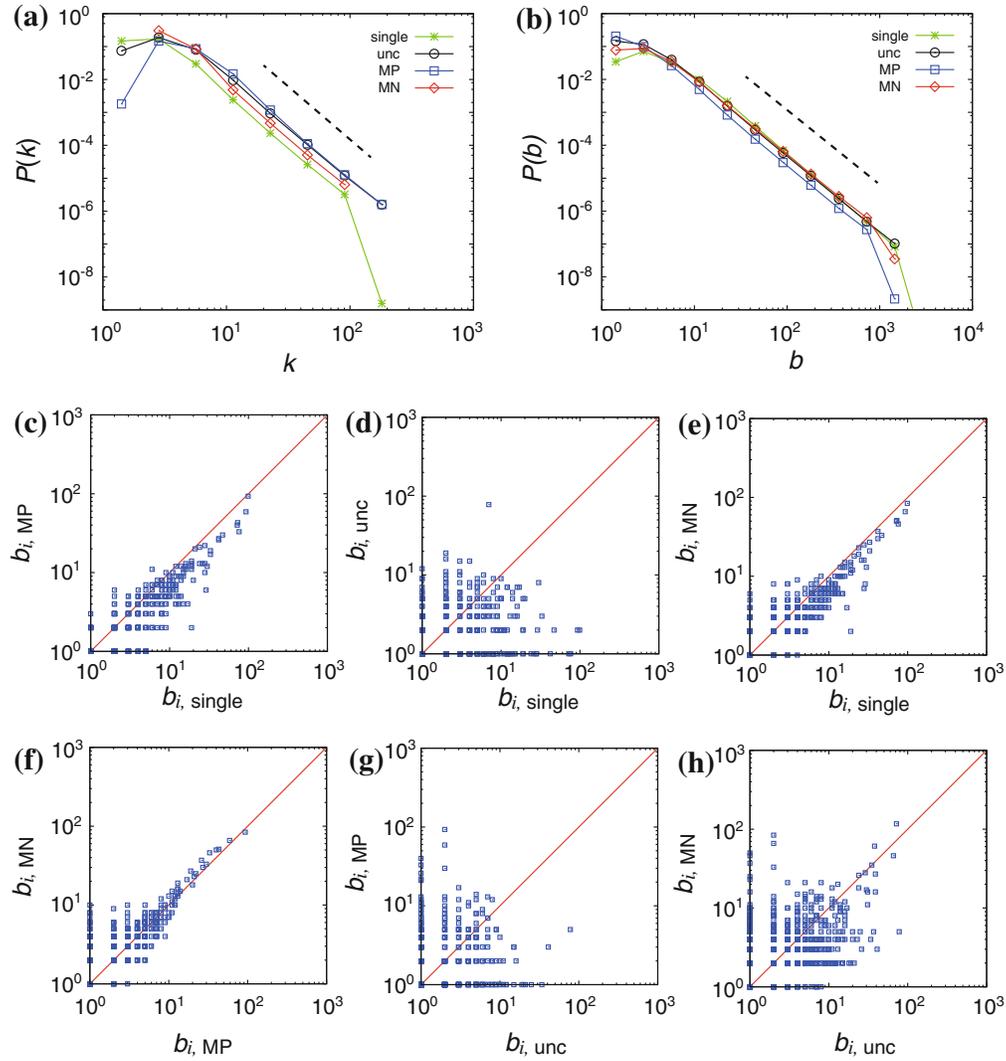
Fig. 3.8 **a** Total degree distribution and **b** betweenness distributions of duplex SF networks. Dashed line has the slope -3.0 (**a**) and -2.2 (**b**), drawn as a guide to the eye. **c–h** Scatterplots of betweenness centralities of a node in the two layers for different multiplex couplings. Diagonal lines are drawn as a guide to the eye

two networks are coupled randomly, rather than in a MP or MN way (Fig. 3.8c–h). This suggests that in MP or MN coupling the pathway structure is weakly affected and topological centralities of hub nodes are largely preserved. Concepts of betweenness and load are intimately related with the definition of shortest path. One interesting issue in this regard is the concept of optimal path in multiplex networks with the context and interplay between layers fully taken into account, which deserves further study.
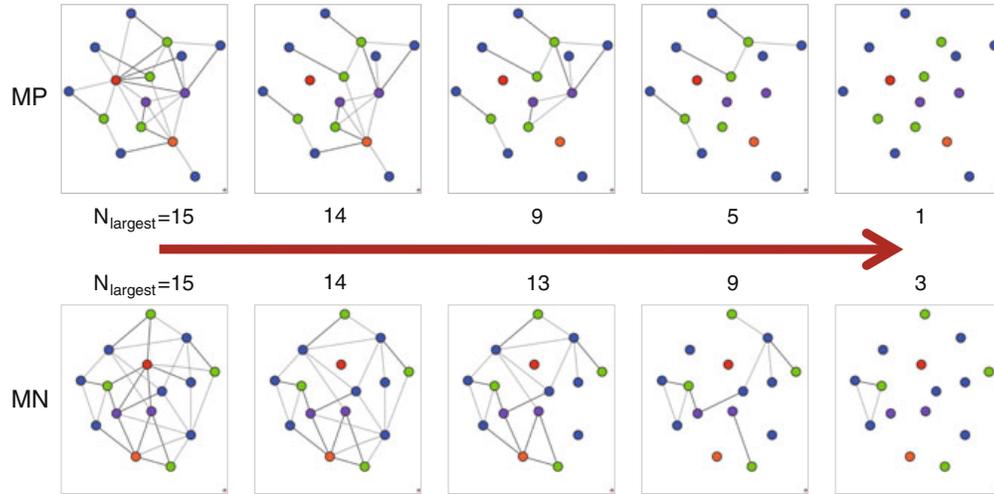
Fig. 3.9 Schematic diagram of intentional attack on MP and MN type multiplex networks. From left to right, nodes are removed in descending order of the total degree to simulate an intentional attack, and the size of largest connected component in the remaining superposed network is monitored

3.7 Robustness of Multiplex Networks

Having established that correlated multiplexity can significantly affect the overall connectivity of multiplex networks, the next question we might have is its impact on network robustness against random failures or intentional attack [2]. For example, as the cartoon diagram in Fig. 3.9 shows, the way how the network layers are multiplex-coupled can alter the resilience of the superposed network against attack. It has also been shown that robustness of interdependent networks to cascade of failures can be affected by the correlated coupling [14, 30].

As a preliminary case study, here we use the multiplex coauthorship network introduced in Sect. 3.2 and examine the topological robustness under various failure and attack scenarios. We construct the SOC-KPZ coauthorship network, consisting of the nodes participating in both layers and the links among them. Then we simulate virtual random node or link failures and degree-based intentional node attacks, and measure the fraction of nodes in the initial largest component that still form largest component in the remaining superposed network, denoted S/S_0, as a function of the fraction of removed nodes or links f. We also compare the results against those obtained from three shuffled networks, in which the two layers are MP, uncorrelated, and MN-coupled (obtained by shuffling the node names in each layer according to the coupling rule, while controlling the link density of the superposed network to be equal) (Fig. 3.10). It is noteworthy that even though the coauthorship networks show positive correlated multiplexity (Fig. 3.2), the topological robustness properties do not always correspond to those of MP-correlated networks. For example, the real coauthorship network is more vulnerable, albeit slightly, to random link removals than its uncorrelated versions, in contrast to the higher robustness of MP-correlated
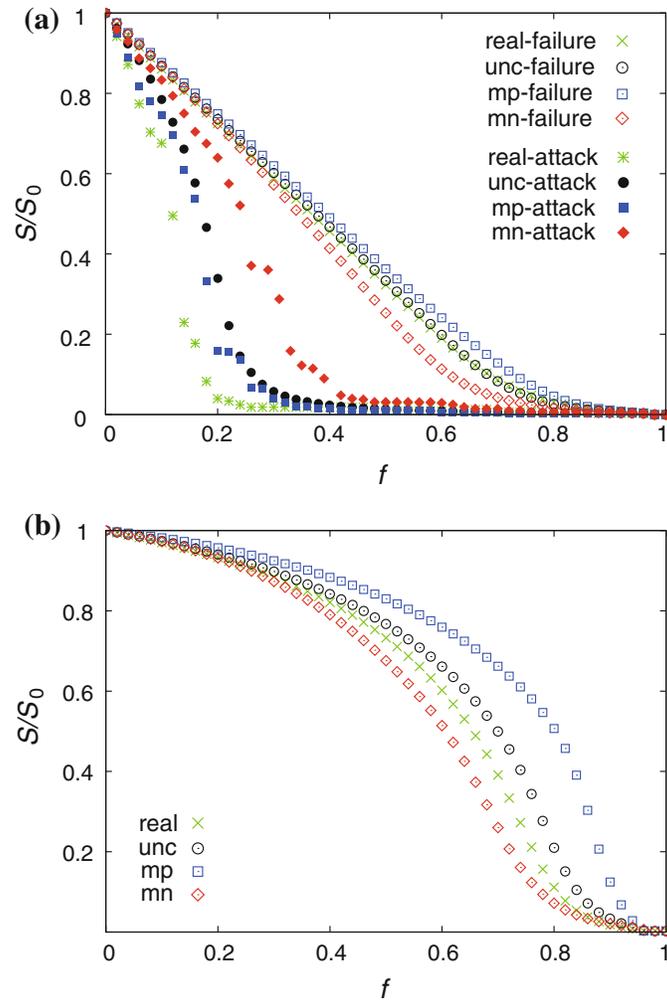
Fig. 3.10 Topological robustness of correlated multiplex networks. **a** Relative size of largest component S/S_0 of the superposed network under random (failure) and intentional (attack) removal of nodes of fraction f. The intentional attack was simulated by removing nodes in descending order of total degree. Shown are results for the multiplex coauthorship network (SOC-KPZ) ($\times$, $*$) and its three shuffled versions, MP (*square*), uncorrelated (*circle*), and MN-coupled networks (*textitdia-mond*). **b** Same plots for random link removals. Data are averaged over 10^4 independent simulation runs

networks than the uncorrelated ones (Fig. 3.10b). Such discrepancy indicates the presence of higher-order correlations in the coupling structure of real multiplex networks, beyond the degree correlation. More systematic investigation on this topic using model networks is currently underway (B. Min et al., arXiv:1307.1253).

3.8 Dynamics on Multiplex Networks

Multiplexity can also have impact on network dynamics [31]; in fact it is one of the ultimate goals of the study of multiplex networks to understand what the generic effect of correlated multiplexity on various dynamic processes occurring on top of real-world multiplex complex systems. This may have implications on many profound real-world complex systems problems, such as understanding, predicting, and controlling systemic risk and collective social movement. Dynamics with multiplexity in general, poses the question of how the interplay of different network layers can bring about emergent dynamic consequences, and in many cases calls for development of new theoretical tools, similarly to what we did in Sect. 3.4 for structural analysis, which raises theoretical challenge as well.

Study of dynamical processes on multiplex networks is still in its infancy, yet is rapidly growing over the years [32–39]. Surveying all these recent effort would already require a separate contribution; here we could merely compile them with a brief summary of key findings. Given the obvious relevance of multiplex-network framework for many real-world problems, such as social cascades in social networks [5] or dynamics of systemic risk [40], this list is expected to expand quickly so is by no means meant to be exhaustive.

One of the first studies on multiplex dynamics was the study of sandpile dynamics [32], where it is found that the scaling behavior of avalanche does not change by the multiplex coupling, despite alterations in the detailed cascade dynamics. Generalized models of behavioral cascades in multiplex social networks [33, 34] showed that the multiplexity can facilitate global cascades compared to null models of simplex networks. In the study of random Boolean network on multiplex networks [35], the multiplex coupling is shown to support stabilization of the system even when each single layer is in the unstable chaotic state. In studies of evolutionary dynamics on multiplex networks, it was shown that the cooperative behavior is enhanced when individuals interact through multiple network layers [36, 37]. In the study of diffusion dynamics on multiplex networks [38], the existence of multiple channels of diffusive motion is shown to speed up the diffusion process. These studies collectively highlight how the dynamical properties on multiplex networks can differ from those of a single or simplex network.

3.9 Summary and Outlook

In summary, we have surveyed recent studies on multiplex networks, the networks with explicit multiple types of links, which is a better representation of real-world complex systems. Particularly emphasized are the notion of correlated multiplexity and its effect on the structural properties of multiplex network system. We have introduced the random graph models of correlated multiplex networks and developed analytical formalism to study its structural properties. Applications to multiplex

ER and SF networks demonstrated that the correlated multiplexity can dramatically change the properties of the giant component. This shows that a multiplex complex system can exhibit structural properties that cannot be represented by its individual network layer's properties alone. Such nontrivial, emerging multiplex structure should entail significant impact on dynamical processes occurring on it, opening a vast avenue of future studies on the impact of correlated multiplexity on network dynamics and function [14, 30].

The concepts and tools for the multiplex network should also be useful in the study of related subjects of recent interest such as layered [8], multi-type [23], interacting [9, 41], and interdependent networks [10, 24, 42], which share similar theoretical framework and mathematical techniques. Notable areas for further investigation would be, to name but a few, the multiplex network evolution [43] and the role of negative or antagonistic interactions between layers [11, 44]. Altogether, these studies will cooperatively help establish unified framework for the emerging paradigm of "network of networks," and the concept of network multiplexity will play an essential role in this collective endeavor.

References

1. M. E. J. Newman, *Networks: An introduction* (Oxford University Press, Oxford, 2010)
2. S. Havlin and R. Cohen, *Complex neworks: Structure, robustness, and function* (Cambridge Unviersity Press, Cambridge, 2010).
3. K.-M. Lee, J. Y. Kim, W.-K. Cho, K.-I. Goh, and I.-M. Kim, New J. Phys. **14**, 033027 (2012).
4. S. Wasserman and K. Faust, *Social network analysis* (Cambridge University Press, Cambridge, 1994).
5. M. O. Jackson, *Social and economic networks* (Princeton University Press, Princeton, NJ, 2008).
6. Z. Maoz, *Networks of nations: The Evolution, structure, and impact of international networks, 1816-2001* (Cambridge University Press, Cambridge, 2010).
7. M. Buchanan, G. Caldarelli, P. De Los Rios, F. Rao, and M. Vendruscolo, *Networks in cell biology* (Cambridge University Press, Cambridge, 2010).
8. M. Kurant and P. Thiran, Phys. Rev. Lett. **96**, 138701 (2006).
9. E. A. Leicht and R. M. D'Souza, e-print (2009) arXiv:0907.0894.
10. S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, Nature **464**, 1025 (2010).
11. M. Szell, R. Lambiotte, and S. Thurner, Proc. Natl. Acad. Sci. U.S.A. **107**, 13636 (2010).
12. D. Lee, K.-I. Goh, B. Kahng, D. Kim, Phys. Rev. E **82**, 026112 (2010).
13. A. Lomi, P. Pattison, Org. Sci. **17**, 313 (2006).
14. R. Parshani, C. Rozenblat, D. Ietri, C. Ducruet, and S. Havlin, EPL **92**, 68002 (2010).
15. R. G. Morris and M. Barthelemy, Phys. Rev. Lett. **109**, 128703 (2012).
16. H. W. Han, et al. Nucl. Acids Res. **41**, 9209 (2013).
17. M. Barigozzi, G. Fagiolo, and D. Garlaschelli, Phys. Rev. E **81**, 046104 (2010).
18. B. Soderberg, Phys. Rev. E **68**, 015102 (2003).
19. D.-H. Kim, B. Kahng, and D. Kim, Eur. Phys. J. B **38**, 305 (2004).

20. M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
21. M. Molloy and B. Reed, Random Struct Algorithms **6**, 161 (1995).
22. J. P. Gleeson, S. Melnik, and A. Hackett, Phys. Rev. E **81**, 066114 (2010).
23. A. Allard, P.-A. Noël, L. J. Dubé, and B. Pourbohloul, Phys. Rev. E **79**, 036113 (2009).
24. S.-W. Son, G. Bizhani, C. Christensen, P. Grassberger, and M. Paczuski, EPL **97**, 16006 (2012).
25. M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
26. P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).
27. K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
28. R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
29. L. C. Freeman, Sociometry **40**, 35 (1977).
30. S. V. Buldyrev, N. Shere, and G. A. Cwilich, Phys. Rev. E **83**, 016112 (2011).
31. A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamic processes on complex networks* (Cambridge University Press, Cambridge, 2008).
32. K.-M. Lee, K.-I. Goh, and I.-M. Kim, J. Korean Phys. Soc. **60**, 641 (2012).
33. C. D. Brummitt, K.-M. Lee, and K.-I. Goh, Phys. Rev. E **85**, 045102(R) (2012).
34. O. Yağan and V. Gligor, Phys. Rev. E **86**, 036103 (2012).
35. E. Cozzo, A. Arenas, and Y. Moreno, Phys. Rev. E **86**, 036115 (2012).
36. J. Gómez-Gardeñes, I. Reinares, A. Arenas, and L. M. Floria, Sci. Rep. **2**, 620 (2012).
37. J. Gómez-Gardeñes, C. Gracia-Lázaro, L. M. Floría, and Y. Moreno, Phys. Rev. E **86**, 056113 (2012).
38. S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas, Phys. Rev. Lett. **110**, 028701 (2013).
39. S. Shai and S. Dobson, Phys. Rev. E **86**, 066120 (2012).
40. K.-M. Lee, J.-S. Yang, G. Kim, J. Lee, K.-I. Goh, and I.-M. Kim, PLoS ONE **6**, e18443 (2011).
41. J. F. Donges, H. C. H. Schultz, N. Marwan, Y. Zou, and J. Kurths, Eur. Phys. J. B **84**, 635 (2011).
42. D. Zhou, H. E. Stanley, G. D'Agostino, and A. Scala, Phys. Rev. E **86**, 066103 (2012).
43. J. Y. Kim and K.-I. Goh, Phys. Rev. Lett. **111**, 058702 (2013).
44. K. Zhao and G. Bianconi, J. Stat. Mech. P05005 (2013).

Chapter 4
Modeling Interdependent Networks as Random Graphs: Connectivity and Systemic Risk

R. M. D'Souza, C. D. Brummitt and E. A. Leicht

Abstract Idealized models of interconnected networks can provide a laboratory for studying the consequences of interdependence in real-world networks, in particular those networks constituting society's critical infrastructure. Here we show how random graph models of connectivity between networks can provide insights into shifts in percolation properties and into systemic risk. Tradeoffs abound in many of our results. For instance, edges between networks confer global connectivity using relatively few edges, and that connectivity can be beneficial in situations like communication or supplying resources, but it can prove dangerous if epidemics were to spread on the network. For a specific model of cascades of load in the system (namely, the sandpile model), we find that each network minimizes its risk of undergoing a large cascade if it has an intermediate amount of connectivity to other networks. Thus, connections among networks confer benefits and costs that balance at optimal amounts. However, what is optimal for minimizing cascade risk in one network is suboptimal for minimizing risk in the collection of networks. This work provides tools for modeling interconnected networks (or single networks with mesoscopic structure), and it provides hypotheses on tradeoffs in interdependence and their implications for systemic risk.

R. M. D'Souza (✉) · C. D. Brummitt
University of California, Davis, CA 95616, USA
e-mail: raissa@cse.ucdavis.edu

C. D. Brummitt
e-mail: cbrummitt@math.ucdavis.edu

E. A. Leicht
CABDyN Complexity Center, University of Oxford, Oxford OX1 1HO, UK
e-mail: elizabeth.leicht@wolfson.ox.ac.uk

4.1 Introduction

Collections of networks occupy the core of modern society, spanning technological, biological, and social systems. Furthermore, many of these networks interact and depend on one another. Conclusions obtained about a network's structure and function when that network is viewed in isolation often change once the network is placed in the larger context of a network-of-networks or, equivalently, when viewed as a system composed of complex systems [13, 15]. Predicting and controlling these über-systems is an outstanding challenge of increasing importance because system interdependence is growing in time. For instance, the increasingly prominent "smart grid" is a tightly coupled cyber-physical system that relies on human operators and that is affected by the social networks of human users. Likewise, global financial markets are increasingly intertwined and implicitly dependent on power and communication networks. They are witnessing an escalation in high frequency trades executed by computer algorithms allowing for unanticipated and uncontrolled collective behavior like the "flash crash" of May 2010. Reinsurance companies uncannily forecast the increase of extreme events (in particular in the USA) just weeks before the onslaught of Superstorm Sandy [59] and stressed the urgent need for new scientific paradigms for quantifying extreme events, risk, and interdependence [54].

Critical infrastructure provides the substrate for modern society and consists of a collection of interdependent networks, such as electric power grids, transportation networks, telecommunications networks, and water distribution networks. The proper collective functioning of all these systems enables government operations, emergency response, supply chains, global economies, access to information and education, and a vast array of other functions. The practitioners and engineers who build and maintain critical infrastructure networks have long been cataloging and analyzing the interdependence between these distinct networks, with particular emphasis on failures cascading through coupled systems [19, 21, 29, 42, 51, 55, 56, 60, 61, 63].

These detailed, data driven models are extremely useful but not entirely practical due to the diversity within each infrastructure and due to difficulty in obtaining data. First, each critical infrastructure network is independently owned and operated, and each is built to satisfy distinct operating regimes and criteria. For instance, consider the distinct requirements and constraints of a municipal transportation system versus a region of an electric power grid. Even within a municipal transportation system there exist multiple networks and stakeholders, such as publicly funded road networks and private bus lines and train networks. Second, there are few incentives for distinct operators to share data with others, so obtaining a view of a collection of distinctly owned systems is difficult. Third, the couplings between the distinct types of infrastructure are often only revealed during extreme events; for instance, a natural gas outage in New Mexico in February 2011 caused rolling electric power blackouts in Texas [16]. Thus, even given the most detailed knowledge of individual critical infrastructure systems, it is still difficult to anticipate new types of failures mechanisms (i.e., some failure mechanisms are "unknown unknowns").

Idealized models for interdependent networks provide a laboratory for discovering unknown couplings and consequences and for developing intuition on the new emergent phenomena and failure mechanisms that arise through interactions between distinct types of systems. In fact, the idea of modeling critical infrastructure as a collection of "complex interactive networks" was introduced over a decade ago [3]. Yet idealized models are only starting to gain traction [58, 71], and they are largely based on techniques of random graphs, percolation and dynamical systems (with many tools drawn from statistical physics). Despite using similar techniques, these models can lead to contrasting conclusions. Some analytic formulations show that interdependence makes systems radically more vulnerable to cascading failures [15], while others show that interdependence can confer resilience to cascades [13].

Given a specified set of network properties, such as a degree distribution for the nodes in the network, random graph models consider the ensemble of all graphs that can be enumerated consistent with those specified properties. One can use probability generating functions to calculate the average or typical properties of this ensemble of networks. In the limit of an infinitely large number of nodes, the generating functions describing structural and dynamic properties are often exactly solvable [52], which makes random graphs appealing models that are widely used as simple models of real networks. Of course there are some downsides to using the random graph approach, which will require further research to quantify fully. First, in the real-world we are typically interested in properties of individual instances of networks, not of ensemble properties. Second, percolation models on random graphs assume local, epidemic-like spreading of failures. Cascading failures in the real-world, such as cascading blackouts in electric power grids, often exhibit non-local jumps where a power line fails in one location and triggers a different power line hundreds of miles away to then fail (e.g., see Ref. [1]). This issue is discussed in more detail below in Sect. 4.3.4.1. Nonetheless, random graphs provide a useful starting point for analyzing the properties of systems of interdependent networks.

Here, in Sect. 4.2 we briefly review how random graphs can be used to model the structural connectivity properties between networks. Then, in Sect. 4.3 we show how, with the structural properties in place, one can then analyze dynamical process unfolding on interconnected networks with a focus on cascades of load shedding.

4.2 Random Graph Models for Interconnected Networks

Our model of "interconnected networks" consists of multiple networks (i.e., graphs) with edges introduced between them. Thus, the system contains multiple kinds of nodes, with one type of node for each network, and one type of edge. A simple illustration of a system of two interconnected networks is shown in Fig. 4.1. (A related class of graphs called multiplex networks considers just one type of node but multiple kinds of edges [49, 70].) This general framework can model different kinds of systems that have connections to one another, or it can capture mesoscopic structure in a single network, such as communities and core-periphery structure.

Fig. 4.1 A stylized illustration of two interconnected networks, *a* and *b*. Nodes interact directly with other nodes in their immediate network, yet also with nodes in the second network
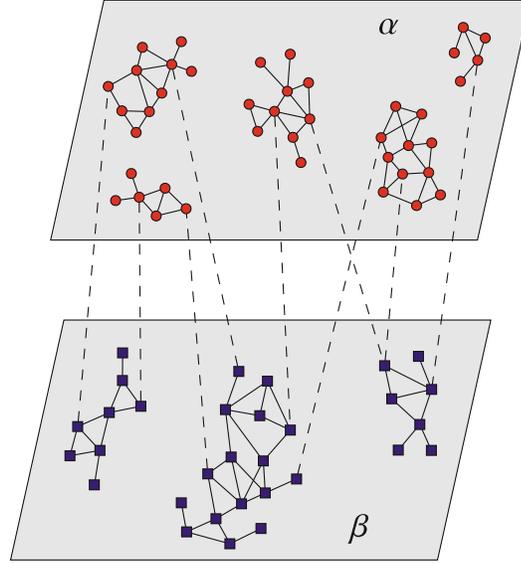
4.2.1 Mathematical Formulation

Here we briefly review the mathematics for calculating the structural properties of interconnected networks as discussed in Ref. [40]. In a system of $d \geq 2$ interacting networks, an individual network μ is characterized by a multi-degree distribution $\{p_{\mathbf{k}}^{\mu}\}$, where $\mathbf{k}$ is a d-tuple, $(k_1, \ldots, k_d)$, and $p_{\mathbf{k}}^{\mu}$ is the probability that a randomly chosen node in network μ has k_{ν} connections with nodes in network ν. A random graph approach considers the ensemble of all possible networks consistent with this multi-degree distribution. To realize a particular instance of such a network we take the "configuration model" approach [10, 47]. Starting from a collection of isolated nodes, each node independently draws a multi-degree vector from $\{p_{\mathbf{k}}^{\mu}\}$. Next, each node is given k_{ν} many "edge stubs" (or half-edges) of type ν. We create a graph from this collection of labeled nodes and labeled edge stubs by matching pairs of compatible edge stubs chosen uniformly at random. For instance, an edge stub of type ν belonging to a node in network μ is compatible only with edge stubs of type μ belonging to nodes in network ν. Generating functions allow us to calculate the properties of this ensemble.

The generating function for the $\{p_{\mathbf{k}}^{\mu}\}$ multi-degree distribution is

$$G_{\mu}(\mathbf{x}) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_d=0}^{\infty} p_{\mathbf{k}}^{\mu} \prod_{\nu=1}^{d} x_{\nu}^{k_{\nu}}, \tag{4.1}$$

where $\mathbf{x}$ is the d-tuple, $\mathbf{x} = (x_1, \ldots, x_d)$. This is a generating function for a probability distribution already known to us (our multi-degree distribution for network μ), and thus not terribly informative on its own. However, we can derive additional generating functions for probability distributions of interest, such as the distribution of sizes of connected components in the system. However, we much first derive
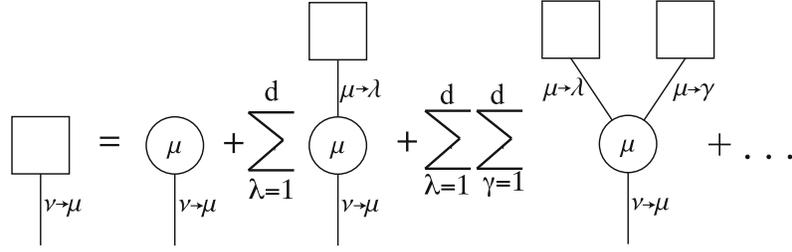
Fig. 4.2 A diagramatical representation of the topological constraints placed on the generating function $H_{\mu\nu}(\mathbf{x})$ for the distribution of sizes of components reachable by following a randomly chosen ν-μ edge. The labels attached to each edge indicate type or *flavor* of the edge, and the sum runs over over all possible flavors

two intermediate generating function forms, one for the probability distribution of connectivity for a node at the end of a randomly chosen edge and a second for the probability distribution of component sizes found at the end of a random edge. Reference [52] contains a clear and thorough discussion of this approach for a single network, which we apply here to multiple networks.

First consider following an edge from a node in network ν to a node in network μ. The μ node is k_ν times more likely to have ν-degree k_ν than degree 1. Thus the probability $q_{\mathbf{k}}^{\mu\nu}$ of reaching a μ-node of ν-degree k_ν is proportional to $k_\nu p_{k_1\cdots k_\nu\cdots k_d}^\mu$. Accounting for the fact that we have followed an edge from a node in ν to a node in μ, the properly normalized generating function for the distribution of additional edges from that μ-node is

$$G_{\mu\nu}(\mathbf{x}) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_d=0}^{\infty} \frac{(k_\nu + 1)p_{k_1\cdots(k_\nu+1)\cdots k_l}^\mu}{\overline{k}_{\mu\nu}} \prod_{\gamma=1}^{d} x_\gamma^{k_\gamma} = \frac{G_\mu'^\nu(\mathbf{x})}{G_\mu'^\nu(\mathbf{1})}. \quad (4.2)$$

Here $\overline{k}_{\mu\nu} = \sum_{k_1} \cdots \sum_{k_d} k_\nu p_{\mathbf{k}}^\mu$ is the normalization factor accounting for $G_{\mu\nu}(\mathbf{1}) = 1$ and $\overline{k}_{\mu\nu}$ is also the average ν-degree for a node in network μ. We use $G_\mu'^\nu(\mathbf{x})$ to denote the first derivative of $G_\mu(\mathbf{x})$ with respect to x_ν and thus $G_\mu'^\nu(\mathbf{1}) = \overline{k}_{\mu\nu}$. A system of d interacting networks has d^2 excess degree generating functions of the form shown in Eq. 4.2.

Now consider finding, not the connectivity of the μ-node, but the size of the connected component to which it belongs. This probability distribution for sizes of components can be generated by iterating the random-edge-following process described in Eq. 4.2, where we must consider all possible types of nodes that could be attached to that μ-node. For an illustration see Fig. 4.2. In other words, the μ-node could have no other connections; it might be connected to only one other node and that node could belong to any of the d networks; it might be connected to two other nodes that could each belong to any of the d networks; and so on. Iterating the

random-edge construction for each possibility leads to a generating function $H_{\mu\nu}$ for the sizes of components at the end of a randomly selected edge

$$H_{\mu\nu}(\mathbf{x}) = x_\mu q_{0\cdots0}^{\mu\nu} \tag{4.3}$$

$$+ x_\mu \sum_{k_1\ldots k_d=0}^{1} \delta_{1,\sum_{\lambda=1}^{d} k_\lambda} q_{k_1\cdots k_d}^{\mu\nu} \prod_{\gamma=1}^{d} H_{\gamma\mu}(\mathbf{x})^{k_\gamma}$$

$$+ x_\mu \sum_{k_1,\ldots,k_d=0}^{2} \delta_{2,\sum_{\lambda=1}^{d} k_\lambda} q_{k_1\cdots k_d}^{\mu\nu} \prod_{\gamma=1}^{d} H_{\gamma\mu}(\mathbf{x})^{k_\gamma} + \cdots,$$

where δ_{ij} is the Kronecker delta. Reordering the terms, we find that $H_{\mu\nu}$ can be written as a function of $G_{\mu\nu}$ as follows:

$$H_{\mu\nu}(\mathbf{x}) = x_\mu \sum_{k_1=0}^{\infty} \cdots \sum_{k_d=0}^{\infty} q_{k_1\cdots k_d}^{\mu\nu} \prod_{\gamma=1}^{d} H_{\gamma\mu}(\mathbf{x})^{k_\gamma}$$

$$= x_\mu G_{\mu\nu}[H_{1\mu}(\mathbf{x}), \ldots, H_{d\mu}(\mathbf{x})]. \tag{4.4}$$

Here again, for a system of d networks, there are d^2 self-consistent equations of the form shown in Eq. 4.4.

Now instead of selecting an edge uniformly at random, consider a node chosen uniformly at random. This node is either isolated or has edges leading to other nodes in some subset of the d networks in the system. The probability argument above allows us to write a self-consistency equation for the distribution in component sizes to which a randomly selected node belongs:

$$H_\mu(\mathbf{x}) = x_\mu G_\mu[H_{1\mu}(\mathbf{x}), \ldots, H_{d\mu}(\mathbf{x})]. \tag{4.5}$$

With this relation for H_μ, we can now calculate the distribution of component sizes and the composition of the components in terms of nodes from various networks. However, our current interest is not in finding the exact probability distribution of the sizes of connected components, but in finding the emergence of large-scale connectivity in a system of interacting networks. To address this problem, we need only to examine the average component size to which a randomly chosen node belongs. For example, the average number of ν-nodes in the component of a randomly chosen μ-node is

$$\langle s_\mu \rangle_\nu = \frac{\partial}{\partial x_\nu} H_\mu(\mathbf{x})\bigg|_{\mathbf{x}=\mathbf{1}}$$

$$= \delta_{\mu\nu} G_\mu[H_{1\mu}(\mathbf{1}), \ldots, H_{d\mu}(\mathbf{1})]$$

$$+ \sum_{\lambda=1}^{d} G_\mu'^\lambda[H_{1\mu}(\mathbf{1}), \ldots, H_{d\mu}(\mathbf{1})]H_{\lambda\mu}'^\nu(\mathbf{1})$$

$$= \delta_{\mu\nu} + \sum_{\lambda=1}^{d} G_{\mu}^{'\lambda}(\mathbf{1}) H_{\lambda\mu}^{'\nu}(\mathbf{1}). \tag{4.6}$$

Table 4.1 shows the explicit algebraic expressions derived from Eq. 4.6 for a system of $d = 2$ networks with two different forms of internal degree distribution and types of coupling between networks. Where the algebraic expression for $\langle s_\mu \rangle_\nu$ diverges marks the percolation threshold for the onset of a giant component. For instance, the first case shown in Table 4.1 is for two networks, a and b, with internal Poisson distributions, coupled by a third Poisson distribution. For this situation, the percolation threshold is defined by the expression $(1 - \bar{k}_{aa})(1 - \bar{k}_{bb}) = \bar{k}_{ab}\bar{k}_{ba}$.

4.2.2 Consequences of Interactions

To quantify the consequences of interaction between distinct networks, we want to compare results obtained from the calculations above to a corresponding baseline model of a single, isolated network. Interesting differences already arise for the case of $d = 2$ interacting networks, which we focus on here. Consider two networks, a and b, with n_a and n_b nodes respectively. They have multi-degree distributions $p_{k_a k_b}^a$ and $p_{k_a k_b}^b$ respectively. The reference single network, $\mathcal{C}$, neglects the network membership of the nodes. It is of size $n_\mathcal{C} = n_a + n_b$ nodes, and has degree distribution

$$p_k = \left[f_a \sum_{k_a, k_b = 0}^{k} \left(p_{k_a k_b}^a \delta_{k_a + k_b, k} \right) + f_b \sum_{k_a, k_b = 0}^{k} \left(p_{k_a k_b}^b \delta_{k_a + k_b, k} \right) \right],$$

where $f_a = n_a/(n_a + n_b)$ and $f_b = n_b/(n_a + n_b)$. In other words, network $\mathcal{C}$ is a composite view that neglects whether a node belongs to network a or b. So a node that had degree $\{k_a, k_b\}$ in the interacting network view has degree $k = k_a + k_b$ in the composite, $\mathcal{C}$, view. We compare the properties of the ensemble of random graphs constructed from the interconnected networks multi-degree distribution, $\left\{ p_{k_a k_b}^a, p_{k_a k_b}^b \right\}$, to the properties of the ensemble constructed from the composite, p_k, degree distribution (Fig. 4.3).

In Ref. [39], we analyze the situation for two networks with distinct internal Poisson distributions coupled together via a third Poisson distribution. We show that large-scale connectivity can be achieved with fewer total edges if the network membership of the node is accounted for (i.e., the composite $\mathcal{C}$ view requires more edges to achieve a giant component).

Next we show that other effects are possible for different types of networks. For instance, the degree distributions that are a truncated power law describe many real-world networks, such as the connectivity between Autonomous Systems in the Internet and connectivity patterns in social contact networks [20]. Yet many critical infrastructure networks (such as adjacent buses in electric power grids) have very

Table 4.1 Expressions for average component size by node type for three different interacting network topologies

	Network topology			Average node count by type and initial network
a-a	*a-b*	*b-a*	*b-b*	
	Distribution parameters			$\langle s_a \rangle_a$
	Generating functions			$\langle s_a \rangle_b$
				$\langle s_b \rangle_a$
				$\langle s_b \rangle_b$

Poisson row:

a-a	*a-b*	*b-a*	*b-b*
Poisson	Poisson	Poisson	Poisson
$\bar{k}_{aa}$	$\bar{k}_{ab}$	$\bar{k}_{ba}$	$\bar{k}_{bb}$

$$G_a(x_a,b) = e^{\bar{k}_{aa}(x_a-1)}\, e^{\bar{k}_{ab}(x_b-1)}$$

$$G_b(x_a,b) = e^{\bar{k}_{ba}(x_a-1)}\, e^{\bar{k}_{bb}(x_b-1)}$$

$$\langle s_a \rangle_a = 1 + \frac{\bar{k}_{aa}[1-\bar{k}_{bb}] + \bar{k}_{ab}\bar{k}_{ba}}{(1-\bar{k}_{aa})(1-\bar{k}_{bb}) - \bar{k}_{ab}\bar{k}_{ba}}$$

$$\langle s_a \rangle_b = \frac{\bar{k}_{ab}}{(1-\bar{k}_{aa})(1-\bar{k}_{bb}) - \bar{k}_{ab}\bar{k}_{ba}}$$

$$\langle s_b \rangle_a = \frac{\bar{k}_{ba}}{(1-\bar{k}_{aa})(1-\bar{k}_{bb}) - \bar{k}_{ab}\bar{k}_{ba}}$$

$$\langle s_b \rangle_b = 1 + \frac{\bar{k}_{bb}[1-\bar{k}_{aa}] + \bar{k}_{ab}\bar{k}_{ba}}{(1-\bar{k}_{aa})(1-\bar{k}_{bb}) - \bar{k}_{ab}\bar{k}_{ba}}$$

Power-law row:

a-a	*a-b*	*b-a*	*b-b*
Power-law	Poisson	Poisson	Poisson
τ_a, κ_a	$\bar{k}_{ab}$	$\bar{k}_{ba}$	$\bar{k}_{bb}$

$$G_a(x_a,b) = \frac{\mathrm{Li}_{\tau_a}(x_a e^{-1/\kappa_a})}{\mathrm{Li}_{\tau_a}(e^{-1/\kappa_a})}\, e^{\bar{k}_{ba}(x_b-1)}$$

$$G_b(x_a,b) = e^{\bar{k}_{ba}(x_a-1)}\, e^{\bar{k}_{bb}(x_b-1)}$$

$$\langle s_a \rangle_a = 1 + \frac{\bar{k}_{aa}[1-\bar{k}_{bb}] + \bar{k}_{ab}\bar{k}_{ba}[1 - G'^{\alpha}_{\alpha\alpha}(1) + G'^{\alpha}_{\alpha\beta}(1)]}{[1 - G'^{\alpha}_{\alpha\alpha}(1)][1-\bar{k}_{bb}] - \bar{k}_{ab}\bar{k}_{ba}[1 - G'^{\alpha}_{\alpha\alpha}(1) + G'^{\alpha}_{\alpha\beta}(1)]}$$

$$\langle s_a \rangle_b = \frac{\bar{k}_{ab}[1 - G'^{\alpha}_{\alpha\alpha}(1) + G'^{\alpha}_{\alpha\beta}(1)]}{[1 - G'^{\alpha}_{\alpha\alpha}(1)][1-\bar{k}_{bb}] - \bar{k}_{ab}\bar{k}_{ba}[1 - G'^{\alpha}_{\alpha\alpha}(1) + G'^{\alpha}_{\alpha\beta}(1)]}$$

$$\langle s_b \rangle_a = \frac{\bar{k}_{ba}[1 + \bar{k}_{aa} - G'^{\alpha}_{\alpha\alpha}(1)]}{[1 - G'^{\alpha}_{\alpha\alpha}(1)][1-\bar{k}_{bb}] - \bar{k}_{ab}\bar{k}_{ba}[1 - G'^{\alpha}_{\alpha\alpha}(1)]}$$

$$\langle s_b \rangle_b = 1 + \frac{\bar{k}_{bb}[1 - G'^{\alpha}_{\alpha\alpha}(1)] + \bar{k}_{ab}\bar{k}_{ba}[1 - G'^{\alpha}_{\alpha\alpha}(1) - G'^{\alpha}_{\alpha\beta}(1)]}{[1 - G'^{\alpha}_{\alpha\alpha}(1)][1-\bar{k}_{bb}] - \bar{k}_{ab}\bar{k}_{ba}[1 - G'^{\alpha}_{\alpha\alpha}(1) + G'^{\alpha}_{\alpha\beta}(1)]}$$

Fig. 4.3 Comparing random graph models which account for interacting networks (*red line*) to random graph models with the identical degree distribution, but which neglect network membership (*dashed black line*). **a** The fraction of nodes in the largest connected component for two interconnected networks with Poisson degree distribution, as edges are added to network b. Accounting for network structure allows for a giant component to emerge with fewer edges. Here $n_a = 4n_b$. **b** The corresponding fractional size of the giant component for a network with a Poisson degree distribution coupled to a network with a truncated power law degree distribution as the power law regime is extended. Here was see the opposite effect to **a**, where large scale connectivity is delayed by accounting for network membership

narrow degree distributions, which we approximate here as Poisson. Thus, we are interested in the consequences of coupling together networks with these different types of distributions. Let network a have an internal distribution described by a truncated power law, $p_{k_a}^a \propto k_a^{-\tau_a} \exp(-k/\kappa_a)$, and network b have an internal Poisson distribution. Coupling these networks via a distinct Poisson distribution is described by the second case shown in Table 4.1. Here, the composite $\mathcal{C}$ view requires fewer edges to achieve a giant component, so large-scale connectivity requires more edges if the network membership of the nodes is accounted for. The effects in shifting the percolation transition can be amplified if the networks are of distinct size, $n_a \neq n_b$. For more details on these percolation properties of interconnected networks, see Refs. [39, 40]. Also, see Ref. [38] for a discussion of how correlations in *multiplex* networks can alter percolation properties.

4.3 Application: Sandpile Cascades on Interconnected Networks

Equipped with a random graph model of interconnected networks and an understanding of its percolation properties, we now use this framework to analyze systemic risk by studying a dynamical process occurring on such interconnected networks. Here we seek a model that captures risk of widespread failure in critical infrastructures.

4.3.1 The Sandpile Model as a Stylization of Cascading Failure in Infrastructure

A common feature of many infrastructures is that their elements hold load of some kind, and they can only hold a certain amount of it. For example, transmission lines of power grids can carry only so much electricity before they trip and no longer carry electricity [18]; banks can withstand only so much debt without defaulting [30]; hospitals can hold only so many patients; airports can accommodate only so many passengers per day. When a transmission line, bank, hospital or airport partially or completely fails, then some or all of its load (electricity, debt, patients or travelers) may burden another part of that network or a completely different kind of network. For instance, when a transmission line fails, electricity quickly reroutes throughout the power grid (the same network), whereas when an airport closes due to a catastrophe like a volcano eruption [31] travelers may overwhelm railway and other transportation networks.

In addition to loads and thresholds, another commonality among certain risks of failure in infrastructure are heavy-tailed probability distributions of event size. In electric power systems, for instance, the amount of energy unserved during 18 years of North American blackouts resembles a power law over four orders of magnitude, and similarly broad distributions are found in other measures of blackout size [18]. In financial markets, stock prices and trading volume show power law behavior, in some cases with exponents common to multiple markets [22, 26]. In interbank credit networks, most shocks to banks result in small repercussions, but the 2008 financial crisis demonstrates that large crises continue to occur. Similarly broad distributions of event sizes also occur in natural systems such as earthquakes [64], landslides [32] and forest fires [45, 65]. Some evidence suggests that engineered systems like electric power grids [18] and and financial markets [22], not to mention natural catastrophes like earthquakes [64], landslides [32] and forest fires [45, 65], all show heavy-tailed event size distributions because they self-organize to a critical point.

An archetypal model that captures these two features—of units with capacity for load and of heavy-tailed event size distributions—is the Bak-Tang-Wiesenfeld (BTW) sandpile model [5, 6]. This model considers a network of elements that hold load (grains of sand) and that shed their load to their neighbors when their load exceeds their capacity. Interestingly, one overloaded unit can cause a cascade (or avalanche) of load to be shed, and these cascades occur in sizes and durations distributed according to power laws. This deliberately simplified model ignores detailed features of real systems, but its simplicity allows comprehensive study that can in turn generate hypotheses to test in more realistic, detailed models, which we will discuss in Sect. 4.3.4.

4.3.2 Defining the Sandpile Model on Networks

First studied on a two-dimensional lattice [5, 6], the BTW model has recently been studied on random graphs [11, 13, 17, 27, 28, 35–37, 41, 53], in part because many critical infrastructure like power, water, transportation and finance have network structure. There are different ways to implement the sandpile model on a network, but these implementations only differ in specifics. Here we study the following natural formulation [27, 28, 36, 37].

Each node holds grains of sand, which we interpret as units of load. Nodes can hold only a certain number of grains. When the number of grains equals or exceeds the node's threshold, then the node topples and moves sand to its neighbors. A natural choice for the threshold of a node is its degree, so that when a node topples it sends one grain to each of its neighbors. Other thresholds have been studied [27, 36], but these other rules for the threshold require nodes to shed sand to (for example) a random subset of their neighbors.

The BTW sandpile model consists of a sequence of cascades (avalanches), defined as follows. First, drop a grain of sand on a node chosen uniformly at random. If the node's number of grains is greater than or equal to its threshold (i.e., its degree), then that node is considered overwhelmed or unstable, and that node sheds (moves) all its load to its neighbors by sending one grain to each neighbor. These neighbors may in turn exceed their thresholds and have to topple, and subsequently their neighbors may topple, and so on. Once no node exceeds its threshold, we record the number of nodes that toppled (the cascade size), and the process begins again by dropping a grain on a random node.

In order to prevent the system from becoming inundated with sand, grains of sand must somehow be removed. Following [28], we choose the following rule for dissipation of sand: whenever a grain of sand is sent from one node to another node, with some small, fixed probability that grain is removed from the system.

The quantities of interest are measured in the dynamical equilibrium state that the system reaches after many cascades have occurred, because the system self-organizes to a critical point. Specifically, if the network begins without any sand, then sand slowly builds up in the system. After a large number of cascades (e.g., an order of magnitude more than the number of nodes), the system appears to reach a dynamical equilibrium at which the amount of sand does not change significantly relative to the system size. On one hand, large cascades tend to destroy lots of sand because of the rule for dissipating sand described above. On the other hand, when the amount of sand in the system is low, then cascades are typically smaller, so the amount of sand tends to increase. These effects balance so that the system seeks a critical point at which the probability distribution of cascade size (and of other measures like cascade duration) show power law behavior [28, 53]. These power law–distributed cascades can therefore serve as a useful, stylized model for risk of large cascading failures in infrastructures.

4.3.3 Results for the Sandpile Model on Interconnected Networks: Optimal Interconnectivity, the Yellowstone Effect, and Systemic Risk

In this subsection, we highlight three results from Ref. [13]. Next, in Sect. 4.3.4, we comment on current and future work to understand the sandpile model on isolated and interconnected networks, as well as on work to understand risk in interdependent infrastructures and other examples of optimal, intermediate amounts of connectivity.

We begin by studying one of the simplest interconnected networks, two random 3-regular graphs a and b with edges introduced between them. Specifically, each node in network a (b) has 3 neighbors in network a (b, respectively). Networks a and b have identical number of nodes. Next, a fraction p of nodes in a have one edge to a neighbor in the other network. (In the notation of the join degree distributions in Sect. 4.2.1, the degree distribution of network a is $p_{k_a,k_b}^a = \delta_{k_a,3}\left[p\delta_{k_b,1} + (1-p)\delta_{k_b,0}\right]$, and vice versa for network b.) This "interconnectivity" parameter $p \in [0, 1]$ measures the coupling between the two networks. The threshold of each node is its total degree.

One motivating example for this choice of interconnected networks are power grids. The degree of a typical node in the transmission grid (the part of a power grid that moves electric energy at high voltage) is approximately 3 [13], so we chose to study random 3-regular graphs. (Using 3-regular graphs rather than, say, Erdős-Rényi random graphs, simplifies the degree distribution to delta functions and hence simplifies branching process approximations of cascades [13].) Moreover, power grids have modular structure because they consist of "regions" or "control areas". Historically, each region was its own grid, and then these grids began connecting with one another, so that now one grid can span an entire continent. Each region of the grid is typically more densely connected within the region than with other regions. Furthermore, this modular structure is not static: grids continue to build new transmission lines between regions in order to, for example, accommodate wind power [34]. Increasing the interconnectivity p in our model vaguely captures the construction of new transmission lines between regions of a power grid.

Other infrastructures, from interbank credit networks [46] to transportation [48], exhibit modular structure at different scales. In some cases, these modules are becoming more interconnected over time, as lending, travel and trade become more global. Understanding how this increase in connectivity affects systemic risk is a problem that transcends disciplines. Though the sandpile model does not capture any one of these infrastructures accurately, it self-organizes to a critical point at which cascades occur in sizes described by power laws, and this behavior vaguely resembles large fluctuations in many engineered and natural systems. Thus, the sandpile model can be useful for generating hypotheses to test in more realistic models. Next we highlight three such hypotheses.
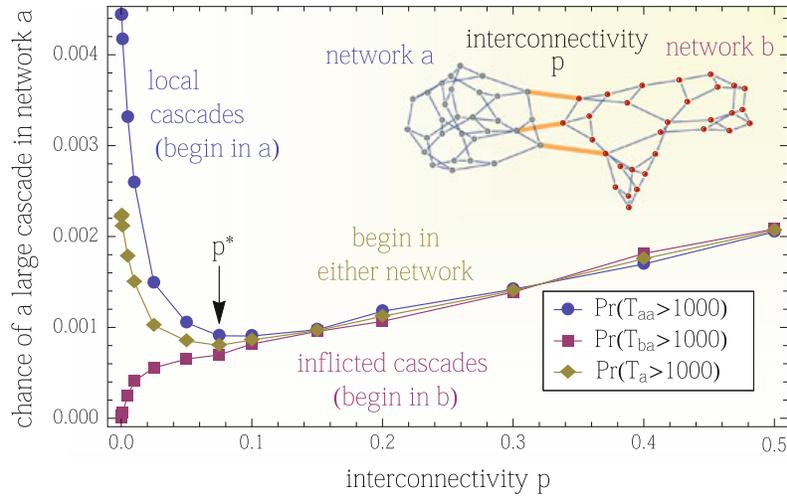
Fig. 4.4 The chance that a network a connected to another network b suffers a cascade larger than half its network [gold curve, $\Pr(T_a > 1000)$] has a minimum at a critical amount of interconnectivity p^*. Networks that want to mitigate their largest cascades would prefer to build or demolish interconnections to operate at this critical point p^*. The *blue* (*red*) *curve* is the chance $\Pr(T_{aa} > 1000)$ [$\Pr(T_{ba} > 1000)$] that a cascade that begins in a (b) topples at least 1000 nodes in a. Increasing interconnectivity only exacerbates the cascades inflicted from b to a (*red*), but interestingly it initially suppresses the local cascades in a (For each p, we run a simulation on one realization of two random 3-regular graphs with 2000 nodes each; each node has a neighbor in the other network with probability p. The dissipation parameter is 0.01, the amount that makes the largest cascades slightly smaller than the size of the system. The *inset* depicts a small example with 30 nodes per network and $p = 0.1$.)

4.3.3.1 Optimal Interconnectivity

Suppose each network a, b is a region of a power grid and that each region is owned by a different utility. (To reiterate, the sandpile model misses crucial features of power grids, described below in Sect. 4.3.4.1, but we use the power grid example to facilitate interpretation of results.) If each network (think "each utility in the power grid") a, b wants to reduce the risk of cascading failure in its own network, then how many interconnections (edges between the networks) would they want?

Figure 4.4 shows the striking result that each network a, b would want to build some interconnections but not too many. Specifically, define a large cascade in a network as a cascade that topples at least half of the network. In Fig. 4.4, a has 2000 nodes, so a large cascade in a is one that causes at least 1000 toppling events in a. (The results are rather insensitive to changes in this cutoff for calling cascades large; see [13, Fig. 4]. Also, Sect. 4.3.3.2 explores the risk of small cascades.) The chance of a large cascade in a network is a measure of that network's risk. The gold curve of Fig. 4.4 shows that a network's risk decreases and then increases with the interconnectivity p, with the minimum occurring at an intermediate interconnectivity p^*. Thus, two initially isolated networks would want to build interconnections up to p^* in order to reduce their own risk of large cascades.

The 70 % drop in the risk of either network due to increasing interconnectivity p from 0.001 to $p^* = 0.075 \pm 0.01$ is significant. If these cascades were blackouts, then utility a (say) would experience 70 % fewer large blackouts. Why? By building $p^* N_a = 150 \pm 20$ edges (transmission lines) with its neighboring network b, the networks can collectively share their risk of large blackouts.

To further illustrate this "optimal interconnectivity" p^*, we distinguish cascades that begin in network a (the blue curve labeled "local cascades" in Fig. 4.4) from cascades that begin in network b (the red curve labeled "inflicted cascades"). As interconnectivity p increases, the chance of a large inflicted cascade increases monotonically, because building interconnections opens new avenues for large cascades to spread to the other network.

More interestingly, building some interconnections (but not too many) suppresses local cascades. That is, when interconnectivity p is small, the more edges a has with b, the lower the chance that a cascade begun in a topples a number of nodes in a greater than half the size of a. One reason for this suppression of local cascades is that nodes with an edge to the other network have larger threshold (because their degree is 4 rather than 3), so they topple less often when they receive sand. (However, the repercussions of toppling a degree-4 node are worse because they hold more sand.) Another reason that some interconnectivity suppresses local cascades is that more interconnections make the cascades less confined to one network and instead become more spread out among the two networks (see [13, Fig. S10]). This phenomenon of sharing risk resembles the tenet of diversification in investment portfolios in finance [2, 9].

Before proceeding, we note a similarity between optimal interconnectivity and equilibria in economics. Just as rational agents seek more of something as long as the marginal benefits exceed the marginal costs, a network would seek more interconnectivity as long as the marginal benefits exceed the marginal costs. In the sandpile model, building interconnections confers more benefits than costs initially, where benefits are reduction in risk of large cascades. In a competitive market, consumers and firms converge on the optimal price p^* at which the marginal benefit of the last unit consumed equals the marginal cost. Analogously, two networks seeking to mitigate their risk of large cascades converge on the optimal interconnectivity p^* at which the marginal benefits of the last edge built equal the marginal cost. More realistic models of connections within and between networks would also incorporate the costs of building and maintaining a new link, and this cost would presumably change the optimal number of links p^*.

Perhaps many interconnected networks are what Nassim Taleb calls "antifragile", meaning that they become more robust against large-scale catastrophes [66] if they have some variability [67] from input from external networks (e.g., interconnectivity $p^* > 0$).
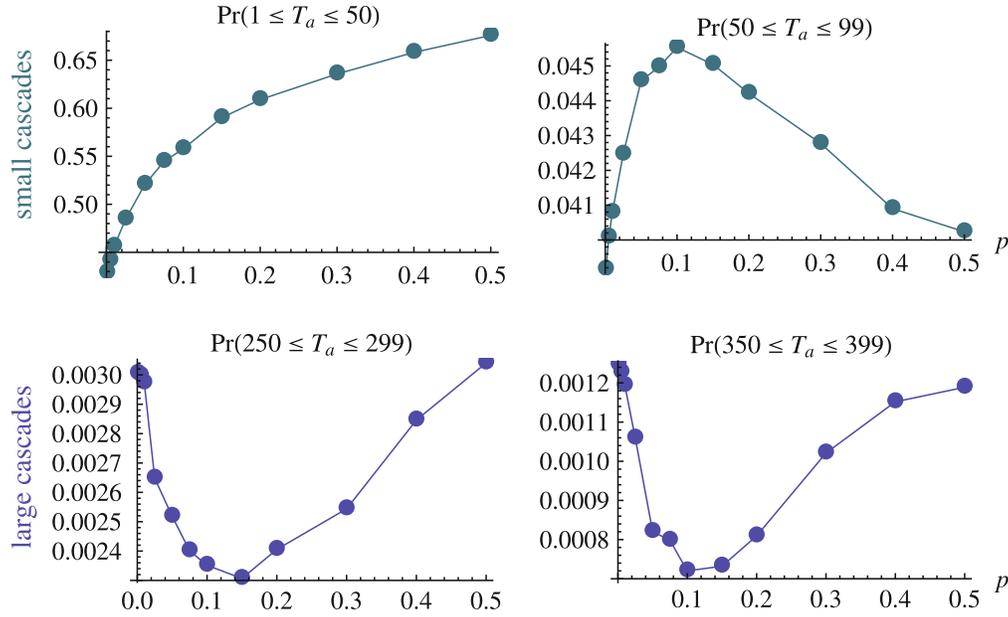
Fig. 4.5 a Networks mitigating the smallest cascades of size $T_a \in [1, 50]$ seek isolation $p = 0$, while **b** networks suppressing intermediate cascades $T_a \in [50, 100)$ seek isolation $p = 0$ or strong coupling $p = 1$, depending on the initial interconnectivity p in relation to the unstable critical point $p^* \approx 0.12 \pm 0.02$. But networks like power grids that mitigate large cascades **c**, **d** would seek interconnectivity at the stable equilibrium $p^* \approx 0.12 \pm 0.02$. The qualitative shape of the plots in the *bottom* figures and the location of p^* are robust to changes in the window $\ell \leq T_a \leq \ell + 50$ for all $200 \leq \ell \leq 800$ (Here we show results from simulations on two random 3-regular graphs with 1000 nodes each, which is half the network size as in Fig. 4.4, to show how p^* decreases with system size.)

4.3.3.2 Yellowstone Effect: Why Suppressing Small Cascades Can Increase the Risk of Large Ones

Rather than seeking to mitigate their risk of large cascades (and hence seeking interconnectivity p^*), what if the two networks a, b seek to mitigate their risk of small cascades? Figure 4.5 shows that the risk of small cascades increases monotonically with interconnectivity p. Thus, $p = 0$ minimizes the risk of small cascades.

However, by Fig. 4.4, $p = 0$ is a local maximum in the risk to each network. Thus, by seeking $p = 0$ to mitigate their own small cascades, networks a, b would increase their risks of large cascades. The same phenomenon is thought to occur in suppressing blackouts [18] and forest fires [45]. In fact, this phenomenon has been given the name the "Yellowstone effect" because suppressing small forest fires in Yellowstone National Park, WY, in the twentieth century densified forest vegetation and hence increased the risk of massive forest fires [45]. What Fig. 4.5 demonstrates is that interconnectivity is another mechanism that can cause the Yellowstone effect. This result suggests that we should look for similar phenomena in more realistic models.
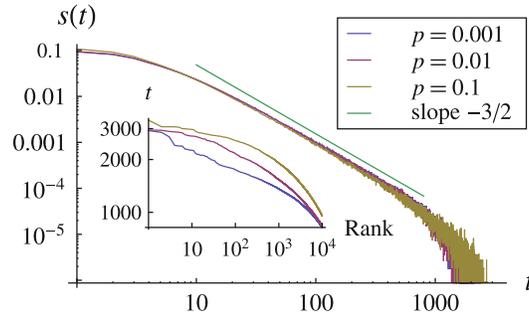
Fig. 4.6 Increasing the interconnectivity p between two random 3-regular graphs extends the tail of the total avalanche size distribution $s(t)$, which does not distinguish whether toppled nodes are in network a or b. The *inset* shows a rank-size plot on log-log scales of the number of topplings t in the largest 10^4 avalanches (with 2×10^6 grains of sand dropped), showing that adding more edges between random 3-regular graphs enlarges the largest global cascades by an amount on the order of the additional number of interconnections. As expected theoretically [28], when a and b nodes are viewed as one network, $s(t) \sim t^{-3/2}$ for large t (*green line*)

4.3.3.3 Risk-Avoiding Networks Can Exacerbate System-Wide Risk

If two networks act in a greedy, rational manner to mitigate their own risk of large cascades, without regard to the risk of the other network, then by Sect. 4.3.3.1 each network would seek the optimal, intermediate amount of interconnectivity p^*. What is the effect of this self-interested behavior on the system as a whole?

Figure 4.6 shows that every increase in interconnectivity p tends to increase the size of the largest cascades in the whole system (where the size of the cascade no longer distinguishes types of nodes). The main plot of Fig. 4.6 is the cascade size distribution $s(t)$, which is the probability of observing a cascade with t many toppling events (in the equilibrium state of the system after many cascades have been run without collecting statistics), for interconnectivity $p = 10^{-3}, 10^{-2}, 10^{-1}$. (As expected [28], the avalanche size distribution shows power law behavior with exponent $-3/2$ over at least two orders of magnitude, and more detailed theoretical arguments confirm this conclusion [53].)

To illustrate the tail behavior of the cascade size distribution $s(t)$, the inset of Fig. 4.6 shows a rank-size plot of the largest cascades in the whole system. This plot shows that, as p increases, global cascades become larger by an amount on the order of the additional number of interconnections. Because each interconnection confers an addition to the threshold of nodes and hence to the capacity of the system to hold sand, the system holds more sand in the dynamic equilibrium state, so the largest cascades can involve more sand. Similar phenomena occur in transportation systems and electric power grids. Building a new bridge to Manhattan, for example, can paradoxically worsen traffic because people use the new capacity (the so-called Braess' Paradox [12]). Similarly, the benefits of upgrades to power grids are often fleeting, because operators seek to efficiently use their costly infrastructure [18].

4.3.4 Optimal Connectivity in Models Other Than the Sandpile Model

Because the sandpile model self-organizes to a critical point, it has relatively few parameters, so its behavior can be explored somewhat comprehensively. By contrast, detailed models of real systems can have so many parameters that they are difficult to comprehend, and many parameters are difficult to measure in the real world. Thus, the interesting behavior of the sandpile model—such as power law distributions of cascade size, optimal interconnectivity, the Yellowstone effect and system-wide risk—can serve as hypotheses to test and refine in more realistic models.

Next we review recent work in calculating optimal interdependence and optimal connectivity in power grids, bank networks and social systems.

4.3.4.1 Cascading Failures in Power Grids: Why Topological Models do not Suffice, and What Details are Needed

One of the promises of the burgeoning field of complex networks is to simplify the complicated behavior of real systems. Unfortunately, power grids are one example for which naïve, topological network approaches do not appear to suffice [14, 33].

Furthermore, most of these topological models, like the sandpile model, treat a cascading failure like an epidemic that spreads between adjacent nodes. By contrast, failures in the power grid spread non-locally: when a node in a power grid (such as a bus or a substation) fails or, more commonly, an edge (a transmission line) trips, the electric power re-routes almost instantly to all parallel paths, inversely proportionally to the impedances on lines. Thus, a failure can trigger other failures hundreds of kilometers away [24]. Models that lack this non-locality (or that have non-local failures but via completely different mechanisms) offer little insight into cascading failures in power grids [33].

What then is needed to offer insight for blackouts in power grids? A first step and essential ingredient are the linearized direct current (DC) equations, an approximation of the physics of electric power. These equations require data on the power injections at every node (which is positive for generators, negative for load buses) and the impedances and capacities of lines. Thus, the topological structure of a power grid is insufficient to run the DC power flow equations; these "thin" networks need to be augmented with data on the buses and transmission lines. (The MATLAB software MATPOWER [72] provides a useful starting point because it contains data on the Polish power grid. Another approach is to generate statistically correct power grid topologies [69].)

Equipped with a physics-based model of electric power flow in a grid, one must choose what failure mechanisms to model. Unfortunately, the number of failure mechanisms is large; examples include thermal overloads, relay failure, voltage collapse, dynamic instability and operator error [23]. The state-of-the-art is to model a subset of these failure mechanisms (see, e.g., [7, 23, 50]).

Such a detailed, physics-based, data-driven model of cascading failures might find rather rich pictures of optimal interconnectivity between regions of a power grid. The model space would likely be much richer than that of the sandpile model. But solving this problem once is not enough because modern power grids are changing. For instance, rapid deployment of smart grid technologies enable greater control and measurement of the grid. Renewable energy will stress the grid in new ways, as generation becomes more intermittent and increasingly located in sunny, windy locations, thereby changing the import of power between regions. These changes to the grid make studies of optimal grid structure all the more timely and important.

4.3.4.2 Optimal Interconnectivity in Bank Networks, Coupled Transportation Networks and Social Systems

The notion of optimal, intermediate amounts of connectivity is not new. For example, Battiston et al. [9] found that a network of banks is most resilient to cascading default if banks have an intermediate amount of risk diversification. What made this result novel was its contrast with the common wisdom in the financial literature that more diversification is always better [2]. In another model of bank default, if banks lend to an intermediate number of other banks, then the banks can be the most fragile and still not suffer cascading default [8].

Optimal coupling has also been found in a model of transportation on coupled spatial networks [48]. If a transportation administrator wishes to minimize both the average travel time and the amount of congestion, then a nontrivial, optimal "coupling" between, say, a rail network and a road network can emerge. Like in the sandpile model on interconnected graphs [13], two competing forces (efficiency and congestion in the transportation model) can balance at optimal amounts of coupling.

Optimal connectivity has also been found in strategic games played on networks. For example, a social network playing the minority game is most efficient at an intermediate amount of connectivity [4, 44]. Optimal interconnectivity between two networks has been found in the public goods game, where the interconnectivity $p = 1/2$ maximizes cooperation [68].

These results in financial and social models suggest that optimal connectivity (and interconnectivity) may be common in networks. If the dynamics occurring on some network have opposing tradeoffs, then these tradeoffs may balance at critical points. Whether the corresponding real systems—such as power grids, bank networks or social networks—can sense and respond to these tradeoffs and hence operate at the optimal points remains an open question. The answers would likely be essential to any effort to control such systems [43].

4.4 Conclusion and Future Work

Why might two networks build connections between them? This chapter demonstrates two reasons: to efficiently provide global connectivity and to reduce the risk that either system suffers a large cascading failure.

This work belongs to a movement to study mesosopic, intermediate-scale structure in networks, and not just global structure (like degree distributions) and microscopic structure (like clustering coefficients). Two prominent examples of mesoscopic structure in networks are community (or modular) structure and core-periphery. There exist many tools for finding community structure in networks (see the reviews [25, 57]) and comparatively fewer tools for finding core-periphery structure [62]. But we are only just beginning to learn about the effect of this mesoscopic structure on the system's percolation properties (Sect. 4.2, Ref. [40]) and on dynamics occurring on the network (Sect. 4.3, Ref. [13]).

Another challenge is to study the converse: how the dynamics on the network affect its mesoscopic structure. In the sandpile model on interconnected networks [13], large cascades in one network may convince it to build more interconnections and hence to change the mesoscopic structure. Similarly, in power grids, large blackouts can provoke upgrades to the grid, which can include new transmission lines that change the structure of the grid. Large financial crises alter web of financial interactions among banks [46]. Widespread defection in a social network may alter its social ties. Thus, rare, catastrophic events [66] may be a sign of a network in the throes of its path toward optimal connectivity, if one exists.

References

1. *Review of Selected 1996 Electric System Disturbances in North America*.North American Electric Reliability Council, 2002.
2. F. Allen and D. Gale.Financial contagion.*Journal of Political Economy*, 108(1):1–33, 2000.
3. M. Amin.National infrastructure as complex interactive networks.In T. Samad and J. Weyrauch, editors, *Automation, control and complexity: an integrated approach*, pages 263–286. John Wiley & Sons,Inc., 2000.
4. M. Anghel, Z. Toroczkai, K. Bassler, and G. Korniss. Competition-Driven Network Dynamics: Emergence of a Scale-Free Leadership Structure and Collective Efficiency. *Physical Review Letters*, 92(5):058701, Feb. 2004.
5. P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality: An explanation of $1/f$ noise. *Physical Review Letters*, 59(4):381–384, 1987.
6. P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality. *Physical Review A*, 38(1):364–374, 1988.
7. R. Baldick, B. Chowdhury, I. Dobson, Z. Dong, B. Gou, D. Hawkins, H. Huang, M. Joung, D. Kirschen, F. Li, J. Li, Z. Li, C.-C. Liu, L. Mili, S. Miller, R. Podmore, K. Schneider, K. Sun, D. Wang, Z. Wu, P. Zhang, W. Zhang, and X. Zhang. Initial review of methods for cascading failure analysis in electric power transmission systems ieee pes cams task force on understanding, prediction, mitigation and restoration of cascading failures. In *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, pages 1–8, July 2008.

8. S. Battiston, D. D. Gatti, M. Gallegati, B. Greenwald, and J. E. Stiglitz. Default cascades: When does risk diversification increase stability? *Journal of Financial Stability*, 8(3):138–149, Sept. 2012.

9. S. Battiston, D. D. Gatti, M. Gallegati, B. Greenwald, and J. E. Stiglitz. Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of Economic Dynamics and Control*, 36(8):1121–1141, Aug. 2012.

10. B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1:311, 1980.

11. E. Bonabeau. Sandpile dynamics on random graphs. *Journal of the Physical Society of Japan*, 64(1):327–328, 1995.

12. D. Braess, A. Nagurney, and T. Wakolbinger. On a Paradox of Traffic Planning. *Transportation Science*, 39(4):446–450, Nov. 2005.

13. C. D. Brummitt, R. M. D'Souza, and E. A. Leicht. Suppressing cascades of load in interdependent networks. *Proc. Natl. Acad. Sci. U.S.A.*, 109(12):E680–E689, Feb. 2012.

14. C. D. Brummitt, P. D. H. Hines, I. Dobson, C. Moore, and R. M. D'Souza. A transdisciplinary science for 21st-century electric power grids. Forthcoming, 2013.

15. S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464:1025–1028, 2010.

16. M. Chediak and L. M. Cold snap causes gas shortages across u.s. southwest. *Bloomberg News*, (http://www.bloomberg.com/news/2011-02-04/cold-snap-causes-gas-shortages-across-u-s-southwest.html), Feb, 2011

17. J. de Arcangelis and H. J. Herrmann. Self-organized criticality on small world networks. *Physica A*, 308:545–549, 2002.

18. I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman. Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos*, 17(026103), 2007.

19. I. Dobson, B. A. Carreras, and D. E. Newman. A branching process approximation to cascading load-dependent system failure. In *Thirty-seventh Hawaii International Conference on System Sciences*, 2004.

20. R. M. D'Souza, C. Borgs, J. T. Chayes, N. Berger, and R. D. Kleinberg. Emergence of tempered preferential attachment from optimization. *Proc. Natn. Acad. Sci. USA*, 104(15):6112–6117, 2007.

21. L. Dueñas-Osorio and S. M. Vemuru. Cascading failures in complex infrastructures. *Structural Safety*, 31(2):157–167, 2009.

22. B. Dupoyet, H. R. Fiebig, and D. P. Musgrove. Replicating financial market dynamics with a simple self-organized critical lattice model. *Physica A*, 390(18–19):3120–3135, Sept. 2011.

23. M. J. Eppstein and P. D. H. Hines. A "Random Chemistry" Algorithm for Identifying Collections of Multiple Contingencies That Initiate Cascading Failure. *Power Systems, IEEE Transactions on*, 27(3):1698–1705, 2012.

24. Federal Energy Regulatory Commission. *Arizona-Southern California Outages on September 8, 2011*, Apr. 2012.

25. S. Fortunato. Community detection in graphs. *Physics Reports*, 486:75–174, 2010.

26. X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423(6937):267–270, 2003.

27. K. Goh, D. Lee, B. Kahng, and D. Kim. Cascading toppling dynamics on scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 346(1–2):93–103, 2005.

28. K.-I. Goh, D.-S. Lee, B. Kahng, and D. Kim. Sandpile on Scale-Free Networks. *Physical Review Letters*, 91(14):148701, Oct. 2003.

29. T. H. Grubesic and A. Murray. Vital nodes, interconnected infrastructures and the geographies of network survivability. *Annals of the Association of American Geographers*, 96(1):64–83, 2006.

30. A. G. Haldane and R. M. May. Systemic risk in banking ecosystems. *Nature*, 469:351–355, Jan 2011.

31. M. Hanlon. How we could all be victims of the volcano... and why we must hope for rain to get rid of the ash. *Daily Mail*, April 2010. http://www.dailymail.co.uk/news/article-1267111/

32. S. Hergarten. Landslides, sandpiles, and self-organized criticality. *Natural Hazards and Earth System Sciences*, 3:505–514, 2003.
33. P. Hines, E. Cotilla-Sanchez, and S. Blumsack. Do topological models provide good information about electricity infrastructure vulnerability? *Chaos*, 20(033122), Jan 2010.
34. C. Joyce. Building power lines creates a web of problems. *NPR*, April 2009. http://www.npr.org/templates/story/story.php?storyId=103537250
35. J. Lahtinen, J. Kertész, and K. Kaski. Sandpiles on Watts-Strogatz type small-worlds. *Physica A*, 349:535–547, 2005.
36. D. Lee, K. Goh, B. Kahng, and D. Kim. Sandpile avalanche dynamics on scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 338(1–2):84–91, 2004.
37. K.-M. Lee, K.-I. Goh, and I. M. Kim. Sandpiles on multiplex networks. *Journal of the Korean Physical Society*, 60(4):641–647, Feb. 2012.
38. K.-M. Lee, J. Y. Kim, W.-K. Cho, K. Goh, and I. Kim. Correlated multiplexity and connectivity of multiplex random networks. *New Journal of Physics*, 14(3):033027, 2012.
39. E. A. Leicht and R. M. D'Souza. Random graph models of interconnected networks. *Forthcoming*.
40. E. A. Leicht and R. M. D'Souza. Percolation on interacting, networks. arXiv:0907.0894, July 2009.
41. S. Lise and M. Paczuski. Nonconservative earthquake model of self-organized criticality on a random graph. *Physical Review Letters*, 88:228301, 2002.
42. R. G. Little. Controlling cascading failure: Understanding the vulnerabilities of interconnected infrastructures. *Journal of Urban Technology*, 9(1):109–123, 2002.
43. Y.-Y. Liu, J.-J. Slotine, and A.-Á. Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, May 2011.
44. T. Lo, K. Chan, P. Hui, and N. Johnson. Theory of enhanced performance emerging in a sparsely connected competitive population. *Physical Review E*, 71(5):050101, May 2005.
45. B. D. Malamud, G. Morein, and D. L. Turcotte. Forest Fires: An Example of Self-Organized Critical Behavior. *Science*, 281(5384):1840–1842, sep 1998.
46. C. Minoiu and J. Reyes. A Network Analysis of Global Banking: 1978–2010. *Journal of Financial Stability*, 2013. in press.
47. M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6:161–180, 1995.
48. R. G. Morris and M. Barthelemy. Transport on Coupled Spatial Networks. *Physical Review Letters*, 109(12):128703, Sept. 2012.
49. P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.
50. D. P. Nedic, I. Dobson, D. S. Kirschen, B. A. Carreras, and V. E. Lynch. Criticality in a cascading failure blackout model. *International Journal of Electrical Power & Energy Systems*, 28(9):627–633, 2006.
51. D. E. Newman, B. Nkei, B. A. Carreras, I. Dobson, V. E. Lynch, and P. Gradney. Risk assessment in complex interacting infrastructure systems. In *Thirty-eight Hawaii International Conference on System Sciences*, 2005.
52. M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64(2):026118, 2001.
53. P.-A. Noël, C. D. Brummitt, and R. M. D'Souza. Controlling self-organizing dynamics using self-organizing models. Forthcoming, 2013.
54. W. of the ETH Risk Center. New Views on Extreme Events: Coupled Networks, Dragon Kings and Explosive Percolation. October 25–26th, 2012.
55. S. Panzieri and R. Setola. Failures propagation in critical interdependent infrastructures. *Int. J. Modelling, Identification and, Control*, 3(1):69–78, 2008.
56. P. Pederson, D. Dudenhoeffer, S. Hartley, and M. Permann. Critical infrastructure interdependency modeling: A survey of u.s. and international research. *Idaho National Laboratory, INL/EXT-06-11464*, 2006.

57. M. A. Porter, J.-P. Onnela, and P. J. Mucha. Communities in networks. *Notices of the American Mathematical Society*, 56(9):1082–1097, 2009.
58. E. Quill. When networks network. *Science News*, 182(6), 2012.
59. M. RE. North America most affected by increase in weather-related natural catastrophes. http://www.munichre.com/en/media_relations/press_releases/2012/2012_10_17_press_release.aspx, Oct. 17, 2012
60. S. Rinaldi, J. Peerenboom, and T. Kelly. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems Magazine*, December:11–25, 2001.
61. S. M. Rinaldi. Modeling and simulating critical infrastructures and their interdependencies. In *38th Hawaii International Conference on System Sciences*, pages 1–8, Big Island, Hawaii, 2004.
62. M. P. Rombach, M. A. Porter, J. H. Fowler, P. J. Mucha. Core-Periphery Structure in, Networks. arXiv:1202.2684, Feb. 2012.
63. V. Rosato, L. Issacharoff, F. Tiriticco, S. Meloni, S. D. Procellinis, and R. Setola. Modelling interdependent infrastructures using interacting dynamical models. *Int. J. Critical Infrastructures*, 4(1/2):63–79, 2008.
64. A. Saichev and D. Sornette. Anomalous power law distribution of total lifetimes of branching processes: Application to earthquake aftershock sequences. *Physical Review E*, 70(4):046123, Oct 2004.
65. P. Sinha-Ray and H. J. Jensen. Forest-fire models as a bridge between different paradigms in self-organized criticality. *Physical Review E*, 62(3):3216, sep 2000.
66. N. N. Taleb. *The Black Swan: The Impact of the Highly Improbable*. Random House Inc., New York, NY, 2007.
67. N. N. Taleb. *Antifragile: Things That Gain from Disorder*. Random House Inc., New York, NY, November 2012.
68. B. Wang, X. Chen, and L. Wang. Probabilistic interconnection between interdependent networks promotes cooperation in the public goods game. arXiv:1208.0468, Nov. 2012.
69. Z. Wang, A. Scaglione, and R. J. Thomas. Generating statistically correct random topologies for testing smart grid communication and control networks. *IEEE Transactions on Smart Grid*, 1:28–39, 2010.
70. S. Wasserman and K. Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.
71. N. Wolchover. Treading softly in a connected world. *Simons Science News*, (http://www.wired.com/wiredscience/2013/03/math-prevent-network-failure/), 2013
72. R. Zimmerman, C. Murillo-Sánchez, and R. Thomas. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *Power Systems, IEEE Transactions on*, 26(1):12–19, feb. 2011.

Chapter 5
Thresholds and Complex Dynamics of Interdependent Cascading Infrastructure Systems

B. A. Carreras, D. E. Newman, I. Dobson, V. E. Lynch and Paul Gradney

Abstract Critical infrastructures have a number of the characteristic properties of complex systems. Among these are infrequent large failures through cascading events. These events, though infrequent, often obey a power law distribution in their probability versus size which suggests that conventional risk analysis does not apply to these systems. Real infrastructure systems typically have an additional layer of complexity, namely the heterogeneous coupling to other infrastructure systems that can allow a failure in one system to propagate to the other system. Here, we model the infrastructure systems through a network with complex system dynamics. We use both mean field theory to get analytic results and a numerical complex systems model, Demon, for computational results. An isolated system has bifurcated fixed points and a cascading threshold which is the same as the bifurcation point. When systems are coupled, this is no longer true and the cascading threshold is different from the bifurcation point of the fixed point solutions. This change in the cascading threshold caused by the interdependence of the system can have an impact on the "safe operation" of interdependent infrastructure systems by changing the critical point and even the power law exponent.

B. A. Carreras
BACV Solutions, Inc., Oak Ridge, TN 37830, USA

D. E. Newman (✉) · P. Gradney
Physics Department, University of Alaska, Fairbanks, AK 99775, USA
e-mail: denewman@alaska.edu

I. Dobson
ECE Department, Iowa State University, Ames, IA 50011, USA

V. E. Lynch
Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

5.1 Introduction

Many critical infrastructure systems exhibit the type of behavior that has come to be associated with "Complex System" dynamics. These systems range from electric power transmission and distribution systems, through communication networks, commodity transportation infrastructure and arguably all the way to the economic markets themselves. There has been extensive work in the modeling of some of these different systems. However, because of the intrinsic complexities involved, modeling of the interaction between these systems has been limited [1–3]. At the same time, one cannot simply take the logical view that the larger coupled system is just a new larger complex system because of the heterogeneity introduced through the coupling of the systems. While the individual systems may have a relatively homogeneous structure, the coupling between the systems is often both in terms of spatial uniformity and in terms of coupling strength, fundamentally different. Understanding the effect of this coupling on the system dynamics is necessary if we are to accurately develop risk models for the different infrastructure systems individually or collectively.

We have already investigated [4, 5] some of the effects of the coupling between systems by using a dynamical model of coupled complex systems, the Demon model. This model is an extension of the Complex System Models used to study forest fires [6, 7]. Here, we will focus on some particular aspects of this model, for which the coupling introduces some fundamental changes on the properties of the system.

This type of model is characterized by the existence of a bifurcated equilibrium. Here one equilibrium solution is such that all components of the system are working. The second type of equilibrium has a fraction of the components failed. As the load on the system increases (or the probability of failure propagation) there is a transition from the first type of equilibrium to the second, at a critical loading [8, 9]. In a single system, this transition point is also the threshold for cascading events of all sizes, that is, transitioning between "normally distributed events" [10] and large-scale failures.

The coupling between the systems can modify the system's behavior and therefore importantly, conditions for safe operation. In this model we introduce a possibility of failure propagation from one system to another not only when a component fails but also when a component is out of working order. This has two different effects. One is a tendency to keep some components failed while still in normal operation. How many depends on the ratio between the strength of the coupling and the repair rate. As we see later in this chapter, the "critical loading" bifurcation point of the equilibrium is reduced by a function of this ratio.

The second effect of the coupling is it allows propagation of failures from one system to the other during a cascading event. Therefore the cascading threshold is also lowered by an amount proportional to this coupling. Since the parameter controlling this effect is not the same as the one controlling the equilibrium bifurcation, the equilibrium bifurcation point and the cascading threshold are now different.

Because of these changes, the often used metrics [11–13] for determining the threshold for large scale cascading events in the system will be re-examined and we will study the effect of the coupling of the systems on these measures.

The rest of the chapter will be organized as follows: Sect. 5.2 gives a description of the coupled infrastructure model, Demon, and a summary of some of the results from that model. Section 5.3 introduces a mean field version of the model and uses it to study the possible steady state solutions. The dynamics from the perspective of the mean field theory is described in Sect. 5.4 and in Sect. 5.5 the results of this analysis is compared with the numerical solutions of the mean field model. Then in Sect. 5.6, the results of the mean field theory are compared with the results of the dynamical model Demon. Finally, in Sect. 5.7, a discussion of the implications of these results and conclusions are presented.

5.2 The Demon Model

The infrastructure model discussed here, the Demon model, is based on the forest fire model of Bak et al. [7] with modifications by Drossel and Schwabl [6].

For a single system, the model is defined on a user defined 2-D network. An example of such network is shown in Fig. 5.1. Nodes represent components of the infrastructure system and lines represent the coupling between components. These components can be operating, failed or failing. The rules of the model for each time step are:

(1) A failed component is repaired with probability P_r.
(2) A failing component becomes a failed one.
(3) An operating component fails with a probability P_n if at least one of the nearest components is failing.
(4) There is a probability P_f that any operating component fails.

The Demon model [4] considers a coupled system by taking two of these 2-D networks and adding another rule:

(5) A component in System 1 can fail with a coupling parameter c, if the associated component in System 2 is failed or failing. The same applies for a component in system 2.

The ordering of the four parameters in the model is very important as discussed in [6]. Here, for the particular infrastructure problem, the different probabilities can be directly related to the characteristic times of repair, failure, and propagation of failure. It is worth noting that the propagation of failure parameter, P_n, is closely related to the loading of the system in a real infrastructure or a more realistic infrastructure model such as those described in [8, 14, 15]. This means that in the real infrastructures and more realistic models, there is an additional feedback that moves the system to near its critical point. We will use data from the power transmission system as guidance for those values. A more difficult parameter to characterize is the parameter that measures the coupling between the systems. The ability to explore the couplings between systems is an important flexibility in Demon as real world systems can have a wide variety of couplings that can impact their dynamics. For example they can be coupled mono-directionally (often,
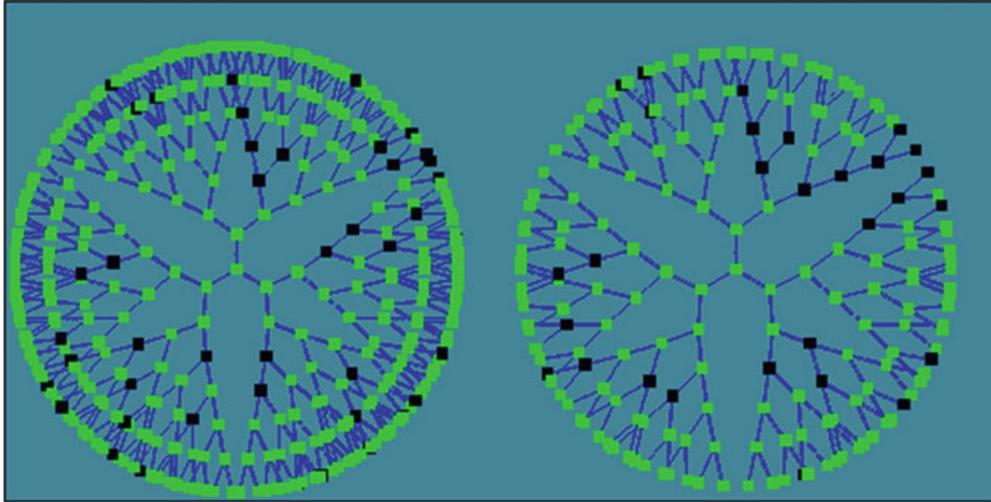
**Fig. 5.1** A pair of tree networks used for the modeling as an example
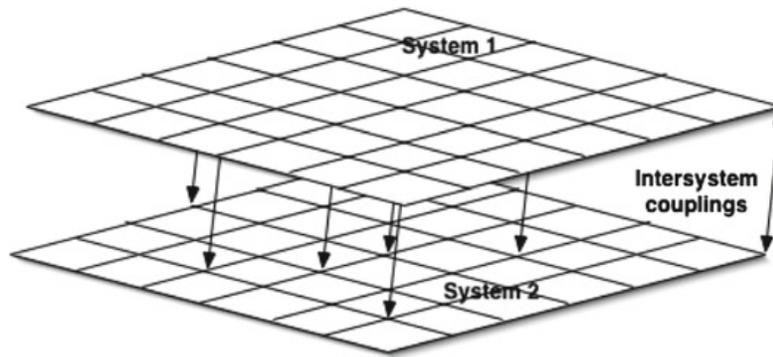
Fig. 5.2 A cartoon of the coupled networks, note that the number of nodes coupled between systems can be varied as can the strength, sign and directionality of the coupling

though not always, for pipeline-communications systems) or bi-directionally (most other systems, i.e. power transmission-communications systems), fully symmetric or asymmetric coupling strengths (failure in power transmission system has stronger impact on communications system then the other way around), homogeneously or heterogeneously (general spatial or course grained in one direction), negative rein-forcement (power transmission-communications) or positive (perhaps infrastructure systems—decision making "system"). A cartoon of this type of coupled system is shown in Fig. 5.2. For most of the work described here we will use the simplest types of couplings, namely symmetric, homogeneous and with negative reinforcement.

Using these rules, numerical calculations can be carried out, the dynamics and critical behavior investigated and impacts of system structure explored.

This model is an extension of a previous model [4] based on square grid networks to consider arbitrary network structures. Therefore, the basic coupling was from each node to four neighbors. The model in [4] was in turn a simple extension of the

Table 5.1 Network properties

Type	K	Number of nodes
Open 3-branch tree	2	3070
Closed 3-branch tree	3	3070
Open 5-branch tree	4	190
Square	3.96	10000
Hexagon	5.9	4681

Drossel model [6] for forest fires with the added rule number 5 that leads to nontrivial differences between those models. In the forest fire model, the propagation velocity of a disturbance is $P_n f$ where P_n is the probability of a disturbance to propagate from a node to another node and f the number of available nodes to propagate to, from a given node. In this model f is an important parameter to understand the propagation of the disturbances and it is not well determined. If K is the averaged number of nodes coupled to a single node in a given network, a first guess for f is $f = K - 1$, because the disturbance is already coming from one of the nodes that the failing node is coupled. In the case of the square network it was found [6] that $f = 2.66$ is a better value than 3. Therefore, we vary K in order to understand what the possible values are for f. In Table 5.1 we have summarized some of the properties of the different networks that we have considered in this chapter.

We will briefly look first at some of the results from this model, then we will investigate the mean field theory for this model and finally in Sect. 5.6 we will discuss the comparison between the mean field theory and the Demon model results.

When the control parameter, P_n, exceeds a critical (percolation) value the coupled system exhibits characteristics of a critical complex system. This critical value for an uncoupled system is given approximately by $P_n f = 1$, which is when the failures have a non-zero probability of propagating across the entire system. For the full-coupled system, the coupling between the two systems modifies this value. If the cross system coupling were the same as the coupling between nodes in each system, this would be the same as a larger system whose average node degree (effectively K) is increased by one. When the two systems being coupled are identical, but with a coupling strength different, typically much smaller, then P_n, the size distribution of failures obeys a power law which is close to -1 for all of the network structures examined. Below this critical value, the systems display an exponential distribution of failure size. An instructive exercise can be carried out by having the probability of random failures, P_f, non-zero in only one of the coupled systems. In this case it is found that if P_n is above the critical value and the coupling between the systems is also non-zero, the system in which there are no random failures also exhibits the characteristic power law size distribution (Fig. 5.3). This means that systems that look robust can actually be vulnerable when coupled making analysis of the entire coupled system critical. This cross system propagation is of course due to the coupling and can be seen in the synchronization of the failures in the two systems.

Using a measure developed by Gann et al. [16] for synchronization, which is basically an average normalized difference between events in the two systems, we

Fig. 5.3 The probability
distribution functions of the
failure sizes for a coupled
system in which only system
1 has random failures but
system 2 still has a power
law PDF and the combined
system has a heavier tail then
an uncoupled system would
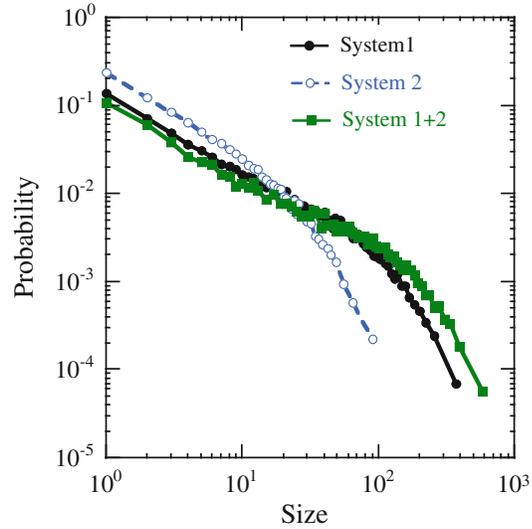
Fig. 5.4 The synchronization
function as a function of size
for a coupled system and three
values of the coupling coef-
ficient. Since 0 is fully syn-
chronized and 1 is completely
unsynchronized, it can be seen
that the synchronization is
stronger for larger failures
and increases then saturates
for larger values of coupling
strength
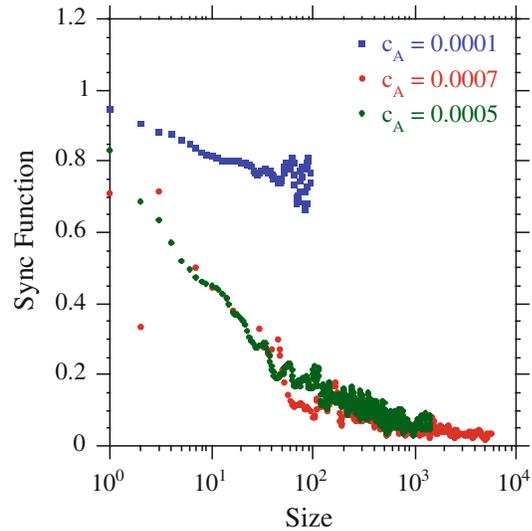
investigate this effect. For this measure, a value of 1 means the difference is effectively
100 % or no synchronization, while a value of 0 means all events are the same in the
two systems, or they are synchronized. For the Demon model it is found that large
failures are more likely to be "synchronized" across the two dynamical systems,
Fig. 5.4, as seen by the decrease in the synchronization function (which is an increase
in the actual synchronization) as a function of size. This means that in the coupled
systems there is a greater probability of large failures and lesser probability of smaller
failures. This in turn causes the power law found in the probability of failure with
size to be less steep, Fig. 5.5, with the coupling (i.e. the risk of larger failures is
even higher in the coupled system). Above a certain value of the coupling, this effect
saturates as the largest events are fully synchronized. The value of slope of the power
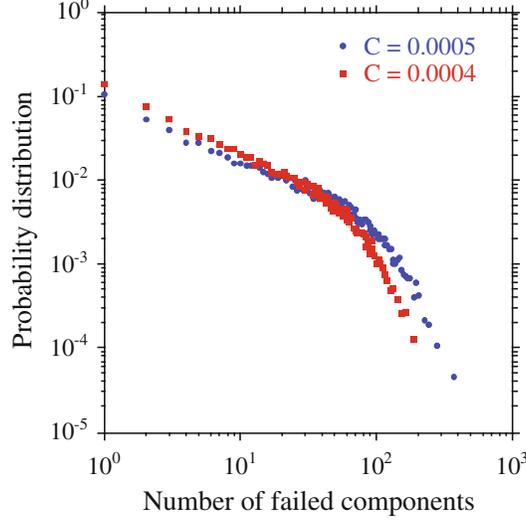
Fig. 5.5 PDF tail gets heavier as the coupling strength increases and then, as with the synchronization, saturates and stops changing

law for the coupled square grid with parameters given earlier approaches ~ -0.8 in contrast to ~ -1.0 for the uncoupled system.

The other major impact of the coupling on the system characteristics is the reduction of the critical point. As the coupling increases, the critical value of P_n, and by extension the loading, rapidly decreases (Fig. 5.6). This means that in an infrastructure system which by itself is nominally subcritical, the coupling, even weakly, to another infrastructure can make the entire system critical. This reduction will be further discussed in the next section on mean field theory of the coupled systems.

5.3 Mean Field Theory: Steady State

Let us consider first the mean field theory for two coupled systems. This is a generalization of the calculation as done in [6]. Let $O^{(i)}(t)$ be the number of operating components in system i at time t normalized to the total number of components $N^{(i)}$. In the same way, we can define the normalized number of failed components, $F^{(i)}(t)$, and the failing ones, $B^{(i)}(t)$. The mean field equations for this coupled system are:

$$B^{(1)}(t+1) = P_f^{(1)} O^{(1)}(t) + P_n^{(1)} f^{(1)} O^{(1)}(t) B^{(1)}(t) + \frac{c^{(1)}}{\mathcal{K}} g_2 O^{(1)}(t) \left(B^{(2)}(t) + F^{(2)}(t) \right)$$

$$(5.1)$$

$$F^{(1)}(t+1) = \left(1 - P_r^{(1)} \right) F^{(1)} + B^{(1)}(t) \tag{5.2}$$

$$O^{(1)}(t+1) = \left(1 - P_f^{(1)} \right) O^{(1)}(t) + P_r^{(1)} F^{(1)}(t) - P_n^{(1)} f^{(1)} O^{(1)}(t) B^{(1)}(t)$$
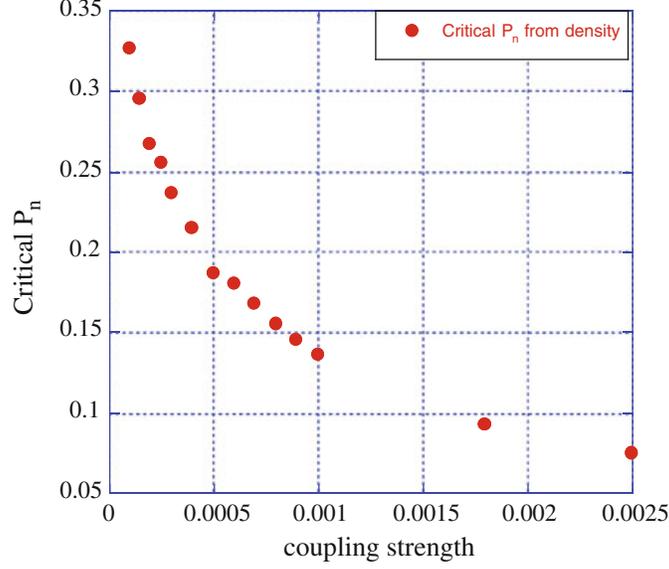
Fig. 5.6 The critical point decreases rapidly as the coupling strength increases. Even the maximum coupling strength is much less then the propagation coefficient within one system but the critical parameter has fallen by more than a factor of three

$$-\frac{c^{(1)}}{\mathcal{K}}g_2 O^{(1)}(t)\left(B^{(2)}(t)+F^{(2)}(t)\right) \tag{5.3}$$

$$B^{(2)}(t+1)=P_f^{(2)} O^{(2)}(t)+P_n^{(2)} f^{(2)} O^{(2)}(t)B^{(2)}(t)+\mathcal{K}c^{(2)}g_1 O^{(2)}(t)\left(B^{(1)}(t)+F^{(1)}(t)\right) \tag{5.4}$$

$$F^{(2)}(t+1)=\left(1-P_r^{(2)}\right)F^{(2)}(t)+B^{(2)}(t) \tag{5.5}$$

$$O^{(2)}(t+1)=\left(1-P_f^{(2)}\right)O^{(2)}(t)+P_r^{(2)}F^{(2)}(t)-P_n^{(2)} f^{(2)} O^{(1)}(t)B^{(2)}(t)$$

$$-\mathcal{K}c^{(2)}g_1 O^{(2)}(t)\left(B^{(1)}(t)+F^{(1)}(t)\right). \tag{5.6}$$

Here $\mathcal{K}=N^{(1)}/N^{(2)}$, g_1 is the fraction of nodes in system 1 coupled to system two, and g_2 is the fraction of nodes in system 2 coupled to system 1. Of course, these equations are consistent with the conditions:

$$O^{(i)}(t)+B^{(i)}(t)+F^{(i)}(t)=1 \tag{5.7}$$

In the limit with no failure triggers, $P_f^{(i)}=0$, and for a steady state solution, the system of equations can be reduced to two coupled equations,

$$\left[1-P_n^{(1)} f^{(1)} O^{(1)}\right]\left(1-O^{(1)}\right)=\frac{a^{(1)}}{\mathcal{K}}g_2\left(1-O^{(2)}\right)O^{(1)} \tag{5.8}$$

$$\left[1 - P_n^{(2)} f^{(2)} O^{(2)}\right]\left(1 - O^{(2)}\right) = \mathcal{K} a^{(2)} g_1 \left(1 - O^{(2)}\right) O^{(2)} \tag{5.9}$$

where

$$a^{(i)} = \frac{c^{(i)}\left(1 + P_r^{(i)}\right)}{P_r^{(i)}}. \tag{5.10}$$

It is important to note that the relevant parameter involves the ratio of the coupling between the systems to the repair rate. The reason for that is the particular form of rule (5) that assumes that a failure can be triggered by both failed and failing components in the other system. If only failing components had been considered, the relevant parameter would be the coupling. For real systems, a realistic rule should probably be in between these two.

If $a^{(i)} \neq 0$ and $\kappa = 1$, then $O^{(1)} = 1$ implies $O^{(2)} = 1$, that is, the systems are decoupled. Therefore, to have truly coupled systems, system 1 must be in a supercritical state. Such case with $a^{(i)} \neq 0$ is more complicated to solve.

First, we assume identical systems symmetrically coupled. That is, all parameters are the same for the two systems, $f^{(1)} = f^{(2)}$, $a^{(1)} = a^{(2)}$, $\kappa = 1$ and $P_{(n)}^{(1)} = P_{(n)}^{(2)}$. This leads to identical solutions for the two systems in steady state. Therefore, we have the following solutions:

$$O_{1eq}^{(i)} = 1, \quad F_{1eq}^{(i)} = 0, \quad B_{1eq}^{(i)} = 0 \tag{5.11}$$

and

$$O_{2eq}^{(i)} = \frac{1}{\hat{g}}, \quad F_{2eq}^{(i)} = \frac{\hat{g} - 1}{\hat{g}\left(1 + P_r\right)}, \quad B_{2eq}^{(i)} = \frac{\hat{g} - 1}{\hat{g}\left(1 + P_r\right)} P_r \tag{5.12}$$

The second solution is only valid for $\hat{g} > 1$. Here, $\hat{g}$ is the control parameter and is given by

$$\hat{g} = P_n f + \frac{c\left(1 + P_r\right)}{P_r} \tag{5.13}$$

In Eqs. (5.11) and (5.12) the subindex eq indicate that is an equilibrium solution.

The bifurcation point of the fixed point, $\hat{g} = 1$, has been decreased from the decoupled case, $P_n f = 1$, by a term proportional to c/P_r. Therefore, in general this reduction is considerably larger than the magnitude of the coupling itself.

We have tested the results from the mean field theory by comparing them with numerical results from some of the two-coupled identical systems networks listed in Table 5.1. The results for the averaged number of operating components are shown in Fig. 5.7. Results have been obtained for fixed $P_r = 0.001, c = 0.0005, P_f = 0.00001$ (for system 1) and $P_f = 0$ (system 2), and we have varied the propagation parameter P_n. The numerical results show very good agreement with the mean field theory results as K increases. For $K = 2$, the systems are practically one-dimensional and the mean field theory is not really applicable.
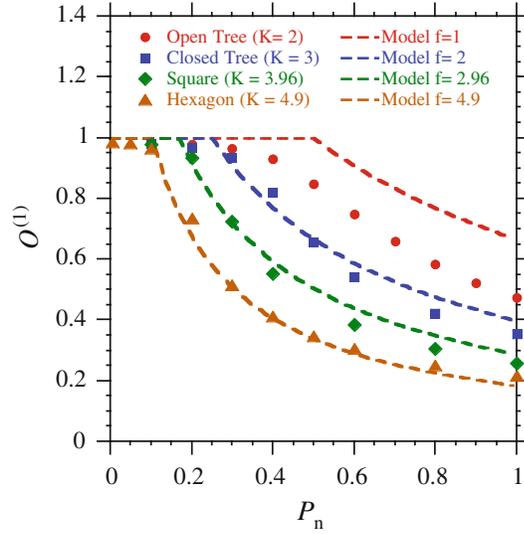
The density of operating components is practically the same in both systems. This is logical because they are identical systems the only symmetry breaking feature is the probability of spontaneous failures that is zero in the second system.

5.4 Mean Field Theory: Time Evolution

We will continue to assume the two systems are identical, in this way, we can simplify the system of Eqs. (5.1)–(5.6) to the following system:

$$B(t + 1) = P_n f O(t) B(t) + c O(t)\big(B(t) + F(t)\big) \tag{5.14}$$

$$F(t + 1) = \big(1 - P_r\big) F(t) + B(t) \tag{5.15}$$

$$O(t) + F(t) + B(t) = 1. \tag{5.16}$$

We have eliminated the super-indices indicating the system because we assume that the two systems are identical at all times. This system of equations has two fixed points or equilibrium solutions, which are the same as before and given by Eqs. (5.11) and (5.12). For $\hat{g} < 1$, there is a single fixed point, but for $\hat{g} > 1$ there are two steady state solutions. We can study the stability of the solutions by linearizing Eqs. (5.14)–(5.16):

$$\begin{pmatrix} B(t+1) \\ F(t+1) \end{pmatrix} = \begin{pmatrix} (P_n f + 2c)O_{jeq} - P_n f B_{eq} - c & 2c O_{jeq}^{(i)} - P_n f B_{jeq} - c \\ 1 & 1 - P_r \end{pmatrix} \begin{pmatrix} B(t) \\ F(t) \end{pmatrix}.$$

At the standard operation equilibrium, no failures fixed point, the linearization becomes

$$
\begin{pmatrix} B(t+1) \\ F(t+1) \end{pmatrix} = \begin{pmatrix} P_n f + c & c \\ 1 & 1 - P_r \end{pmatrix} \begin{pmatrix} B(t) \\ F(t) \end{pmatrix} \tag{5.17}
$$

and the eigenvalues are

$$
\gamma_{1+} = \frac{1}{2} \left(1 - P_r + P_n f + c + \sqrt{(1 - P_r - P_n f - c)^2 + 4c} \right) \tag{5.18}
$$

$$
\gamma_{1-} = \frac{1}{2} \left(1 - P_r + P_n f + c - \sqrt{(1 - P_r - P_n f - c)^2 + 4c} \right). \tag{5.19}
$$

For $\hat{g} = 1$ the largest eigenvalue γ_{1+} goes through 1. This indicates that the fixed-point solution Eq. (5.11) becomes unstable at this point.

Similar calculation evaluating the linearization at the second fixed point, Eq. (5.12), shows that this second fixed point is stable for $\hat{g} > 1$. The bifurcation is essentially a transcritical bifurcation and the stability is transferred from the fixed point Eq. (5.11) to the appearing second fixed point Eq. (5.12) as increases through $\hat{g} = 1$.

The left eigenvectors corresponding to the eigenvalues Eqs. (5.18) and (5.19) are

$$
\vec{V}_{i+} = (\gamma_{i+} - 1 + P_r, c), \quad \vec{V}_{i-} = (\gamma_{i-} - 1 + P_r, c) \tag{5.20}
$$

We can use these eigenvectors to calculate the eigenvalues from measured quantities, because by applying them on the left of Eq. (5.17), we obtain

$$
\gamma_{i+} = \frac{(-1 + P_r + \gamma_{i+}, c) \cdot \begin{pmatrix} B(t+1) \\ F(t+1) \end{pmatrix}}{(-1 + P_r + \gamma_{i+}, c) \cdot \begin{pmatrix} B(t) \\ F(t) \end{pmatrix}} \tag{5.21}
$$

$$
\gamma_{i-} = \frac{(-1 + P_r + \gamma_{i-}, c) \cdot \begin{pmatrix} B(t+1) \\ F(t+1) \end{pmatrix}}{(-1 + P_r + \gamma_{i-}, c) \cdot \begin{pmatrix} B(t) \\ F(t) \end{pmatrix}} \tag{5.22}
$$

From these expressions, we can derive a diagnostic to determine the eigenvalues from the numerical calculations. The expressions are

$$
[\gamma_{i\pm}] = \frac{1}{2} \left\{ 1 - P_r + \frac{B(t+1)}{B(t)} - c\frac{F(t+1)}{B(t)} \right.
$$
$$
\left. \pm \sqrt{\left[-1 + P_r + \frac{B(t+1)}{B(t)} - c\frac{F(t)}{B(t)} \right]^2 - 4\frac{(-1 + P_r)F(t) + F(t+1)}{B(t)}c} \right\} \tag{5.23}
$$

Here, we use the square brackets around the γ's to indicate that these values will be obtained from numerical results. They are diagnostics and should not be confused with the analytical value of the eigenvalues. Note that for $c = 0$, these two eigenvalues are

$$\left[\gamma_{1-}\right] = \frac{B(t+1)}{B(t)} \quad \text{and} \quad \left[\gamma_{1+}\right] = 1 - P_r. \tag{5.24}$$

The first one is identical to the standard metric used in determining criticality with respect to cascading events [11–13]. This metric is defined as

$$\lambda_B(t+1) = \frac{B(t+1)}{B(t)} \tag{5.25}$$

and measures the propagation of the failures. If λ_B is greater than 1, the number of failures increases with time and there is the possibility of a large cascading event. However, if λ_B is less than one, the failures will stop propagating and the failure size remains small. This measure has been introduced [17–19] on the basis of a branching process [20].

Having diagonalized the matrix in Eq. (5.17), one can solve the linear equations by iteration and one obtains

$$\begin{pmatrix} B(t+1) \\ F(t+1) \end{pmatrix} = \frac{B(1)}{\gamma_{i+} - \gamma_{i-}} \begin{pmatrix} \left[\gamma_{i+} - 1 + P_r\right]\gamma_{i+}^t - \left[\gamma_{i-} - 1 + P_r\right]\gamma_{i-}^t \\ \frac{\gamma_{i+}^t - \gamma_{i-}^t}{O_{ieq}} \end{pmatrix}. \tag{5.26}$$

From this solution, we can calculate the propagation of the failures

$$\lambda_{iB}(t+1) = \frac{B(t+1)}{B(t)} = \frac{\left[\gamma_{i+} - 1 + P_r\right]\gamma_{i+}^t - \left[\gamma_{i-} - 1 + P_r\right]\gamma_{i-}^t}{\left[\gamma_{i+} - 1 + P_r\right]\gamma_{i+}^{t-1} - \left[\gamma_{i-} - 1 + P_r\right]\gamma_{i-}^{t-1}}. \tag{5.27}$$

However, this is the solution of the linear problem, it only make sense for $t \to \infty$ in the initial, linear, phase of the evolution. The asymptotic values for are meaningless. This ratio of failing components gives a measure of the propagation of the failures. Here we want to examine the relation between the cascading point, $\lambda_{ib} = 1$, and the equilibrium bifurcation point, $\hat{g} = 1$, which for $c = 0$ were the same. The important question is what is the proper diagnostic to measure the cascading threshold.

From Eq. (5.27) the first two values for the rate of propagation of failures are

$$\lambda_{1B}(2) = P_n f + c \tag{5.28}$$

$$\lambda_{1B}(3) = P_n f + c + \frac{c}{P_n f + c}. \tag{5.29}$$

As one can see in the case of coupled systems, Eq. (5.27) gives λ_{1B} as an increasing function of t. If the first value of λ_{1B} is greater than 1, the cascade will go on. This

is a sufficient condition for the cascade threshold. It is however not a necessary condition, because in a few initial steps the number of failures may first decrease till λ_{1B} becomes greater than one and increase again. How many steps can λ_{1B} be less than 1 without extinguishing the cascade is not clear, it will depend on the size of the initial perturbation. If we assume that three steps are sufficient, we can use Eq. (5.29) as a typical parameter controlling the cascade, then the cascading threshold is

$$P_n f = \frac{1}{2}\left(1 + \sqrt{1 - 4c}\right) - c. \tag{5.30}$$

Since $c \ll 1$ in the relevant cases, we can write this threshold in a more general way

$$P_n f = 1 - \mu c \tag{5.31}$$

where μ is a number of order 1 to be determined by numerical calculations. Here, we see that the effect of the coupling is to reduce the cascading threshold by a factor of the order of c, while the equilibrium bifurcation point was reduced by a larger term of the order c/P_r.

5.5 Mean Field Theory: Numerical Solution

The mean field theory system of equations, Eqs. (5.14)–(5.16), can be solved numerically without any further assumptions. The nonlinear solutions of these equations will allow us to evaluate better the meaning of the analytical results described in the previous section and the validity of the linear approximations. Here, we consider systems with 10^4 components and the values of the couplings are $c = 0.0005$ and $P_r = 0.001$. For these parameters, the equilibrium bifurcation point, $\hat{g} = 1$, is at $P_n f = 0.4995$.

In Fig. 5.8, we have plotted the fraction of failing components as a function of the iteration for different values of $P_n f$. This plot gives a good description of the propagation of the failures. For all cases we have the used the same initial condition:

$$O_{init} = 1, B_{init} = \frac{1}{N}, \text{ and } F_{init} = 0 \tag{5.32}$$

where N is the total number of components. In this case, if B goes below 10^{-4} the cascade has effectively extinguished because the system has 10^4 components. However, for an initial condition with n failures, the cascade is extinguished for $B = n/N$.

In looking at the Fig. 5.9, it is clear that the cascading threshold is close to $P_n f = 1$. We have repeated the calculation for $c = 0.05$ and $P_r = 0.1$. These parameters are unrealistically large, and the steady state value of B is considerably larger than the initial value. In this case, below the cascading threshold the number of failures dips

Fig. 5.8 The fraction of failing components as a function of the iteration for different values of $P_n f$. In this case the parameters are $c = 0.0005$ and $P_r = 0.001$ and the critical value of $P_n f$ is about 1
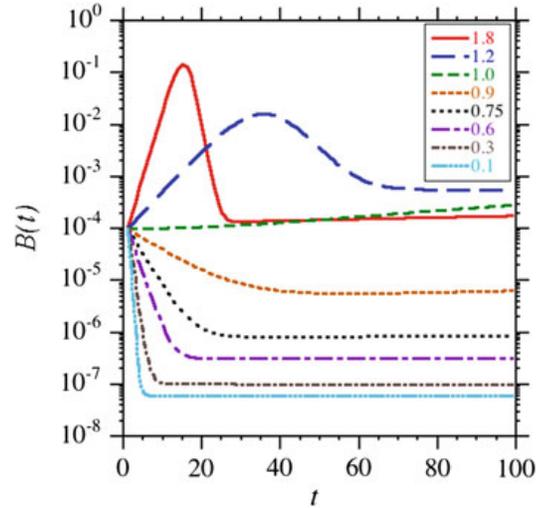

Fig. 5.9 The fraction of failing components as a function of the iteration for different values of $P_n f$. In this case the parameters are $c = 0.05$ and $Pr = 0.1$ and the critical value of $P_n f$ is about 0.9
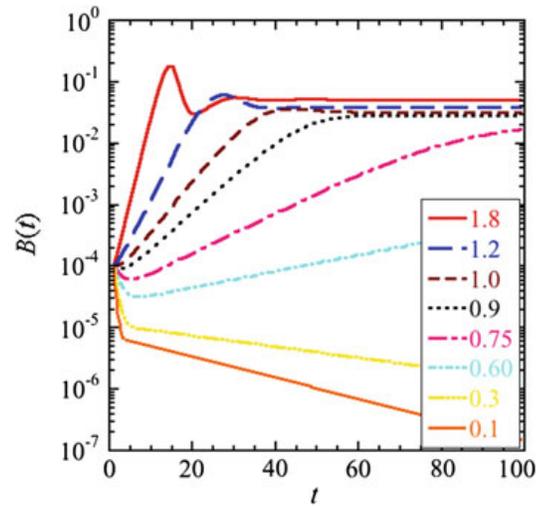
down well below 10^{-4} before rising again to the steady state value. The results are shown in Fig. 5.9. Because the equilibrium bifurcation point depends mostly on the ratio c/P_r, this change of parameters hardly changes the equilibrium bifurcation point. However, the cascading threshold depends on the value of c, therefore and as expected, the threshold for cascading is now close to $P_n f = 0.9$.

From the previous analytical calculations and these numerical results we can draw two conclusions:

(1) The cascading threshold at which failures initially grow is not at the equilibrium bifurcation point. From the numerical calculations we can see that this threshold is consistent with Eq. (5.31). Therefore, the initial cascade propagation does not seem to be linked to the largest eigenvalue of the linear approximation to the mean field equations as it was for the decoupled systems.

Fig. 5.10 The analytic and
computed γ_{1+} eigenvalues for
two values of $P_n f$
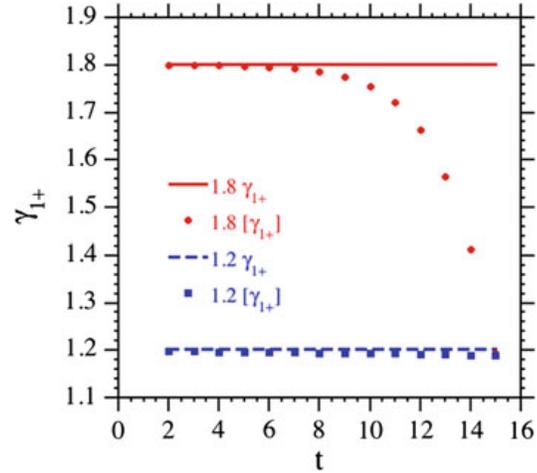
(2) In the cases above the cascading threshold, the cascade starting near the unstable
     fixed point (11) proceeds up to a certain size, and then decreases as the transient
     converges to the stable fixed point (12). This is a transient and nonlinear system
     effect, which is not taken into account in the linearization that is valid only near
     the fixed point (11).

The next step is comparing the calculated eigenvalues Eq. (5.23) with the an-
alytical ones, Eqs. (5.16)–(5.19). We compare the measured eigenvalues from the
numerical solution to the eigenvalues for the first fixed point because of the initial
conditions taken here. For γ_{1+} and $P_n f < 1$, the analytical and numerical values are
very close to 1 and any difference would be small. Therefore, we limit the numerical
comparison to $P_n f > 1$.

In Fig. 5.10, we have plotted the γ_{1+} eigenvalue and the measured one, $[\gamma_{1+}]$ from
the mean field numerical calculations for two values of $P_n f$. There the agreement
is good. In Fig. 5.11, we have the same comparison for γ_{1-}. Again the agreement
is very good. In particular, the agreement is expected to be better for the very low
number of iterations, because no finite size effects are present. Note that γ_{1-} is the
eigenvalue associated with the transition of the fixed point 1 to a fixed point 2.

The next step is to compare the propagation of failures from the linear calculation
with the solution of the mean field theory. In Fig. 5.12, we compare λ_B, as calculated
numerically from the nonlinear mean field theory in Eq. (5.25), with the value in
Eq. (5.27) obtained from the linear approximation Eq. (5.26).

We see that the mean field theory gives a value for λ_B that increases with time.
It is not constant as obtained from a branching process. Therefore this confirms the
previous assumption that the cascade threshold can be calculated by Eq. (5.30).

Fig. 5.11 The analytic and computed γ_{1-} eigenvalues for a number of values of $P_{\mathrm{n}}f$
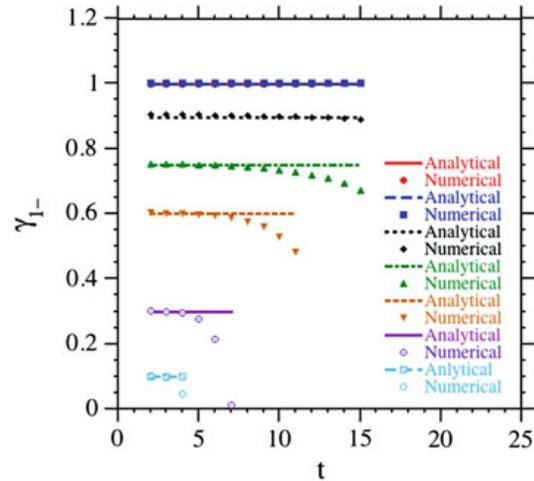
Fig. 5.12 Comparison of λ_B, as calculated numerically from the nonlinear mean field theory in Eq. (5.25), with the value in Eq. (5.27) obtained from the linear approximation
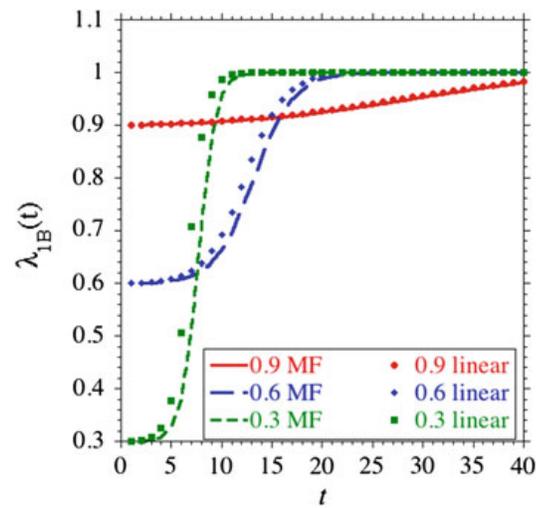
5.6 Application and Comparison to the Demon Model

Now that we have an understanding from the mean field theory of what should be measured, we can apply these measurements to the full dynamical model, Demon. In this case, the measurements will by necessity have a statistical character.

The equilibrium bifurcation is linked to the γ_{1-} eigenvalue. Therefore to get a sense of the equilibrium bifurcation point, we can apply the $[\gamma_{1-}]$ diagnostic, Eq. (5.23), to the Demon numerical calculations. The comparison between the analytical eigenvalue and the measured $[\gamma_{1-}]$ in Demon is shown in Fig. 5.13. The agreement is relatively good for a low number of iterations. As the number of iterations increases, finite size effects become important and the analytical and numerical results diverge, as we should expect. If instead of the linear analytical result we used the mean field theory result the agreement would be better.

Fig. 5.13 The analytic and Demon γ_{1-} eigenvalues for a number of values of $P_n f$
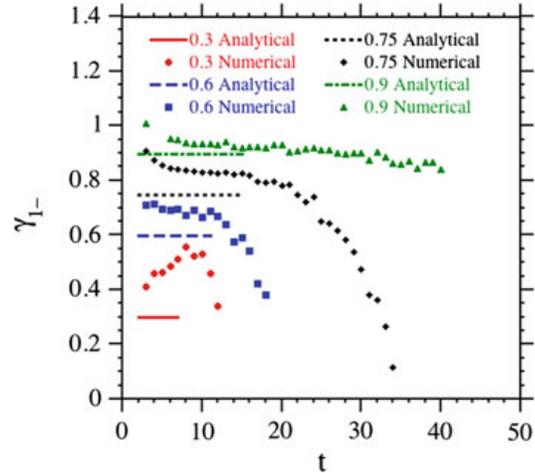
Fig. 5.14 Comparison between Demon and non-linear mean field (Eq. (5.30)) λ_B for a number of values of $P_n f$
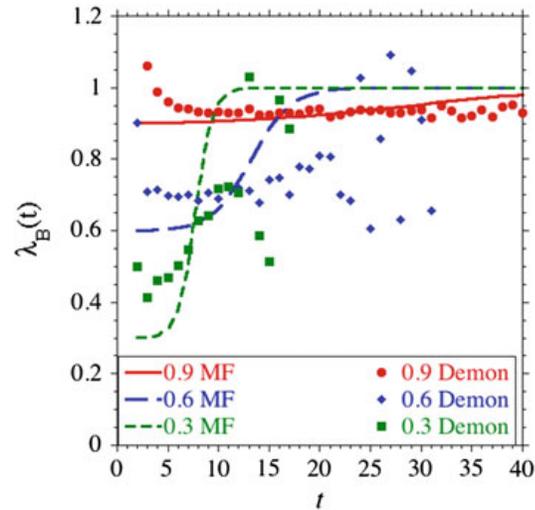
The other relevant parameter is the rate of propagation of the cascades. We can compare λ_B in the Demon model with the one calculated from the mean field theory, Eq. (5.30). The result of this comparison is shown in Fig. 5.14. We can see that there is a basic agreement.

Both comparisons are poorer for low values of γ_{1-} and λ_B, because in this parameter region the data from Demon are scattered. The reason for that is that there is a very small number of cascading events and the statistical evaluation is poor.

5.7 Conclusions

The critical infrastructure systems upon which modern society relies often exhibit characteristics of complex dynamical systems operating near their critical point including heavy, power law, tails in the failure size distribution and long time cor-

relations. We as a matter of course take their smooth operation for granted and are typically shocked when one of these systems fails despite the fact that these failures are a completely inevitable result of the complex dynamical nature of the system. Though failures are inevitable, one can design and operate the systems to reduce the risk or at least be aware of what the risk is, making understanding these systems a high priority for ensuring security and social wellbeing. While modeling these individual systems themselves is a challenging and worthwhile exercise, in the real world they usually do not exist in a vacuum, instead being coupled, sometimes very tightly, to one or more other complex infrastructure systems. This coupling can lead to new behavior including modifications of the critical points and the weight of the tails. Realistically modeling these coupled infrastructure systems in a dynamic manner is a daunting task outside out current capabilities. For example, only recently has a simplified model of the electric power grid alone with cascading overloads and complex dynamics been validated with observed data [21]. Therefore simpler models that can capture some of the important characteristics have a significant role to play in understanding the risks associated with the structure and growth of these critical systems. Even the simple modeling of these coupled system leads to a very large parameter space that must be explored with different regions of parameter space having relevance to different coupled infrastructure systems. Within each of these parameter regimes there is a rich variety of dynamics to be characterized.

This chapter has attempted to look at a simple model, Demon, of coupled infrastructure systems that can both be simulated and attacked analytically using mean field theory. We have found that in the region of parameter space we have explored, the coupling between the systems reduces the critical point (the propagation parameter in this model which is related to the system loading in the real world) and makes the tail heavier. The reduction in the critical point is found both in the mean field theory and the Demon model. This reduction has serious implications for the real world as we load the systems more heavily and as the coupling becomes ever tighter, suggesting that the probability of large failures is likely to become more probable. The mean field theory does a better job of matching the numerical results when taken to higher order and is even able to capture the general time behavior of the propagation metric λ. This metric is one which can in principle be measured in the real world [22] as a state estimator. Using this it may be possible to give a statistical estimate of the risk of failure of various sizes, a needed function given the non-normal nature of these distribution functions. In addition, it has been found that the PDF of the failure sizes gains a heavier tail, with the slope going from ~-1.0 to ~-0.8. While this may not seem like a major change, because this is a power law it implies a significantly higher relative risk of the larger failures which are the failures that dominate the "cost" to society. These models find that even with weak interaction one cannot always safely ignore coupling.

Characterizing the dynamics in the different regimes is more than an academic exercise since as we engineer higher tolerances in individual systems and make the interdependencies between systems stronger we will be exploring these new parameter regimes the hard way, by trial and error. Unfortunately error in this case has the potential to lead to global system failure. By investigating these systems from

this high level, regimes to be avoided can be identified and mechanisms for avoiding them can be explored. These general relationships are then available to be verified either with more physically based models or with real data.

References

1. Richard G. Little, Toward More Robust Infrastructure: Observations on Improving the Resilience and Reliability of Critical Systems, in Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03).
2. S. M. Rinaldi, Modeling and Simulating Critical Infrastructures and Their Interdependencies, in Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04), (Big Island, HI, USA), IEEE Computer Society Press, Jan. 2004.
3. S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, *Identifying, understanding, and analyzing critical infrastructure interdependencies*, IEEE Control Systems Magazine, p. 11, December 2001.
4. D. E. Newman, B. Nkei, B. A. Carreras, I. Dobson, V. E. Lynch, P. Gradney, Risk assessment in complex interacting infrastructure systems, *Thirty-eighth Hawaii International Conference on System Sciences*, Hawaii, January 2005.
5. B. A. Carreras, D. E. Newman, Paul Gradney, V. E. Lynch, and I. Dobson, Interdependent Risk in Interacting Infrastructure Systems, *40th Hawaii International Conference on System Sciences, Hawaii*, Hawaii, Jan 2007.
6. B. Drossel and F. Schwabl, Physica A 199, 183 (1993).
7. P. Bak, K. Chen and C. Tang, Phys. Lett. A 147, 297 (1990).
8. B.A. Carreras, V.E. Lynch, I. Dobson, D.E. Newman, Critical points and transitions in an electric power transmission model for cascading failure blackouts, Chaos, vol. 12, no. 4, December 2002, pp. 985–994.
9. I. Dobson, J. Chen, J.S. Thorp, B. A. Carreras, and D. E. Newman, *Examining criticality of blackouts in power system models with cascading events*, 35th Hawaii International Conference on System Sciences, Hawaii, Hawaii, Jan. 2002.
10. Charles Perrow, *Normal accidents*, Princeton University Press, 1984.
11. I. Dobson, B. A. Carreras, and D. E. Newman, A probabilistic loading-dependent model of cascading failure and possible implications for blackouts, *36th Hawaii International Conference on System Sciences, Maui*, Hawaii, Jan. 2003.
12. I. Dobson, B.A. Carreras, D.E. Newman, Probabilistic load-dependent cascading failure with limited component interactions, *IEEE International Symposium on Circuits and System, Vancouver*, Canada, May 2004.
13. I. Dobson, B. A. Carreras, and D. E. Newman, A loading-dependent model of probabilistic cascading failure, Probability in the Engineering and Informational Sciences 19 (1), 15–32.
14. D.E. Newman, B.A. Carreras, V.E. Lynch, I. Dobson, Exploring complex systems aspects of blackout risk and mitigation, Reliability, IEEE Transactions on 60 (1), 134–143
15. I. Dobson, B.A. Carreras, V.E. Lynch, D.E. Newman, Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization, Chaos, vol. 17 no. 2, 26103
16. R. Gann J. Venable, E.J. Friedman, A.S. Landsberg, Behavior of coupled automata, Phys. Rev. E 69, 046116 (2004).
17. I. Dobson, B.A. Carreras, D.E. Newman, A branching process approximation to cascading load-dependent system failure. *37th Hawaii International Conference on System Sciences*, Hawaii, January 2004.

18. I Dobson, BA Carreras, DE Newman, Branching process models for the exponentially increasing portions of cascading failure blackouts, System Sciences, 2005, Proceedings of the 38th *Hawaii International Conference on System Sciences*.

19. I Dobson, KR Wierzbicki, BA Carreras, VE Lynch, DE Newman, An estimator of propagation of cascading failure, System Sciences, 2006. HICSS'06. Proceedings of the 39th *Hawaii International Conference on System Sciences*

20. T.E. Harris, Theory of branching processes, Dover NY 1989.

21. B.A. Carreras, D.E. Newman, I. Dobson, N.S. Degala, Validating OPA with WECC data, Forty-sixth Hawaii International Conference on System Sciences, Maui, Hawaii, January 2013.

22. I Dobson, Estimating the propagation and extent of cascading line outages from utility data with a branching process, IEEE Transactions on Power Systems 27 (4), 2116–2126

Part II
Applications

One the most exciting challenges is to reduce the complexity of real infrastructures to a simpler and insightful model. In this part of the book we present some applications to real infrastructures, like transport, the electric grid and even the human body.

Chapter 6 builds up an abstract model to predict epidemic-like fault propagation and cascades, thus allowing a statistical description of the real system.

In Chap. 7 the most classical problem of traffic routing and congestion is analysed taking into account the interaction between public and private transportation. At the theoretical level, this corresponds to the interaction between two or more 'quasi-planar' graphs (spatial networks).

Chapter 8 represents a very broad and general introduction to the electrical power system and concentrates on the blackout mitigation strategies based on 'islanding', i.e. the intentional partition of the electric grid into smaller full operating sub-networks.

Chapter 9 highlights how the complexity of the power system does not reduce to the interaction among its physical components. In fact, all real infrastructures require human intervention at different levels, like governance (decision making) or planning (policies), that nowadays have to take into account also the intermingling of economic and social networks influencing the whole system.

Finally, Chap. 10 is not directly related to critical infrastructures; however, it provides a novel paradigm for one of the most complex 'system of systems', i.e. the human body. The results of the chapter are strictly related to one of the most stringent problems in critical infrastructures, i.e. the analysis of the interdependencies among systems through historical data.

Chapter 6
Characterizing Relevant Network Structure with Reliability Polynomials

Stephen Eubank, Mina Youssef and Yasamin Khorramzadeh

6.1 Introduction

Coupled socio-technical infrastructure networks are most usefully represented as complex, irregular graphs with directed, weighted edges that link vertices associated with states. Cascading failures in such systems are an example of a reaction-diffusion process over the graph. Reaction-diffusion have systems, of course, been studied from many perspectives over the past century. It is well known that many kinds of failures (such as illnesses or power outages) that propagate from vertex to vertex along the edges correspond to bond percolation [8]. Our understanding of Ising models and their generalizations thus provides insight into cascading failures. However, applying this insight to existing real-world infrastructure networks presents a problem: how does the structure of a finite, complex network affect phenomena that we have studied, for the most part, on highly regular (e.g. lattices, trees) or essentially unstructured (e.g. random) networks?

Here we re-introduce an approach that was developed to study the reliability of specific networks, making no assumptions about regularities and symmetries. We define and provide algorithms for calculating the reliability, discuss its interpretation in terms of statistical physics, provide illustrative examples on several networks, and show how it can be used to understand complicated phenomena. We illustrate how to compare two graphs and identify the structural differences between them that are most relevant for a given dynamical process. We also indicate how reliability can be used to infer structure, in the sense of network tomography.

S. Eubank (✉) · M. Youssef · Y. Khorramzadeh
Virginia Bioinformatics Institute at Virginia Tech, Blacksburg, VA, USA
e-mail: eubank@vt.edu

M. Youssef
e-mail: myoussef@vbi.vt.edu

Y. Khorramzadeh
e-mail: yasi@vbi.vt.edu

The novelty in this discussion is not the concept of reliability itself—the *IEEE Transactions on Reliability* is now in its 61st year; nor is it in the statistical physics of reliability. It is in our suggestions that

1. coefficients of the reliability polynomial are the best way to characterize graph structure and
2. network analysis in terms of reliability emphasizes global structure, providing insights that elude other approaches.

Reliability refocuses the question of structural effects from the individual inter-actions between elements to global structural properties, suggesting new methods of analysis. By design, a set of reliability coefficients encodes all the structure of a graph that is relevant to a dynamical phenomenon of interest. Hence it is a structural measure that is immediately connected with dynamics. In contrast to static measures like degree, modularity and measures of centrality, no intervening theory is needed to make the connection to dynamical phenomena. Moreover, the reliability can be further decomposed into the product of a purely combinatorial factor and a structure-dependent factor. The latter forms the basis for categorizing graphs in a way that folds together static properties into precisely the combinations that are most relevant to the dynamics.

6.2 Reliability

We have previously defined the concepts of *vulnerability* and *criticality* e.g. [3]. The vulnerability of a set $\mathscr{V}$ of vertices is essentially the probability that all of the vertices' states change, while criticality is the difference between the expected number of state changes in graphs with and without $\mathscr{V}$. These notions distinguish graphs in a way that is immediately relevant to, for example, designing interventions to mitigate an outbreak of infectious disease. Vulnerability is just one example of a class of *reliability* measures, which encode all—and only—the structure of a graph that is relevant for a particular dynamical process. Criticality can be calculated as the difference in reliability between graphs with and without a collection of edges, as in the example in Sect. 6.5 below.

Consider a graph (directed or undirected) $G(V, E)$ with N vertices and M weighted edges. For now, assume the edges are weighted homogeneously with weight $x \in [0, 1]$. The extension to heterogenous weights is possible, but more complicated, as discussed in Sect. 6.2.5. Let the set $\mathscr{S}$ be the set of all subgraphs of G generated by including each edge independently with probability x. There are 2^M elements of this set, and a subgraph s with k edges has a probability of occurrence given by $p_s(x) = x^k(1 - x)^{M-k}$.

Now consider a binary function $r : \mathscr{S} \to [0, 1]$. If $r(s) = 1$, we say that subgraph s is *accepted* by the rule r, or that s is *reliable*. We define the *reliability* $R(G, r, x)$ of a graph G with respect to the acceptance criterion r for a probability of edge failure $1 - x$ as the probability of choosing a subgraph s that will be accepted. Denoting the

set of all subgraphs accepted under r as $\mathcal{R}(G, r) \subseteq \mathcal{S}$, we have:

$$R(G, r, x) \equiv \sum_{s \in \mathcal{S}} r(s) p_s(x) = \sum_{s \in \mathcal{R}} p_s(x). \tag{6.1}$$

We will generally include the dependence on the graph G and the rule r in notation such as $R(G, r, x)$ explicitly only when we wish to distinguish the reliability of two different graphs or two different rules.

6.2.1 Reliability Criteria

Following [4], we describe a few different rules that capture properties important to diffusion dynamics. They overlap with, but are not the same as, the rules he describes.

1. *two terminal reliability*: a subgraph is accepted if it contains at least one directed path from a distinguished vertex S (the *source*) to another distinguished vertex T (the *terminus*);
2. *component size*: a subgraph is accepted if the number of vertices connected to the source is greater than or equal to a specified value;
3. *percolating*: a subgraph is accepted if it contains a path from a source to at least one of the vertices at maximum distance from that source, i.e. the "other side" of the network;
4. *multi-source*: any of the above problems with multiple source vertices;
5. *time dependent*: any of the above rules in which only paths no longer than a given time t are considered in evaluating the rule.

We will make the reasonable assumption that the rule is *coherent*: that if $s \subseteq t$ and $r(s) = 1$, then $r(t) = 1$. In other words, for a coherent rule, adding edges to a reliable graph does not make it unreliable. Other than this assumption, none of what follows is specific to the rule.

6.2.2 The Reliability Polynomial(s)

It is simple to write down a polynomial form for the reliability as a function of x. Partition the set of subgraphs $\mathcal{S}$ into subsets $\mathcal{S}_k$ in which each subgraph has exactly $k \leq M$ edges. There are $\binom{M}{k}$ subgraphs in $\mathcal{S}_k$ and each subgraph appears with probability $p = x^k (1-x)^{M-k}$. Suppose $R_k \equiv |\mathcal{R} \cap \mathcal{S}_k|$ of them are accepted by rule r. Then the total contribution of subgraphs in $\mathcal{S}_k$ to $R(x)$ is $R_k x^k (1-x)^{M-k}$. Summing these contributions over all k gives the *reliability polynomial* (for rule r and graph G):

$$R(x) = \sum_{k=0}^{M} R_k x^k (1-x)^{M-k}. \tag{6.2}$$

It is often convenient to express the k-dependence of R more compactly as a function of $y \equiv x/(1-x)$, in the form

$$R(y) = (1+y)^{-M} \sum_{k=0}^{M} R_k y^k. \tag{6.3}$$

Note that the coefficients R_k, being counts, are non-negative integers.

A great deal is known about reliability polynomials. They are related to the Tutte polynomial and the partition function of a Potts model. Colbourn [4] provides a wonderful introduction to reliability with an emphasis on the computational complexity of evaluating the coefficients R_k and bounding their values. The well-known Max-Flow/Min-Cut theorems discovered by Ford and Fulkerson [6, 7] relate certain of the coefficients. While individual coefficients may be efficiently determined for some classes of graphs, Colbourn notes that the problem of determining all the coefficients of the reliability polynomial for a general graph is #P-complete for many commonly used rules. Page and Perry [10] developed methods to assess the relative importance of edges in a graph, beginning with the differences in reliability illustrated here.

6.2.3 Reliability as a Polynomial Transform

It is instructive to think of the reliability polynomial as a generating function for the coefficients R_k, i.e.

$$R_k = \frac{d^k}{dy^k}(1+y)^M R(y)|_{y=0}. \tag{6.4}$$

Alternatively, the coefficients R_k distill the information in $R(x)$ that depends on the graph G and not x. As with other transformations, e.g. Fourier transformation, we can speak of the k-space and x-space representations of R. It may be the case that information that is spread over many values of x is concentrated in values of R_k for only a few k, and *vice versa*.

At first glance, nothing is accomplished by rewriting $R(x)$ as R_k. The amount of work required to determine $R(x)$ for all x is the same as the amount of work required to evaluate all the coefficients R_k. Furthermore, because of the strong envelope created by the binomial function, $R(x)$ is determined to a large extent by R_{Mx}. However, as we illustrate below, thinking in k-space rather than x-space suggests new proof strategies and styles of argumentation. Whereas perturbative analysis emphasizing local structures and links between vertices is natural for $R(x)$, non-perturbative analysis emphasizing global structures and large-scale features of subgraphs is natural for R_k.

6.2.4 Interpreting Reliability Coefficients

Individual coefficients R_k have direct interpretations in terms of graph structures. For two-terminal reliability, $R_k = 0$ for all k less than the shortest path length d between the source and terminus, and the value of R_d is the number of distinct shortest paths. For k larger than the minimum cut size, all graphs will be accepted, and thus $R_{k'} = \binom{M}{k'}$ for $k' \geq c$. For all-terminal reliability, the minimum k for which $R_k > 0$ is the size of the minimal spanning tree, and its value is the number of such trees.

In general, we expect that a reliability analysis of any problem will draw attention to certain values of k, i.e. to structures in subgraphs of certain specific sizes k'. Examples in Sect. 6.5 illustrate analyses that do this. The focus on k-space leads to useful explanations of otherwise nearly incomprehensible dynamical phenomena. For example, consider two graphs whose reliability polynomials cross. Crossing means that the relative reliability of the two graphs depends on the likelihood of an edge failure. This has immediate and important implications for designing failure-resistant networks. In the context of epidemiology, for example, a public health intervention can be represented as a change in contact network weights and/or structure. If the reliability polynomials for the two contact networks cross, an intervention that is *better* than doing nothing for some values of transmissibility will be *worse* than doing nothing for others. In x-space, crossing behavior has been noted and discussed in studies of network diffusion e.g. [9]. In k-space, it has an immediate analogue: R_k values that cross at a particular value k'. The interpretation is that a certain graph structure of size k' that appears in one graph but not the other contributes in an important way to the overall reliability of the graph. A deeper analysis can identify the nature of that structure and design new interventions that take advantage of this knowledge. Thus, reliability analysis can be thought of as a tool to identify these structural motifs (possibly extremely large ones) that are most relevant to particular dynamical processes.

6.2.5 Inhomogeneous Weights: A Useful Special Case

The simple polynomial expression for reliability in Eq. 6.2 results from the assumption of homogeneous weights. Relaxing this assumption leads to some interesting possibilities. Suppose there is a small number q of different classes of edges with different failure probabilities. For example, a structured population model of epidemiology might distinguish transmission rates between people in $q = 4$ different age groups or susceptibility classes; likewise, a model of coupled infrastructures might consist of interactions among a few different kinds of infrastructure, such as electrical transformers, water pumps, etc. The reliability polynomial in this case will become a multinomial in $x_1, \ldots, x_q$, and the coefficients will be functions of the number of edges of each type, $k_1, \ldots, k_q$.

In Sect. 6.5 we explore the difference in reliability between two graphs $g_1(V, E_1)$ and $g_2(V, E_2)$ by distinguishing $q = 3$ classes of edges. Applications include study-

ing graphs whose reliability polynomials cross, as discussed in Sect. 6.2.4, or analyzing the effect of popular edge-swapping techniques e.g. [5].

If $g_2 \subset g_1$, we proceed by partitioning the edges E_1 that appear in g_1 into two parts: E_2, the ones that appear in g_2; and $E' = E_1 - E_2$, the ones that don't. As usual, M_1 is the number of edges in g_1, and M_2 the number in g_2. We assign all edges in E' the weight x as before, but edges in E_2 are assigned the weight εx. This turns an inherently discrete change from one graph to another into a continuous deformation, and introduces the possibility of local stability/sensitivity analysis. The resulting reliability multinomial can be written as

$$R(x, \varepsilon) = \sum_{k=0}^{M} x^k (1 - x)^{M-k} \sum_{k'=0}^{\min(k,M_2)} R_{k,k'} \varepsilon^{k'} (1 - \varepsilon)^{k-k'}. \tag{6.5}$$

When g_2 is not a subgraph of g_1, we form a convex combination of the two graphs instead. In this case, we partition the set of all edges $E_1 \cup E_2$ into *three* parts:

1. the common edges $E' \equiv E_1 \cap E_2$;
2. $E_1' \equiv E_1 - (E_1 \cap E_2)$, the complement of E_2 in E_1, or the edges that appear in g_1 but not in g_2;
3. E_2' defined analogously for g_2.

As before, we assign all edges in E' the weight x and all edges in E_1' the weight εx, but now there is a third class of edge, those in E_2'. We assign these edges the weight $(1 - \varepsilon)x$. The parameter ε allows us to continuously deform the graph from g_1 for $\varepsilon = 1$ to g_2 for $\varepsilon = 0$. This trick also works for graphs with heterogenous weights. The resulting reliability multinomial can be written as

$$R(x, \varepsilon) = \sum_{k=0}^{M} \sum_{k'=0}^{\min(k,M_1+M_2)} R_{k,k'} x^k (1 - x)^{M_1-k} \varepsilon^{k'} (1 - \varepsilon)^{-k'}. \tag{6.6}$$

In this expression, k' represents the number of edges from E_1' that are in the subgraph and the number of edges from E_2' that are *not* in the subgraph.

In principle, we could study the reliability of this convex combination of graphs directly. However, for a given value of ε, subgraphs of the convex combination contain a confusing mixture of edges from g_1 and edges from g_2. In practice, it is easier to understand the difference between reliabilities for two related graphs A and B: in both of the two graphs, the common edges E' have weight x, but A gives the edges E_1' weight εx while B gives the edges E_2' weight εx. Derivatives of $R^A(x, \varepsilon) - R^B(x, \varepsilon)$ with respect to x and ε are valuable for understanding trade-offs between increasing component reliability and restructuring the network—for epidemiology, between pharmaceutical prophylaxis and social distancing. Analysis of $R_{k,k'}^A - R_{k,k'}^B$ can be used to identify graph structures that significantly influence dynamical phenomena.

6.3 Statistical Physics of Reliability

A graph's reliability polynomial is related to a Potts model's partition function defined on the graph. This connection provides another important perspective on characterizing graph structure. We show the connection explicitly here.

6.3.1 Relation to Ising Model

Equation 6.1 defines reliability as a sum over possible system configurations weighted by the probability of occurrence for each state. Equations 6.2 and 6.3 group system configurations into equi-probability subsets. Thus the reliability function itself is a partition function, the relative probability y^k is related to the energy of a system configuration in a canonical ensemble, and the coefficients R_k are the density of states[1]:

$$Z(y) \equiv \sum_k R_k y^k = R(y).$$ (6.7)

To make contact with the usual partition function for an Ising model, we re-interpret subgraphs with k edges in terms of vertex "states", failed or not. The presence of an edge represents the conditional propagation of failure, conditioned on one of its endpoints having failed. On this view, vertices have two states (the state, s_i, of vertex i is defined to be 1 if it has not failed and -1 if it has), and there is an edge between any pair of vertices in the same state. Finally, we consider the complement of the graph. It will have edges only between vertices that are in different states, and occurs with probability given by replacing x with $1 - x$ or, equivalently, y with y^{-1}. Then emphasizing the vertex states instead of the edges, we can rewrite the factor y^k in Eq. 6.7 in terms of an energy for each configuration which is the sum over the number of frustrated spins (neighboring vertices with different states) of $J \equiv -\frac{1}{4} \ln y$:

$$y^k = e^{k \ln y} \propto \exp -J \sum_{i=1}^{N} \sum_{v_j \in \mathcal{N}(v_i)} s_i s_j.$$ (6.8)

The constant of proportionality can be absorbed into the overall normalization factor, leaving (not surprisingly) an Ising model.

---

[1] The overall factor $(1 + y)^{-M}$ in Eq. 6.3: is a normalization that will be ignored in the partition function, where only relative probabilities are needed.

6.3.2 The Density of States

What of the density of states, R_k? We can further decompose it into two factors, taking

$$R_k = P_k \binom{M}{k} \qquad (6.9)$$

as a definition of P_k. This decomposition splits R_k into what we might call an *entropic* or combinatorial factor $\binom{M}{k}$ and a *structural* factor P_k. The entropic factor simply makes explicit the sharp peak in the number of possible subgraphs with k edges, i.e. the size of the space from which equi-probable system configurations can be drawn. It creates an envelope that windows the effects of $x^k(1-x)^k$ to a small region centered at $k = Mx$. The factor P_k is structural in the sense that it encodes all the information about the specific graph G that is needed to determine its reliability. This is obvious because the probability y^k and the entropic factor $\binom{M}{k}$ are the same for any graph with M edges. The meaning of P_k is also clear—it is the fraction of possible subgraphs with given k that are accepted by the reliability criterion. We can think of this as the probability that a random subgraph with k independently selected edges is reliable. Re-introducing independent edge selection in this way makes analysis easier without sacrificing correctness.

We have shown that all the structural information needed to characterize G is contained in P_k. We claim further that P_k is not only sufficient but necessary. If any of the P_k are unknown, the reliability can only be approximated. Of course, the approximation may be very good for probabilities x far outside the entropic envelope, but the reliability will not be uniformly approximable on the interval $[0, 1]$. Finally, the structural properties encoded in P_k are, by the design of the reliability criterion, exactly those relevant to the dynamical phenomena under study. This is in stark contrast to characterizing a graph by static, statistical properties such as degree distribution, clustering, etc.

It cannot be denied that the coefficients R_k carry identical information to the P_k. Which is the "better" representation is thus largely a matter of taste. We argue here that giving primacy to P_k is a useful perspective because it strips out all the features common to all graphs, leaving only structural information about the specific graph G that is relevant to the rule r.

6.3.3 Partition Functions and Ensembles

A system's partition function serves to normalize the relative probabilities of different system configurations (here, subgraphs g). The expectation value of any function of the configurations $< f >$—e.g. the average size of the largest connected components in reliable subgraphs, or the number of triangles in the graph, or the reliability itself—can be computed with respect to the resulting normalized probability:

$$< f > = \sum_{k} \sum_{g_k} f(g_k) r(g_k) x^k (1-x)^{M-k} \qquad (6.10)$$

There is a great deal of latitude in decomposing $R(x)$ into the product of a density-of-states factor and a probability-of-state factor. This latitude in specifying the probability factor is usually referred to as choosing the ensemble of system configurations.

In a grand canonical ensemble, we might consider the probability factor to be $\binom{N(N-1)/2}{k}$, the probability of choosing any graph on N vertices with k edges. The partition function for this case would be:

$$Z_{grand} \equiv \sum_{k} \sum_{g_k \subset N \times N} x^k (1-x)^{\binom{N}{2}-k}. \qquad (6.11)$$

Calculating $< f >$ over reliable graphs correctly in a grand canonical ensemble requires adding two filters to the density of states: one that excludes graphs that are not subgraphs of G and another that excludes graphs rejected by the reliability criterion: $f(g) \rightarrow f(g)\delta(g \subset G)\delta(r(g))$. In textbook derivations, e.g. [11] these delta function filters are approximated by exponentials and included in the system Hamiltonian. This approach woks well for constraints involving things like energy that can be represented in closed form as functions of the system's configuration, but it is not well-suited for the reliability filter $\delta(r(g))$ or the subgraph filter because the constraints they represent cannot be written in closed form.

We could include the constraint that a graph must be a subgraph of G in the probability factor, as we have done in this chapter, generating a canonical ensemble whose partition function is:

$$Z_{canon} \equiv \sum_{k} \sum_{g_k \subset G} x^k (1-x)^{M-k}. \qquad (6.12)$$

Calculating $< f(g) >$ over reliable graphs in a canonical ensemble would require including only the reliability filter in f: $f(g) \rightarrow f(g)\delta(r(g))$.

Finally, a micro canonical ensemble includes only those subgraphs of G that are reliable. Its partition function is

$$Z_{micro} \equiv \sum_{k} \sum_{g_k \subset G} r(g_k) x^k (1-x)^{M-k}. \qquad (6.13)$$

No adjustments to f are needed to evaluate its expectation over reliable graphs in this ensemble. Clearly the partition function for the micro canonical ensemble *is* the reliability polynomial itself.

However, there is no free lunch—the progression from grand to canonical to micro-canonical ensembles simply shifts where the effort must be applied, from evaluating the reliability or subgraph filters to generating the ensemble. The canonical ensemble is more tractable analytically, being nothing more than a binomial dis-

tribution. Computationally, the micro-canonical ensemble is generally much more efficient, since no time is "wasted" computing f on system configurations that do not contribute to the expectation value.[2] The reliability polynomial provides an analytically tractable way to handle the micro-canonical ensemble by pretending that we know the coefficients R_k. The potential value of this approach is two-fold:

1. analysis may show that only a few coefficients are needed to find an approximate solution to a given problem;
2. the coefficients may well be approximated by functions with only a few parameters.

6.3.4 Phase Transitions

As usual, we can define the location of a phase transition as the point at which $d^2 \ln Z/dx^2$ vanishes. Since all of the x-dependence is contained in the binomial factor $B_k(x) \equiv \binom{m}{k} x^k (1 - x)^k$, we can apply the derivatives to that factor using the useful identity Eq. 6.25 and recover an expression for the location of the phase transition in reliability as zeroes of the polynomial $J(x) \equiv \sum_{k=0}^{M} J_k B_k(x)$, where

$$J_{k-1} \equiv (M - k)(M - k - 1)(P_{k+2} - 2P_{k+1} + P_k). \tag{6.14}$$

Note that J_k vanishes when $P_k = P_{k+1} = P_{k+2}$, that is, when $R_k \propto \binom{M}{k}$, or when the second finite difference of P vanishes, i.e. P_k is changing at a constant rate. The coefficients that remain are the ones where global innovations to the set of reliable subgraphs become important. The exact location of the zeroes is an interplay between the strong enveloping effect of the binomial distribution centered at $k = Mx$ and sudden changes in the fraction of reliable subgraphs as k varies.

If we make an ansatz for values of P_k, we can write down an analytical expression for the critical point. Coherence implies that the P_k increase monotonically. Suppose the rate of increase peaks at k_c. A procedure to generate a parameterized family of sigmoidal functions that reproduce this behavior is given in Appendix 2. This particular family has a parameter, k_c that determines the center of the sigmoid, and another, N, determines the width of the sigmoid.

Combining this ansatz with the identity (see Appendix 1)

$$\sum_{k=0}^{m} k^q B_k(x) = \sum_{l=0}^{q-1} \frac{m!}{(m - l)!} x^{q-l} \tag{6.15}$$

yields a polynomial in x with one root in the interval $(0, 1)$ that is the critical point. Figure 6.1 displays the envelope $d^2 B_k(x)/dx^2$ as a function of k for two values of x, and $R''(x)$ for two choices of k_c and three choices of N. Figure 6.2 exhibits the

---

[2] This rule of thumb may depend on the computational complexity of the reliability criterion.

Fig. 6.1 *Top* The envelope determined by the second derivative of the binomial distribution for two different values of x. Note that for small values of x, the envelope is distinctly asymmetric; *bottom* The second derivative of the reliability polynomial, $R''(x)$, for each of the models for P_k generated by the sigmoidal functions described in Appendix 2, for M = 20 edges. These values of P_k are not derived from a specific graph, but show the effect of varying the sigmoid's center, controlled by k_c, and width, controlled by N. A phase transition occurs at the zero of this function. The zero is given by k_c/M to a first approximation, but is affected by the width of the sigmoid
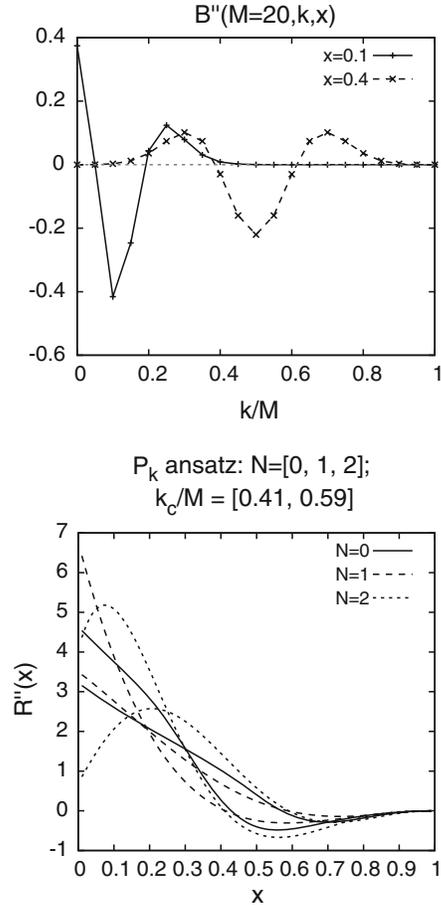

details of $R''(x)$ around its zeroes. Clearly, we could derive a functional form relating the critical point and critical exponents with k_c and N. Refinements and extensions of this procedure can easily be imagined, including a more general ansatz for P_k, approximating the binomials with Poisson distributions, and expanding around the approximate solution $x_c \approx k_c/M$.

6.4 Reliability-Based Methods for Analyzing Graph Structure

We will outline here two related approaches to evaluating the effects of graph structure on reliability. The first is a bottom-up approach that analyzes the number of new, distinct subgraphs generated by adding an edge to an acceptable subgraph; the second is a top-down approach that analyzes the impact of reliable motifs. They both seem to be complementary to—and distinct from—common approaches like mean-field theory, low- and high-temperature expansions, etc. They permit the introduction of new principled expansions in a small parameter that is not the coupling constant.

Fig. 6.2 Detailed view of the
zeroes of $R''(x)$ as shown in
the previous figure. *top k_c =
0.41M; bottom k_c = 0.59M*
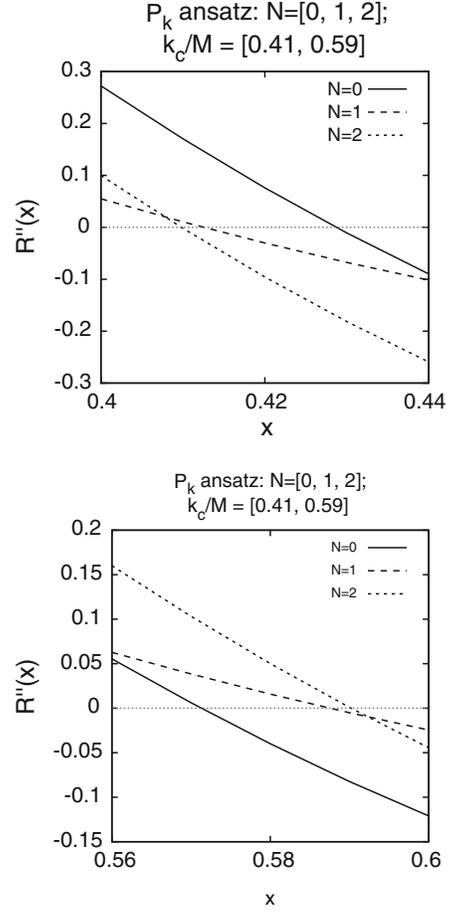
First, we formalize the intuitive notion of distance between two graphs on the
same vertex set $g_1(V, E_1)$ and $g_2(V, E_2)$:

$$d(g_1, g_2) = |E_1 \cup E_2| - |E_1 \cap E_2|, \tag{6.16}$$

i.e. the number of edges in either g_1 or g_2 that are not common to both. Note that if
g_1 and g_2 have the same number of edges, then $d(g_1, g_2)$ is an even number, and if
we construct the sets of edges $F_i \subseteq E_i$ that are in one graph g_i but not in the other,
then $|F_1| = |F_2|$, i.e. the distance is evenly split between the graphs.

Let the set of accepted subgraphs with exactly k edges be denoted $\mathscr{R}_k$. In the
bottom-up approach, we partition the set of pairs of reliable subgraphs $\mathscr{R}_k \times \mathscr{R}_k$ into
non-overlapping subsets that structure relationships among the coefficients R_k for
different values of k. Our partition is based on distances between graphs:

$$\mathscr{D}_n^{(k)} \equiv \{(g_1, g_2) \in \mathscr{R}_k \times \mathscr{R}_k | d(g_1, g_2) = n\}. \tag{6.17}$$

We denote the cardinality of $\mathscr{D}$ by d, that is, $d_n^{(k)} \equiv |\mathscr{D}_n^{(k)}|$. Since every pair falls into one of the sets $\mathscr{D}$, we know that $\sum_{n=2}^{2M} d_n^{(k)} = \binom{R_k}{2}$.

How are $d_n^{(k)}$ and $d_n^{(k+1)}$ related? Consider a pair of subgraphs $(g_1, g_2) \in \mathscr{D}_n^{(k)}$. If we add the *same* edge to both, the distance between the new graphs will be the same as between the old pair. There are $M-k-1$ ways this can be done. If we add to g_1 an edge that was only present in g_2 and *vice versa*, the distance between them will *decrease* by two. There are $\frac{n^2}{4}$ ways this can be done. All other $(M-k)^2 - (M-k-1) - \frac{n^2}{4}$ choices *increase* the distance between the graphs by two. Since not all graphs generated in this process will be distinct, this provides only upper bounds on the contributions of individual graphs to $d_n^{(k+1)}$.

Now consider the overall contribution of subgraphs in $\mathscr{R}_k$ to $\mathscr{R}_{k+1}$. Adding a single one of each of the $M - k$ possible new edges to a graph in $\mathscr{R}_k$ generates $M - k$ subgraphs that, by the principle of coherence, will be elements of $\mathscr{R}_{k+1}$. However, some of these new subgraphs will be generated by more than one of the graphs in $\mathscr{R}_k$. How many? Suppose we add an edge to g_1, creating g_1', and similarly for g_2. It is easy to see that g_1' and g_2' can be identical if and only if $d(g_1, g_2) = 2$. In that case, there is exactly one choice of edge to add to g_1 and one choice to add to g_2 so that $g_1' = g_2'$. Hence the number of new subgraphs in $\mathscr{R}_{k+1}$ generated by a single graph $g \in \mathscr{R}_k$ is $M - k$ *less* the number of times g appears in $\mathscr{D}_2^{(k)}$. Summing over all $\tilde{R}_k$ subgraphs in $\mathscr{R}_k$ gives $(M - k)R_k - d_2^{(k)}$.

We can define a set of $\tilde{R}_k$ subgraphs $\tilde{\mathscr{R}}_k$ as those generated by adding a single edge to graphs in $\mathscr{R}_{k-1}$. $\tilde{R}_{k+1}$ provides a lower bound on R_{k+1}. It is combinatorial, in the sense that no new (i.e. not counting $d_2^{(k)}$) structural information is included in the estimate. Thus the new structural information is all to be found in two places:

1. the difference between the bounds on $d_n^{(k+1)}$ generated above and the actual partition of graphs generated from $\mathscr{R}_k$;
2. what we call the *innovation* $\mathscr{R}_k - \tilde{\mathscr{R}}_k$.

The first of these is related to local structure around existing reliable paths, such as clustering; the second involves global structure. Subgraphs in the innovation will appear in sets $\mathscr{D}_n^{(k+1)}$ only for $n \geq 4$. As we increase k, we can track the appearance of innovations and watch as they inexorably coalesce with other subgraphs—the reliability polynomial keeps track of how their contribution overlaps with that of other structures at every step of the process.

In the top-down approach, we count the number of times a particular motif will appear among all subgraphs of size k. The motif is a subgraph whose presence is sufficient to make any graph containing it reliable. For example, for two-terminal reliability, the motif of interest might be a shortest path between the source and terminus; for all-terminal reliability, it might be a spanning tree. Denote the number of edges in the motif by k'. Then for subgraphs of size $k < k'$, the motif cannot appear. However, the motif is guaranteed to appear in some of the subgraphs of size $k \geq k'$. How many? Consider coloring all k' edges in the motif red and placing them in a bag with all the rest of G's edges, which are colored black. We reach into the bag

and select k edges in all possible ways. This is the enumeration over subgraphs of G with k edges. How many times will the k edges we select include all k' of the red edges? The answer is given by the multinomial coefficient $M!/(M-k)!(M-k-k')!$. Once k is large enough that $R_k > 0$, we know that a reliable motif exists, and this argument provides a lower bound on R_k for higher order coefficients.

This argument can be extended to the case of multiple motifs. If the reliable motifs are entirely edge distinct, the probability that a random subgraph includes one of them is independent of the probability that it includes the other, so we can count the number that include one or the other or both. Similarly, if the distance between the motifs is d, we can also easily adjust the over-counting incurred by summing their individual contributions to the reliability. The bottom-up and top-down approaches are linked together by the knowledge of the innovation and how distant its elements are from existing reliable subgraphs.

These approaches provide new tools to understand the effect of structural changes on reliability. Consider, for example, swapping edges to change the number of triangles in a graph while holding the degree distribution and assortativity-by-degree fixed. For two-terminal reliability, the increased number of triangles reduces the number of innovations (which are due to subgraphs with at least four edges that are "new") and increases the number of distance-two subgraphs.

For most of this discussion, we have assumed the graphs are known, and we are attempting to characterize their structure. Reversing these arguments, we can infer from values of R_k for a range of k's how many reliable motifs with sizes in that range are present. This can form the basis of network tomography, in which we infer the existence of hidden graph structure (e.g. the existence and number of shortest paths) from information about reliability at a few vertices whose states are accessible to observation.

6.5 Examples

We illustrate the discussions above with evaluations and interpretations of the reliability polynomial, the difference in reliability between graphs, and the effect of assortativity-by-degree and number of triangles on several networks.

6.5.1 Toy Network

A networked *SIR* process is a Markov process over system configurations. If $p(c, t)$ denotes the probability of finding the system in configuration c at time t, then we have $p(c', t+1) = \sum_c p(c'|c)p(c, t)$. We will assume uniform edge weights x—that is, each infectious vertex has the same probability of transmitting infection to each of its neighbors. We use the usual rules for transmission: an infectious vertex

Table 6.1 The probabilities of several configurations for the graph in Fig. 6.1 as a function of time. Only configurations that eventually contribute to the two-terminal reliability are shown

Time	s_1	s_2	s_3	s_4	s_T	Probability
1	I	I	S	S	S	x^2
1	I	S	S	S	S	$x(1-x)$
1	S	I	S	S	S	$x(1-x)$
2	R	R	I	I	S	$x^4(2-x)$
2	R	R	I	S	S	$x^3(1-x)(2-x)$
2	R	R	S	I	S	$x^3(1-x)^2$
2	R	S	I	I	S	$x^3(1-x)$
2	R	S	I	S	S	$x^2(1-x)^2$
2	R	S	S	I	S	$x^2(1-x)^2$
2	S	R	I	S	S	$x^2(1-x)$
3	R	R	R	R	I	$x^5(2-x)^2$
3		10 more combinations ...				
5	R	R	R	R	I	$x^5(1-x)^2$

transmits independently to each of its neighbors, and a susceptible vertex is infected independently by each of its infectious neighbors. That is:

1. the probability of a transition from one configuration to another is the product of the probabilities for each vertex to be in the given state, because the event of infecting a vertex is independent of the event of infecting a different vertex.
2. the probability that a vertex is *not* infected depends only on the number of its infectious neighbors, as $(1-x)^n$.

Given these transmission rules, we can evaluate the transition matrix $p(c'|c)$.

Table 6.1 lists probabilities for example configurations as a function of time for the toy network shown in the left panel of Fig. 6.3. The two terminal reliability expressed as the cumulative sum of these probabilities is

$$R(x) = x^5(2-x)^2 + x^4(1-x)(5-3x) + 2x^3(1-x)^2 + x^3(1-x) + x^5(1-x)^2. \quad (6.18)$$

There is no particular interpretation for any of these coefficients, nor any obvious way to simplify the expression.

By contrast, we determine the reliability coefficients for the same two-terminal reliability by counting accepted subgraphs, as shown in Fig. 6.4. We can write down the reliability polynomial by inspection as

$$R(x) = x^3 \left\{ 3(1-x)^4 + 12x(1-x)^3 + 17x^2(1-x)^2 + 7x^3(1-x) + x^4 \right\}. \quad (6.19)$$

The result in Eq. 6.18 can indeed be reduced to this form through some tedious algebra. In this form, we can immediately read off structural information. For example, the coefficient of the first term in braces, 3, is the number of distinct shortest paths
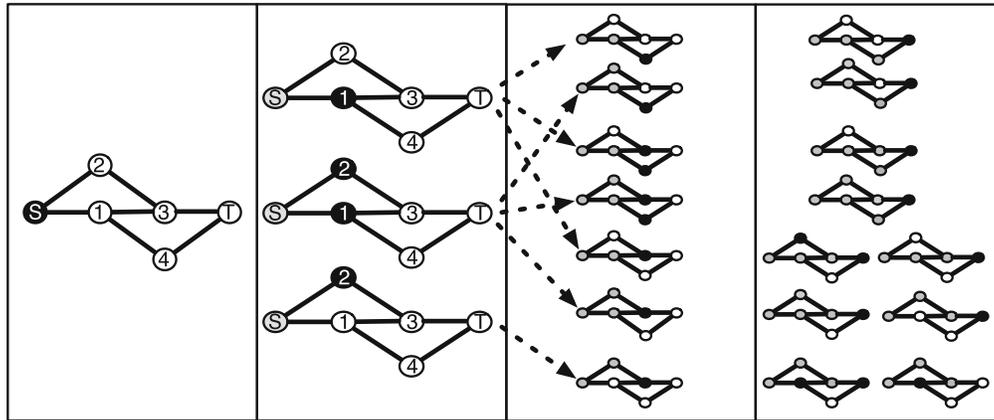
Fig. 6.3 Example vertex-centric calculation of two-terminal reliability in a toy network between the source S and terminus T vertices. Each panel shows the possible configurations at a different time step, ignoring those configurations that will not contribute to the reliability. The vertex color corresponds to its state: *white* for Susceptible, *black* for Infectious, and *gray* for Recovered. The single diagram at the *bottom right* contributes to a length-five path from S to T that is not shown here. Probabilities for each of these configurations are given in Table 6.1
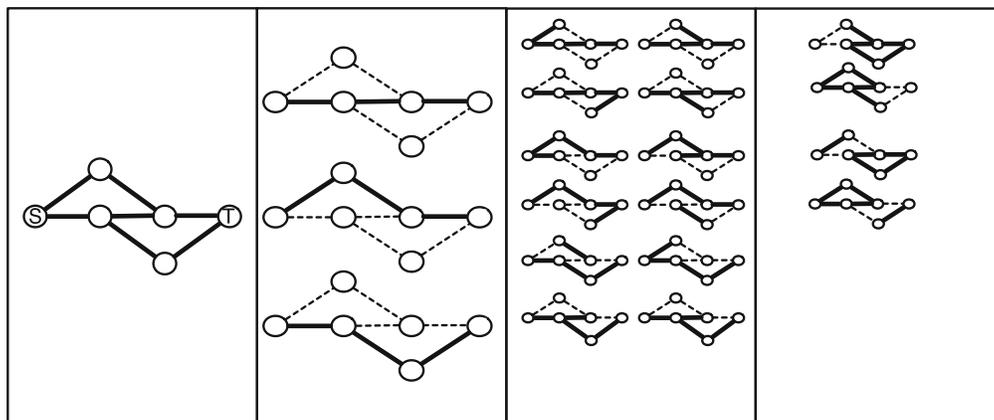

Fig. 6.4 *(left)* Example edge-centric calculation of two-terminal reliability in the toy network of Fig. 6.3. Each panel shows the accepted subgraphs with specific numbers of edges, except the panel on the *right*, which shows the subgraphs that are *not* accepted

from S to T; the coefficients of $x^3(1-x)$ and x^4 are $\binom{7}{6}$ and $\binom{7}{7}$, because the minimum cut size is 2; the coefficient of $x^2(1-x)^2$ is $\binom{7}{5} - 4$, because there are 4 cuts with cut size 2. The three reliable subgraphs for $k = 3$ form three possible pairs; of these, two pairs of graphs are at distance $d^{(3)} = 4$ from each other and one (the center and bottom graphs in the panel) is at distance 6.

Now let us compare the reliability of two graphs obtained by removing either (1) the center edge $(1, 3)$ for G_1 or (2) the lower left edge $(1, 4)$ for G_2 from the toy network. Using the approach described in Sect. 6.2.5, we construct two interpolating graphs by weighting the selected edge with a factor ε. The non-trivial reliability

Table 6.2 Two terminal reliability coefficients for two networks formed by weighting an edge of the toy network by ε, as discussed in Sect. 6.2.5. Coefficients not shown here are either 0, for $k < 3$ or $\binom{7}{k+k'}$, for $k > 4$

k	k'	$R_{k,k'}(G_1)$	$R_{k,k'}(G_2)$
3	0	2	2
3	1	6	6
4	0	6	6
4	1	16	17

coefficients for both these interpolating graphs are shown in Table 6.2. By inspection, the difference in reliability of the two graphs is

$$R_{G_2}(x) - R_{G_1}(x) = \sum_{k=0}^{M-1} \sum_{k'=0}^{1} R_{k,k'} x^k (1-x)^{M-k-1} \varepsilon^{k'} (1-\varepsilon)^{(1-k')} \qquad (6.20)$$

$$= x^4 (1-x)^2 \varepsilon \qquad (6.21)$$

Repeating the analysis above for other edges, we can determine the partial derivatives of $R(x)$ with respect to any single edge weight or any set of edge weights. We consider, as before, the two terminal reliability from S to T, but only include subgraphs with paths of length at most three. As mentioned in the discussion of example criteria, this corresponds to a time-dependent reliability function. (That is why the partials for edges $(1,3)$ and $(1,4)$ below differ from those that would be found using Table 6.2.) These partials are:

$$\begin{aligned}
\partial R(x)/\partial x_{S1} &= x^2(2 - 2x^2 - x^3 + x^4) \\
\partial R(x)/\partial x_{S2} &= x^2(1 - x^2 - x^3 + x^4) \\
\partial R(x)/\partial x_{13} &= x^2(1 - 2x^2 + x^4) \\
\partial R(x)/\partial x_{14} &= x^2(1 - x^2 - x^3 + x^4) \\
\partial R(x)/\partial x_{23} &= x^2(1 - x^2 - x^3 + x^4) \\
\partial R(x)/\partial x_{3T} &= x^2(2 - 2x^2 - x^3 + x^4) \\
\partial R(x)/\partial x_{4T} &= x^2(1 - x^2 - x^3 + x^4)
\end{aligned} \qquad (6.22)$$

In this case, we see that the partials for edges $(S, 1)$ and $(3, T)$ are the same, as are the partials for $(S, 2)$, $(1, 4)$, $(2, 3)$, and $(4, T)$. Inspection of the graph shows that these groupings make sense. We also see that there is a partial order on the partials: $\partial R(x)/\partial x_{S1} > \partial R(x)/\partial x_{S2} = \partial R(x)/\partial x_{14}$, etc. This induces a partial ordering on the edges:

$$(S, 1) > (S, 2) = (1, 4) = (2, 3) = (4, T) > (1, 3) > (3T). \qquad (6.23)$$

It also allows us to define the gradient direction $(1, \alpha, \beta, \alpha, \alpha, 1, \alpha)$, where $\alpha \equiv 1 - \frac{1-x^2}{2-2x^2-x^3+x^4}$ and $\beta \equiv 1 - \frac{1-x^3}{2-2x^2-x^3+x^4}$, and the basis elements are the edges in order as they appear in Eq. 6.22.

The immediate interpretation of the gradient is that small changes to the weights distributed in the proportions given by the gradient will produce the largest change in $R(x)$, the vulnerability of T at time $t = 3$, of any possible distribution. In other words, this is the locally optimal combination of edges to change. Of course, it is not necessarily a global optimum, since the gradient will change as the weights change—this is just a linear analysis. Given the full expression of the gradient for inhomogeneous weights, we could follow the integral curve defined by the gradient at any point to a local optimum.

6.5.2 Estimation

Simulation provides approximate values for the probability of any configuration as a function of time. It is particularly well-suited for estimating reliability coefficients. We can view N simulations for a given edge weight x as N samples of a Bernoulli process with parameter $R(x)$. Thus we can place confidence bounds on the estimated $\hat{R}(x)$. Sweeping across $x \in [0, 1]$ provides data to which we can fit a polynomial, thus obtaining all the coefficients of the reliability polynomial. In principle, since we know the polynomial coefficients must be integer, we can determine *a priori* the number of samples that will be required to generate the exact result with high probability. Unfortunately, since the integers involved are combinatorially large, this provides little benefit in practice.

Any single coefficient P_k can be estimated more directly by a simulation that selects a subgraph with k edges and evaluates the reliability criterion. The computational complexity of this is proportional to k (not M), the number of samples selected, and the complexity of evaluating the criterion.[3] What became of the hardness results? Once again, there is no free lunch—they are implicit here in the number of samples needed to determine P_k with sufficient precision. However, note that this procedure is embarrassingly parallel, so one would expect linear speedups on distributed computers.

6.5.3 Edge Swapping

We have constructed a collection of graphs with the same mean degree, but different degree distributions, assortativity-by-degree, and clustering coefficients. Some examples of these graphs are shown in Fig. 6.5. We started with two base graphs:

---

[3] The complexity of the criterion itself must not be overlooked. In many cases, its evaluation requires partitioning the selected subgraph into connected components.
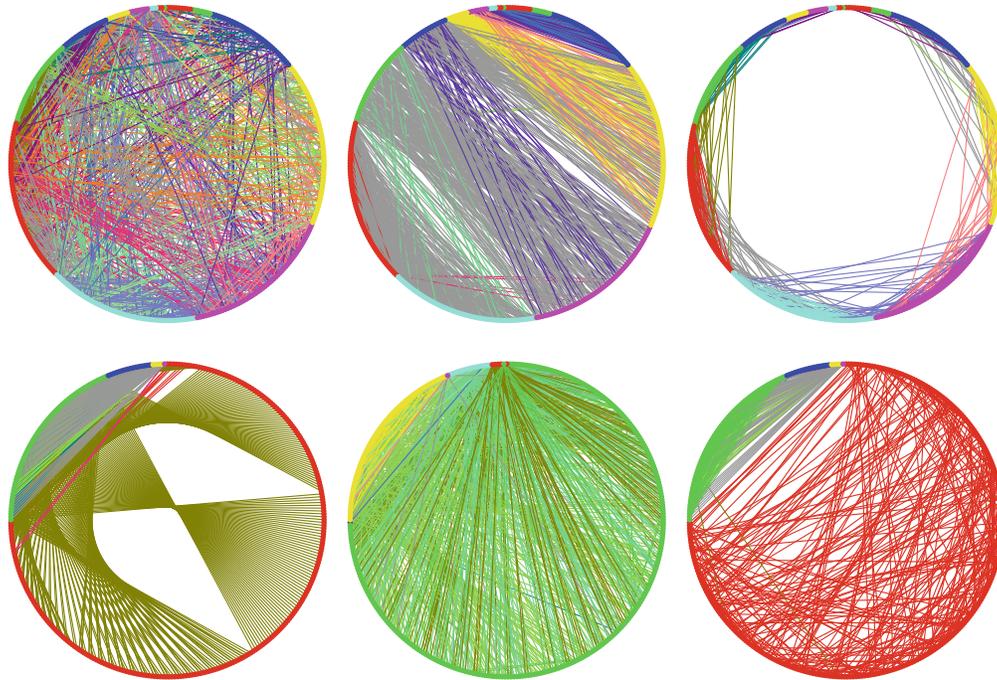
Fig. 6.5 Example graphs used in this edge swapping study. Vertices are colored by degree, and edges are colored by the degrees of the nodes at either end. Each row exhibits, from *left* to *right*, the initial graph, the lowest assortativity graph with the fewest triangles, and the highest assortativity graph with the most triangles. *Top row*, an Erdős-Rényi random graph, (*GNM*); *bottom row*, scale-free-like graphs (*SFL*)

1. one we call "scale-free-like" because the number of vertices with degree d is proportional to 2^{-d}, for $d \in \{4, 8, 16, 32, 64\}$, giving a total of 992 edges and 341 vertices;
2. an Erdős-Rényi random graph with 341 vertices and 992 edges (and therefore the same mean degree, $\sim$5.82, as the scale-free-like graph).

For each of these base graphs, we constructed other graphs by swapping edges in such a way as to preserve the degree distribution. Thus, in particular, the modified scale-free-like graphs all have a power-law degree distribution with exponent -2. Two different kinds of swaps were used: one kind ensures that the number of triangles in the graph changes monotonically (either increasing or decreasing); the other ensures that the assortativity-by-degree changes monotonically (again, either increasing or decreasing). These were both run in both directions for the scale-free-like and $G(N, M)$ base graphs. The resulting collection provides a laboratory for studying the effect of assortativity and clustering on graphs with different structures. It is similar to examining exponential family random graphs satisfying constraints on mean degree, degree distribution, clustering, and assortativity-by-degree, but we expect in addition to see traces of the initial structure in the graphs, a constraint that is hard to formalize mathematically. That is, instead of starting with a random graph and constraining various properties to have the values we want, we start with
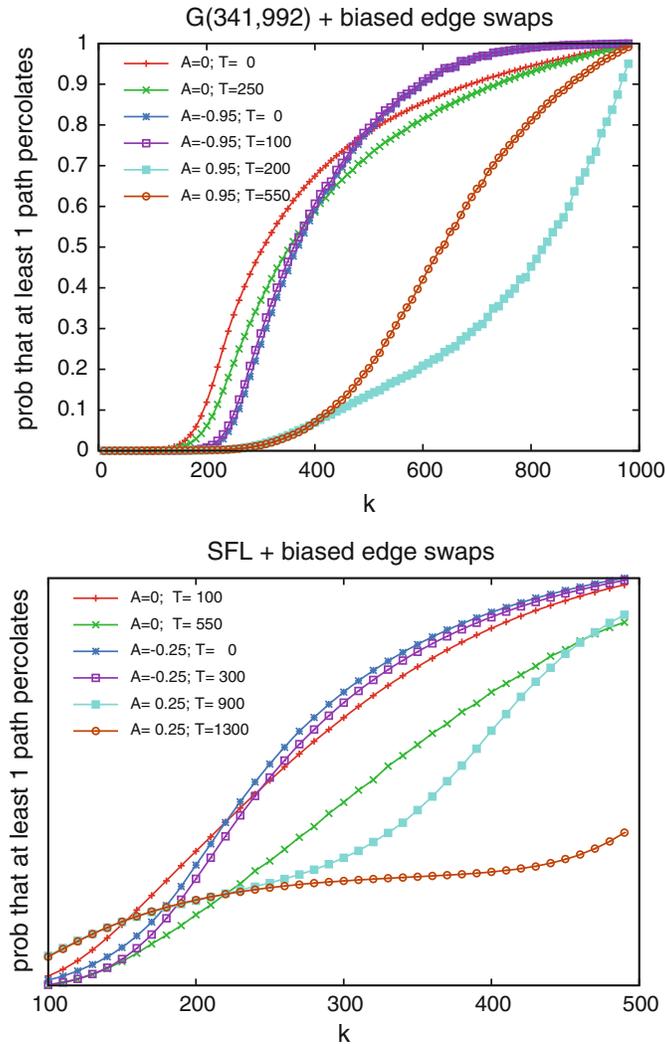
Fig. 6.6 *Top* The percolating reliability coefficients for a set of graphs derived from an Erdős-Rényi random graph by swapping edges to change assortativity and clustering while maintaining the degree distribution invariant. *bottom* The same, for scale-free-like graphs described in the text. In both panels, data are plotted for the most assortative and least assortative graphs as well as a neutrally assortative graph, for graphs with the largest and smallest numbers of triangles consistent with the assortativity and degree distribution. Lines are drawn between points to guide the eye

carefully structured graphs and randomize them while maintaining the same properties invariant. We intend to make this graph library readily available to researchers through the cyberinfrastructure for network science (CINET) web service at http://ndssl.vbi.vt.edu/cinet/cinetproject/ [1].

Figure 6.6 shows reliability coefficients for six matched pairs of graphs. In each pair the two graphs have the same assortativity-by-degree A and the same original structure, *SFL* or *GNM*, but extreme values of the number of triangles, T. Note that the magnitude of the effect of increasing the number of triangles depends strongly

Fig. 6.7 The difference in percolating reliability coefficients for two pairs of graphs whose individual coefficients are shown in Fig. 6.6. Each *curve* displays results for coefficients for subgraphs that include a different number of common edges k. Unlike other figures above, the total number of edges in the subgraph, $k + k'$, varies along each *curve*. Thus the reliability of each graph separately increases monotonically from *top* to *bottom*, but the difference between the two graphs does not

GNM(341, 992) A=0.95

SFL A=0.25
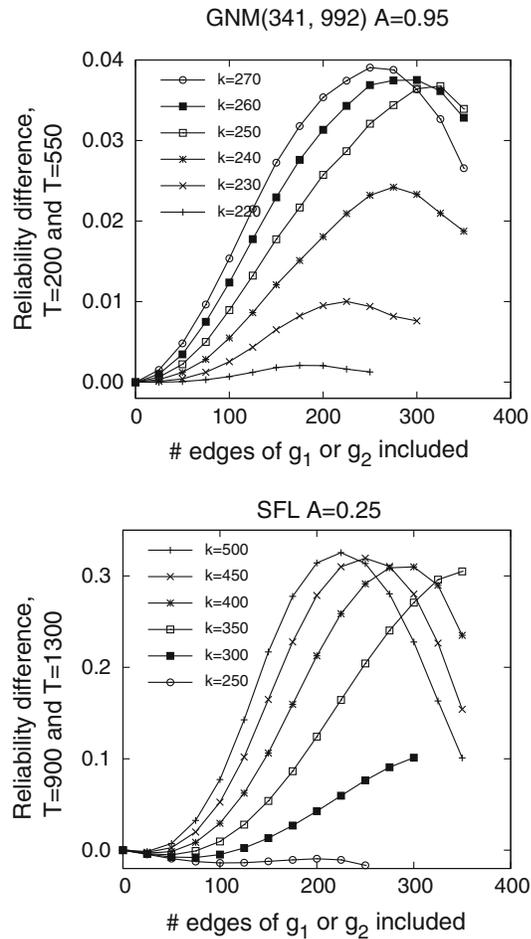
on the assortativity-by-degree: the more assortative the graph, the more pronounced the effect of triangles. Furthermore, the sign of the effect can vary: for five out of the six pairs shown here, the graph with more triangles generally has larger reliability coefficients. The exception is the neutrally assortative ($A = 0$) *GNM* graph. This is in mixed agreement with theoretical results e.g. [2, 9]

Furthermore, T's effect is noticeable at smaller k for neutrally assortative graphs than for either diassortative or assortative graphs. Also note that the curves for the scale-free-like graphs $A = 0$; $T = 550$ and $A = 0.25$; $T = 900$—i.e. the graphs with the largest (respectively, smallest) number of triangles for the neutrally (respectively, highly) assortative cases—cross at least twice, suggesting that their relationship in x-space is very complicated.

As discussed above, these results indicate the existence of a motif with about 200 edges that is not present in the highly assortative Erdős-Rényi-based graphs. In contrast, for the scale-free-like graphs, there is a motif with fewer than 100 edges that is present in the highly assortative graphs, but not the others. Its influence is overcome in the other graphs by other reliable motifs with sizes in the range 150–

225, depending on the assortativity and number of triangles. Likewise, for the highly assortative graphs, one or more motifs (with size approximately 400 edges for Erdős-Rényi and 225 edges for scale-free-like) distinguish the reliability for graphs with high and low numbers of triangles. The motif(s) appear in the graphs with many triangles for Erdős-Rényi-based graphs, but *vice versa* for scale-free-like graphs.

Figure 6.7 shows the difference in reliability for the most highly assortative pair of GNM graphs and the most highly assortative pair of SFL graphs. Note the appearance and growth of a peak in the reliability differences in the curves with k > 230–240. Also note that the location of the peak when it appears is where about 280 edges from one of the two GNM graphs are included. Our interpretation is that there is a structure or motif of about 240 common edges in the GNM graphs that combines with a motif of about 280 edges that appear in only one of the two GNM graphs to create a reliable subgraph; similarly, a motif of about 300 common edges in the SFL graphs combines with a complementary motif of about 300 edges in one of the two SFL graphs to create a reliable subgraph.

In principle, we could identify the motifs by considering intersections of the edge sets of reliable subgraphs with the correct number of edges. The *k*-space analysis helps us hunt for the needle in this haystack by removing large parts of the haystack that don't hide the needle. Note also that a sample of these subgraphs was already generated and tested against the reliability criterion in order to evaluate the difference in reliabilities. Hence, the additional work required to identify the motifs is not large. As these are preliminary results, we have not yet carried out this analysis.

6.6 Conclusion

We have argued that the set of coefficients P_k of a reliability polynomial character-ize an arbitrary graph's structure in ways that are immediately relevant to specific dynamical processes. Moreover, we speculate that the coefficients themselves can be well approximated by a small set of parameters that define a sigmoidal function, e.g. its center and width. We have placed the reliability polynomial in a statistical physics context, as the system's partition function. Consequently, we showed how to derive an analytical expression for the location of the percolation phase transi-tion in reaction-diffusion processes over complex networks. This and other examples presented here strongly suggest that a small set of parameters indeed play the most significant role in shaping dynamical phenomena on the network.

We argue that, although identical statements could be made about the reliability in x-space, the k-space perspective—where k represents subgraph size—suggests fruitful new approaches to analyzing graph structure. We explored the mechanisms by which both local and global structural information are separately incorporated into the coefficients. These mechanisms can relate the value of the small set of parameters to local or global statistics about the graph such as degree distribution, clustering coefficients, modularity, etc. In contrast with these statistics, however, the reliability parameters inherently factor in the complex relationships among the multi-

scale structures they represent and the eventual dynamical phenomena. As examples, we presented comparisons between several pairs of graphs that identify the structural differences which are most relevant to specific dynamical processes. We also showed how to study the effects of assortativity-by-degree, number of triangles, and edge-swapping on dynamics. We indicated how reliability coefficients could be used to identify critical structural motifs with hundreds of edges that dramatically influence diffusion processes.

This work is just the tip of an iceberg. We have not attempted a review of the literature, nor have we followed many tantalizing leads to their conclusion. We hope that the brief overview and summary presented here will persuade the reader to delve deeper into the subject.

Appendix 1: Useful Binomial Identities

We note that the derivative of a binomial can be expressed simply:

$$B_k(x) \equiv B(M, k, x) \equiv \binom{m}{k} x^k (1 - x)^k \tag{6.24}$$

$$\frac{d}{dx} B_k(x) = x^{-1}(1 - x)^{-1}(k - Mx)B_k(x) \tag{6.25}$$

Furthermore, we note that, since $\sum_{k=0}^{M} B_k(x) = 1$, all derivatives of the sum must vanish:

$$\sum_{k=0}^{M} \frac{d^q}{dx^q} B_k(x) = \delta_{(q)} \tag{6.26}$$

It is clear from Eq. 6.25 that we can write the qth derivative of $B_k(x)$ in the form

$$\frac{d^q}{dx^q} B_k(x) = x^{-q}(1 - x)^{-q}[k^q - g_{q,k}(x)]B_k(x) \tag{6.27}$$

where $g_{q,k}(x)$ has no terms in k that are higher order than k^{q-1}. Then, by Eq. 6.26

$$S_q(x) \equiv \sum_{k=0}^{M} k^q B_k(x) = \sum_{k=0}^{M} g_{q,k}(x) B_k(x) \quad q > 0 \tag{6.28}$$

We can develop a recurrence equation for $g_{q,k}(x)$:

$$g_{q,k}(x) = k^q - \frac{1}{B_k(x)} x^q (1-x)^q \frac{d^q}{dx^q} B_k(x)$$

$$= k^q - \frac{1}{B_k(x)} x^q (1-x)^q \frac{d}{dx} \left[x^{-(q-1)} (1-x)^{-(q-1)} (k^{q-1} - g_{q-1,k}(x)) B_k(x) \right]$$

$$= (k^{q-1} - g_{q-1,k}(x)) [(q-1)(1-2x) + Mx]$$

$$+ kg_{q-1,k}(x) + x(1-x) \frac{d}{dx} g_{q-1,k}(x). \tag{6.29}$$

In the sum over k, $S_q(x)$, the first term above drops out, leaving

$$S_q(x) = \sum_{k=0}^{M} g_{q,k}(x) B_k(x) = \sum_{k=0}^{M} \{ x(1-x) \frac{d}{dx} g_{q-1,k}(x) + kg_{q-1,k}(x) \} B_k(x) \tag{6.30}$$

Thus we have a recurrence relation for a set of related functions $\tilde{g}_{q,k}(x)$:

$$\tilde{g}_{q,k} = k \tilde{g}_{q-1,k} + x(1-x) \frac{d}{dx} \tilde{g}_{q-1,k}, \quad k > 1 \tag{6.31}$$

with the first functions in the series

$$\tilde{g}_{0,k}(x) = 0; \quad \tilde{g}_{1,k}(x) = Mx; \quad \tilde{g}_{2,k}(x) = Mx(k+1-x) \tag{6.32}$$

and the summations

$$S_0(x) = 1; \quad S_1(x) = Mx; \quad S_2(x) = M(M-1)x^2 + Mx;$$
$$S_3(x) = M(M-1)(M-2)x^3 + M(M-1)x^2 + Mx \tag{6.33}$$

or, in general,

$$\sum_{k=0}^{M} k^q B_k(x) = \sum_{l=0}^{q-1} \frac{M!}{(M-l)!} x^{q-l}. \tag{6.34}$$

Appendix 2: Sigmoidal Polynomials

We make the ansatz that $P_k = f(k/M)$, where f is a sigmoidal polynomial which satisfies the following constraints:

1. $f'(0) = 0$;
2. $f'(1) = 0$;
3. $f(0) = 0$;
4. $f(1) = 1$;
5. $x_0 = \arg\max f'(x)$ on $[0, 1]$; and
6. $f(x)$ is monotonic non-decreasing on $[0, 1]$.

Fig. 6.8 Examples of a two-parameter family of sigmoidal polynomials. *Top* The "bump" functions that generate the sigmoids. k_c/M is the location of the peak and N controls the shape of the peak. *Bottom* The sigmoidal polynomials for $n = 0, 1$ or 2 and $x_0 = 0.41$ or 0.59
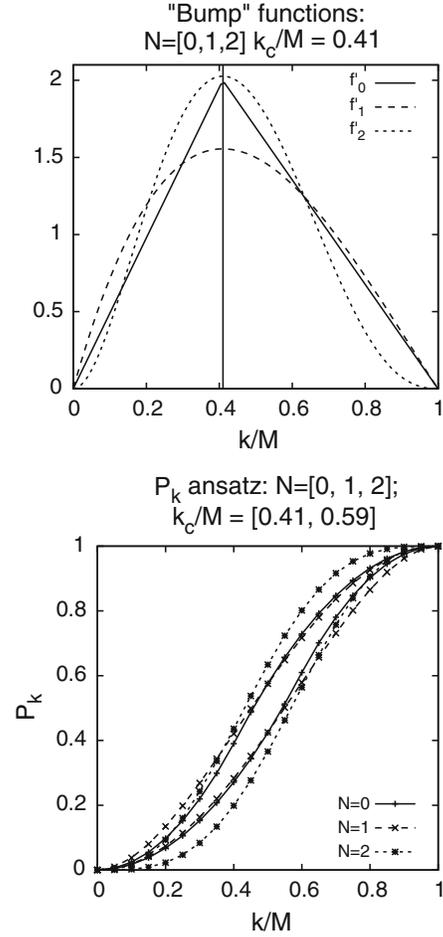
There are many such polynomials, which can be thought of as the cumulative distribution function for unimodal probability distributions. Here we consider one very simple two-parameter family.

We start with a "bump" function $g_n(x) \equiv [x(1-x)]^n(x - c_n)$. Note that $g_n(0) = g_n(1) = 0$. Construct $\tilde{f}(x) = \int_0^x g_n(y)dy$. Then define $f_n(x) \equiv \tilde{f}(x)/\tilde{f}(1)$. Clearly, $f_n(x)$ satisfies constraints 1–4 above. Constraint 5 is also satisfied if we choose a value for c_n to make $g'(x_0) = 0$:

$$g'(x_0) = [x(1-x)]^{n-1}[-n(1-2x)c_n + (n+1-(2n+1)x)x] \qquad (6.35)$$

So x_0 is a zero of $f_n''(x)$ if

$$c_n = \frac{x_0}{n} \frac{n+1-(2n+1)x_0}{(1-2x_0)} \qquad (6.36)$$

There are no free parameters left to adjust in f_n. To ensure that constraint 6 is satisfied, we require that $c_n < 0$ or $c_n > 1$, i.e. $(n+1)/(2n+1) > x_0 > n/(2n+1)$. This is not a general constraint on x_0, just a constraint for the particular family of sigmoidal polynomials we are using. For this family, as n increases, $f_n(x)$ is smoother at $x = 0$ and 1, and rises more rapidly around x_0, but x_0 is confined to a smaller and smaller interval around $\frac{1}{2}$.

Explicit forms for f_1 and f_2 are given below and, along with f_0, plotted in Fig. 6.8:

$$f_1(x) = x^2 \frac{3x^2(2x_0 - 1) - 4x\left(3x_0^2 - 1\right) + 6x_0(3x_0 - 2)}{6x_0^2 - 6x_0 + 1}, \quad \text{for } \frac{1}{3} < x_0 < \frac{2}{3} \tag{6.37}$$

$$\begin{aligned} f_2(x) = x^3(5x_0^2 - 5x_0 + 1)^{-1} \\ \left\{ -10x^3(2x_0 - 1) + 6x^2\left(5x_0^2 + 5x_0 - 4\right) + \right. \\ \left. 15x\left(-5x_0^2 + x_0 + 1\right) + 10x_0(5x_0 - 3) \right\}, \quad \text{for } \frac{3}{5} > x_0 > \frac{2}{5}. \end{aligned} \tag{6.38}$$

References

1. Abdelhamid SE, Alo R, Arifuzzaman SM, Beckman P, Bhuiyan MH, Bisset KR, Fox EA, Fox GC, Hall KD, Hasan S, Joshi A, Khan M, Kuhlman CJ, Lee S, Leidig JP, Makkapati H, Marathe MV, Mortveit HS, Qiu J, Ravi S, Shams Z, Sirisaengtaksin O, Subbiah R, Swarup S, Trebon N, Vullikanti A, Zhao Z (2012) Cinet: A cyberinfrastructure for network science. In: Proceedings of the 8th IEEE International Conference on eScience, 2012, Chicago, IL, October 8–12, 2012.
2. Badham J, Stocker R (2010) The impact of network clustering and assortativity on epidemic behaviour. Journal of Theoretical Population Biology 77:71–75.
3. Barrett CL, Lewis BL, Chen J, Kumar V, Eubank SG, Marathe MV, Mortveit HS (2009) Interactions among human behavior, social networks, and societal infrastructures: A case study in computational epidemiology, Springer Verlag, pp 479–507.
4. Colbourn CJ (1987) The Combinatorics of Network Reliability. Oxford University Press.
5. Eubank SG, Kumar VSA, Khan M, Marathe MV, Barrett CL (2010) Beyond Degree Distributions: Local to Global Structure of Social Contact Graphs, Springer, Berlin/Heidelberg 2010, p 1.
6. Ford L, Fulkerson D (1956) Maximal flow through a network. Canadian Journal of Mathematics 8:399–404.
7. Ford L, Fulkerson D (1962) Flows in Networks. Princeton University Press.

8. Grassberger P (1983), On the critical behavior of the general epidemic process and dynamical percolation. Mathematical Biosciences pp 157–172.
9. Newman M (2002) Assortative mixing in networks. Physical Review Letters 89:1–4.
10. Page L, Perry J (1994), Reliability polynomials and link importance in networks. IEEE Transactions on Reliability pp 51–58.
11. Reif F (1965) Fundamentals of Statistical and Thermal Physics. McGraw-Hill.

Chapter 7
Spatial Effects: Transport on Interdependent Networks

Richard G. Morris and Marc Barthelemy

Abstract Space plays an important role in the behaviour of both individual infrastructures, and the interdependencies between them. In this Chapter, we first review spatial effects, their relevance in the study of networks, and their characterization. The impact of spatial embedding in interdependent networks is then described in detail via the important example of efficient transport (or routing) with multiple sources and sinks. In this case, there is an optimal interdependence which relies on a subtle interplay between spatial structure and patterns of traffic flow. Although simplified, this type of model highlights emergent behaviour and brings new understanding to the study of coupled spatial infrastructures.

7.1 The Importance of Spatial Effects

Catastrophic failures in real world infrastructures are typically a result of consecutive improbable events. However, the chain of these events can often traverse more than one type of system, therefore it is important to understand the role of interdependency. For example, Fig. 7.1 shows a simplified schematic of interdependencies between various systems, and demonstrates how easily failures can propagate from one system to another.

Such interdependencies can be loosely classified into different categories [1]. For example, interdependencies can be 'physical', where the state of each system relies on the physical output of the other. In this case, one can imagine a coal-fired power station might generate the power for a rail network that, in turn, is used to deliver

R. G. Morris
The University of Warwick, Coventry, UK
e-mail: r.g.morris@warwick.ac.uk

M. Barthelemy (✉)
CEA-IPhT, Paris, France
e-mail: marc.barthelemy@cea.fr
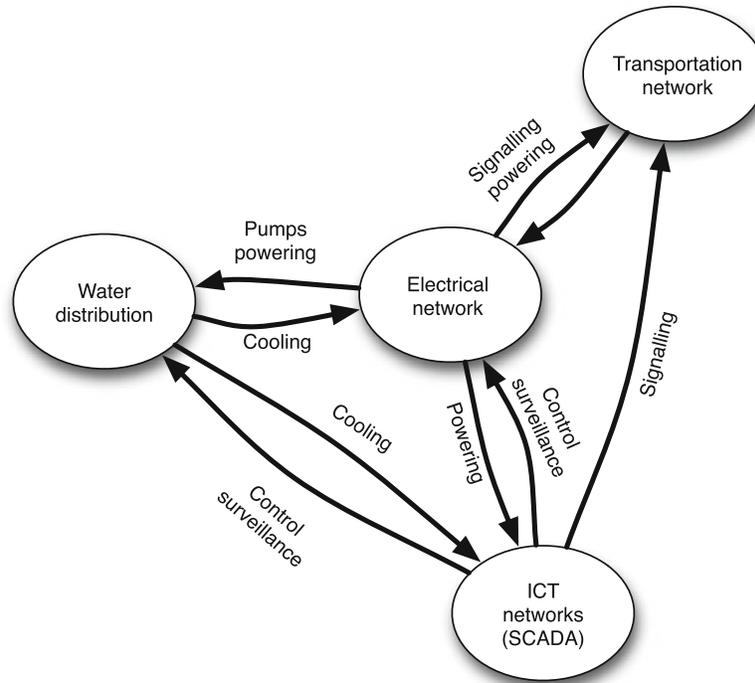
Fig. 7.1 Types of interdependencies between different critical infrastructures—adapted from Ref. [1]. Failure in the water distribution network might cause disruption at a power station, due to lack of cooling. This, in turn, could affect control networks employed to monitor water distribution in the first place, exacerbating the initial failure

coal to that same power station. We may also characterize 'cyber' dependencies, which is the case for example in supervisory control networks. Here, the state of one system relies on information about the state of the other system. The final type of interdependency can be described as 'spatial'. That is, different systems can be affected by a localized disturbance due to spatial proximity. This simplest case relates to scenarios such as seismic failures, explosions, or fires, where an external event directly affects different infrastructures in the same location—and could be the trigger for a chain of failures. However, as we describe in later Sections, localized failure in one infrastructure may also be the cause of localized failure in another. For example, local traffic congestion on a road network can cause train overcrowding in the same region due to more people choosing the train.

This Chapter focusses on the last of these three classes, where transportation systems such as the road network, rail, subway, etc., are an important example. Such urban systems are, by construction, embedded in space and interdependent. However, an assessment of their resilience is very difficult [2], and therefore understanding the effect of interdependence on the stability of such systems is an important task [3]. One of the main problems is that human-mediated interdependency can be the source of counter-intuitive phenomena, such as flash congestion [4]. As a result, many detailed systems-engineering approaches have floundered, whilst the simplified models used

by physicists have proved helpful to identify the mechanisms underlying certain important characteristics. The aim of this Chapter is therefore to describe interesting effects that arise—at the aggregate level—for transportation systems that are both spatial and interdependent. To achieve this, the Chapter is organized as follows. In the next Section, spatial networks will be defined and their key properties reviewed. In Sect. 7.3, we describe, via two examples, how these properties affect systems of coupled networks. The final Section then concludes with a short recap and discussion of the main characteristics of such systems, whilst also highlighting open questions in the field.

7.2 Spatial Networks

Generally speaking, spatial networks are networks for which the nodes are located in a metric space—that is, one that permits the notion of a distance between any two points. Transportation and mobility networks, Internet, mobile phone networks, power grids, social and contact networks, neural networks, are all examples where space is relevant and where topology alone does not contain all the information [5]. For most practical applications though, it suffices to embed nodes in a straightforward two-dimensional euclidean space.

To give an idea of the role played by spatial effects, consider the following simple example. Imagine that a set of nodes are placed at random in the plane, and an edge is created between any pair of nodes according to some probabilistic rule. For spatial networks, this probability might decrease with the euclidean distance between the two nodes, for example. In this case, there is an implicit 'cost' associated with size of each edge, and therefore the connections between nodes are predominantly local. More broadly, the spatial constraints have had a dramatic effect on the resulting topological structure of the network.

Notice that the above definition does not imply that a spatial network is planar. Indeed, the airline passenger network, for example, is a network connecting direct flights through the airports in the world, and is not a planar network. Further to this, it is not even necessary that the embedded space of the network corresponds with a real space: social networks for example connect individuals through a friendship relations. The probability that two individuals are friends however generally decreases with the euclidean distance between them, showing that in social networks, there is an important spatial component (see for example [6]).

Whilst the above exceptions can be both important and interesting, in most systems of interest, both planarity and a real space embedding are natural choices. For example, electricity and gas distribution, roads, rail, and other transportation networks are all, to a very good approximation, spatial and planar networks. Due to the number of relevant examples and the intuitive ease with which they can be understood, we choose to focus primarily on such spatial-planar systems.

In the rest of this Section, therefore, we first review the main types of spatial networks and how they can be characterized. Then, with this in place, we describe

two important classes of problems that commonly feature spatial networks—failure cascade prevention and routing/transportation—and how they can be analyzed.

7.2.1 Types of Spatial Networks

There are, of course, many different types of spatial network. Indeed, as we describe in Sect. 7.2.2.1, choosing appropriate measures to classify different types of spatial networks is still an open area of research. However, for the purposes of this Chapter, it will suffice to look at only the broadest classes of spatial networks.

7.2.1.1 Regular Lattices

The simplest and most commonly used spatial network is the regular lattice— constructed by repeatedly copying a so-called 'unit cell'. Whether a simple square lattice or one comprising more complicated polyhedra, the general properties are: uniform density, periodic structure and high degree of symmetry. In almost all cases, the unit cell is planar and very straightforward, where all nodes of the network have the same degree (although it is possible to use repeating units that are either non-planar or do not have uniform degree). Regular lattices are prevalent for two reasons, primarily due to their simplicity, but also due to the fact that many man-made systems have very regular structures such as the road network in many cities (e.g., Manhattan).

7.2.1.2 Delaunay Triangulations (and Voronoi Tessellations)

If the underlying system of study is planar, but a regular lattice has too much symmetry, one option is to use less regular types planar subdivisions. By far the most well-know of these are the Delaunay triangulation, and its dual, the Voronoi tessellation.

A Delaunay triangulation can be defined for any set of points positioned in the plane, the result being an almost[1] unique triangulation that maximises the smallest angle of all the triangles. That is, it tends to avoid very thin triangles. Given a particular Delaunay triangulation, one may construct the Voronoi diagram—a more general subdivision of the plane that associates a polygon with each node (see Fig. 7.2). There is a great deal of work that concerns the properties of such subdivisions and how to efficiently generate them, primarily due to their importance in problems of finding a so-called 'convex hull' for a discrete set of points. We refer the interested reader to the important work [7] which forms the cornerstone of most modern techniques for generating either Delaunay or Vornoi diagrams.

---

[1] If, in the exceptional circumstances that more than three nodes lie on the same circumcircle (see Fig. 7.2), then the neither the Delaunay or Vornoi diagrams are unique.
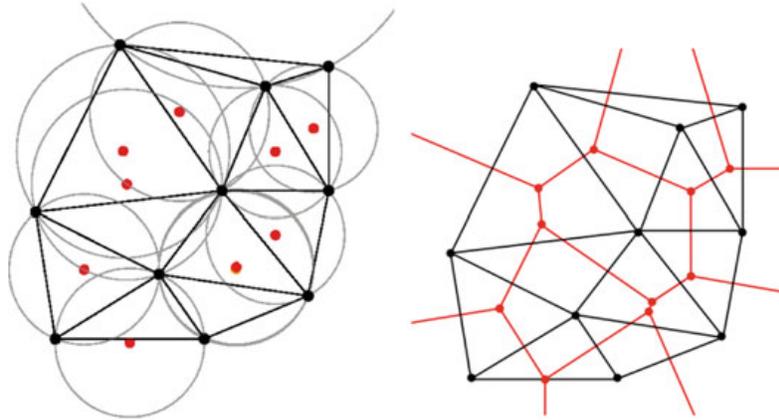
Fig. 7.2 A Delaunay triangulation is a triangulation of points (*black dots*) in the plane where the circumcircle of each triangle does not contain any other points. If the centres of circumcircles (*red dots*) belonging to neighbouring triangles are joined together, the Voronoi diagram is produced (*red lines*)

The benefit of such irregular planar subdivisions to the modeller is that one may specify a non-homogeneous distribution of nodes. For example, when representing a water distribution network, one might expect the result to be planar subdivision, but with greater density around towns and cities.

7.2.1.3 Probabilistic Networks

In order to incorporate more disorder, it is necessary to adopt a fully probabilistic approach to network generation. Here, one typically starts with a set of nodes positioned in the plane, and then, for each pair of nodes, creates an edge according to some probabilistic rule. The example discussed earlier considered a probability decreasing with the distance between two nodes, but this might equally involve more complicated spatio-topological indicators, such as clustering coefficients or average shortest paths (for more details on probabilistic models of spatial networks, we refer the interested reader to the review [5]).

7.2.2 Characterization of Spatial Networks

Whilst the first attempts to characterize spatial networks were made in the area of quantitative geography in the 1960s [8–11], more recently other measures have been popularized by the study of complex networks [12, 13]. Most of the currently used measures—often called indicators—are relatively simple, but still give important information about the spatial structure of the network. Here, we will briefly review the most useful quantities which allow for a good characterization of spatial networks.

7.2.2.1 Spatial and Topological Indicators

The most common quantities used to describe networks typically characterize only topological aspects and are not of particular interest for spatial networks. For example, the degree distribution is usually peaked which results from the fact that physical constraints imply a small cut-off, and the clustering and assortativity are usually flat, a consequence of the fact that connections are predominantly made to neighbors irrespective of their degrees.

A first useful quantity for spatial networks is the average shortest path ℓ which for most random networks scales as $\ell \sim \log N$ (where N is the number of nodes), signalling a small-world type behavior. In contrast, spatial networks are large-world and usually display a lattice-like behavior of the form

$$\ell \sim N^{1/d}, \tag{7.1}$$

where d is the dimension of the embedded space which, for most applications, is two.

Another helpful quantity used to characterize spatial systems, is the total length

$$L_{\text{tot}} = \sum_{e \in E} \ell(e), \tag{7.2}$$

where E is the set of edges and $\ell(e)$ is the Euclidean length of the edge e. Under the assumption of a peaked distribution of $\ell(e)$, the scaling for most networks is of the form

$$L_{\text{tot}} \sim L\sqrt{N}, \tag{7.3}$$

where L is the typical size of the area under consideration. One can then easily construct the minimum spanning tree on the same set of nodes and obtain its length $L_{\text{tot}}(MST)$. The ratio of these lengths

$$C = \frac{L_{\text{tot}}}{L_{\text{tot}}(MST)}, \tag{7.4}$$

is always larger than 1 and is a good measure of how costly a network is.

For some irregular planar subdivisions, like roads and railway lines, the polygons that make up the faces can correspond to important information about the structure of the network. We can characterize the faces by two main quantities, their area A and their shape factor

$$\phi = \frac{4A}{\pi D^2}, \tag{7.5}$$

where D is the largest diameter of the polygon. This quantity ϕ thus indicates how anisotropic the face is: for $\phi \approx 0$ the face is a very elongated rectangle and for $\phi \approx 1$

the face is essentially a disk. We observe for most road networks [5] the following behavior

$$P(A) \sim A^{-\gamma}, \tag{7.6}$$

where $\gamma \approx 2$, a value which probably finds its origin in the node density fluctuations [5].

The shape factor distribution usually displays a peak around $\phi \approx 0.6$ and its time evolution displays some interesting behavior which is so far unexplained [14].

We end this section by noting that the classic quantity, the betweeness centrality (BC) which quantifies the importance of a node (or an edge) in the network behaves very differently in spatial networks. In complex, scale-free networks, the BC scales as a power of the degree. In other words, the larger the degree the larger the BC, indicating that the hubs are the most important nodes in the network. In spatial networks, there is an interplay between the degree and the distance to the barycenter of nodes, leading to the appearance of 'anomalies', nodes with a small degree and a large BC.

7.3 The Effects of Interdependence in Spatial Networks

Studies that incorporate the features of coupled networks with those of spatial networks are small but growing in number. So far, such research has focussed on either failure cascades or transport and routing processes.

For failure cascades, the idea is that either the nodes or edges in the underlying system have an intrinsic carrying capacity which, if exceeded, causes a 'failure'. Once an edge or node has failed, it is removed from the network and then redistributed. This may then cause the overloading and failure of further power lines, and so on. Such cascading failures are important because, under certain circumstances, small isolated failures can result in large system-wide outages. Recent work [15–21] has extended this idea to a system of interdependent networks—that is, the failure of a node in one network causes the immediate failure of the nodes to which it is connected in the second network. By measuring the size of the largest connected component that remains following a cascade, it can be shown that the extent of cascades increases as the number of inter-network connections is increased. In Refs. [23, 24], the authors apply this model of interdependent cascades to a system comprising two interdependent square lattice networks. Here, it is argued that since the model uses percolation techniques, the results—measured in terms of giant connected components—should not depend on the particular realization of the network. That is, it is known that the percolation transition has universal scaling behavior which does not depend on the coordination number and is the same for lattice and off-lattice models, as long as the links have a finite characteristic length [25]. The key aspect of the model is that dependency links between two networks are randomly chosen within a certain distance r. One can then show that percolation for small r is a second-order transition,

and for larger r is a first-order transition. Moreover, the results suggest that systems of this type become most vulnerable when the distance between interdependent nodes is in the intermediate range: greater than zero but much smaller than the size of the system.

As mentioned above, another class of systems that are both interdependent and spatial are general transport processes, or *flows* [22]. Whether flows of people, fluids, or electrical currents, these systems can be characterized by specifying the topology of the underlying network, a source-sink distribution, and a dynamic. To avoid confusion, we only imagine dynamical processes that converge to a steady state—resulting in a stationary distribution of flows over the network. Unfortunately, the methods of analysis mentioned above do not capture many of the typical features one might expect here. For example, it is easy to imagine a simple source-sink distribution that allows the network to be split into two distinct components such that the flows are unaffected. In this case, the size of the giant component may decrease but the network is still operating well.

Since the percolation techniques used to analyze cascading failures are well documented, the rest of this Chapter is devoted to the description of transport and routing processes.

7.3.1 Transport and Routing

One may ask: how should an interacting, or coupled, set of *flow* networks be characterized, and what are the interesting features of such systems? From observing real systems, one expects interesting effects to arise from three main areas:

- Spatial - and localization-effects from network connections.
- Spatial distribution of source and sinks.
- Coupling between the two networks.

Of course, the global behaviour of any real system is intimately linked with the particular form of dynamical interactions involved. However, some understanding of the above points can be gained by investigating the properties of simple examples that are chosen well enough to represent certain classes of systems. In this Section, we recap the results of examining such a 'toy model' [26], where the main idea is intuitively simple. Consider a transport network where there is a choice between travelling by train or by car, or perhaps the routing of packets in Information Communication Technology networks (ICT) where there are two different networks available (a simple schematic of this type of system is shown in Fig. 7.3). For these types of systems, the typical choice is between a 'fast but sparse' network and 'slow but dense' network. It is in this system that interesting effects arise through the interplay of the three main areas outlined above.
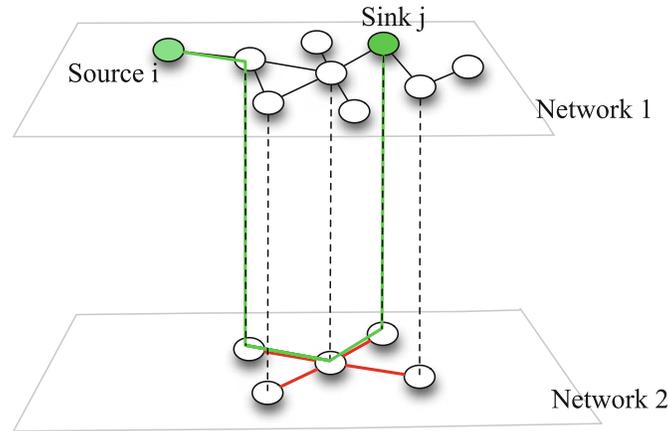
Fig. 7.3 Flows on two interdependent networks: edges of network 1 are shown in black, edges of network 2 are shown in red, and nodes in common to both networks are considered to be coupled (shown by *dashed lines*). Shown in *green*, we represent a path between two nodes, the 'source' i and the 'sink' j
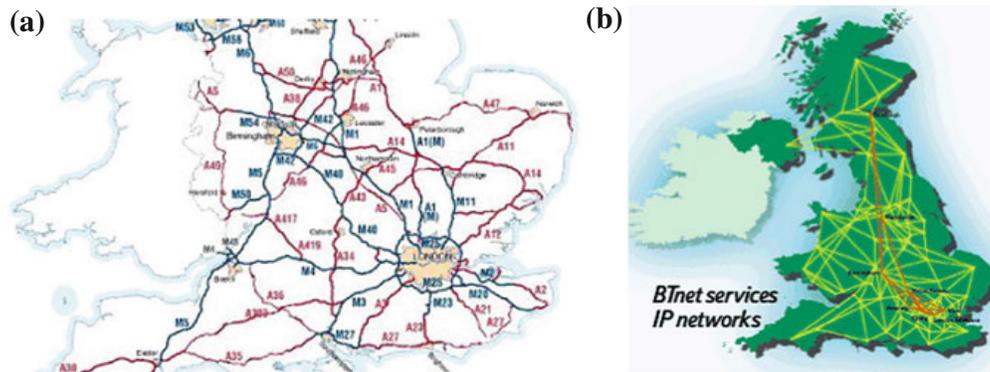

Fig. 7.4 **a** The national road network in England, and **b** the network of major internet servers across the UK operated by British Telecom. These networks are consistent with planar subdivisions on a finite sampling of nodes taken with uniform density

7.3.1.1 Network Structure

Since the motivation here is transport and routing problems, inspiration for the model can be found by looking at real systems. For example, one can argue that schematics of national transport networks or internet server networks resemble planar subdivisions where the nodes have been arranged at random with uniform density (see Fig. 7.4). Drawing from ideas discussed in Sect. 7.2.1, a good approximation for these systems is therefore to use Delaunay triangulations.

   With the aforementioned examples in mind, one can ask: how should two Delaunay triangulations be coupled together? We imagine a road network coupled to a rail or subway network. Here, all the nodes of the road network are not nodes of the rail network, but conversely, all stations are located at points which can be considered

(a)   **(b)**   **(c)**
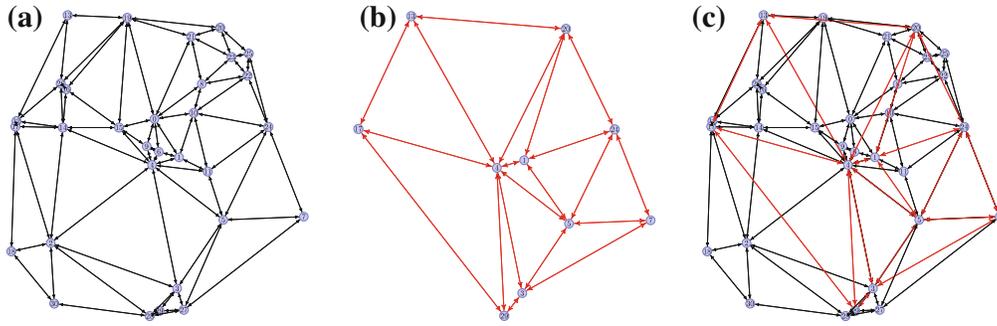
Fig. 7.5 Each instance of the system is generated according to the following process: **a** First, $N^{(1)}$ nodes (here $N^{(1)} = 30$) are positioned at random within the unit disk and the Delaunay triangulation is produced; **b** the second network is then generated by drawing $N^{(2)}$ (here $N^{(2)} = 10$) nodes uniformly from the existing ones ($N^{(2)} \leq N^{(1)}$) and, once again, computing the Delaunay triangulation; **c** the combined system is no longer planar

as nodes in the road network. That is, the nodes of one network are a subset of the nodes of the other. As will be shown, this setup conveniently provides a simple way to realize the 'sparse' versus 'dense' characterization described above.

More mathematically, one can construct two Delaunay triangulations $DT^{(1)}$ and $DT^{(2)}$. The set of nodes of $DT^{(1)}$ are taken to be $N^{(1)}$ points distributed uniformly at random within the unit disk. The nodes of $DT^{(2)}$ are then selected at random from $N^{(1)}$ and we define

$$\beta = \frac{N^{(2)}}{N^{(1)}} \leq 1. \tag{7.7}$$

That is, the model comprises two individual networks that are each planar Delaunay triangulations, forming a combined network that is not necessarily planar (see Fig. 7.5). For the combined network, $N = N^{(1)}$, and $E = E^{(1)} \cup E^{(2)}$. Recalling that Delaunay triangulations are effectively unique for a given set of points, it is then clear that, for a given value of β, the spatial and topological structure is entirely defined by $N^{(1)}$ and $N^{(2)}$.

7.3.1.2 Route Assignment

For modelling a transportation system, it is natural to associate a velocity $v^{(n)}$ with each network $n \in \{1, 2\}$, and to assign weights to each undirected edge $e^{(n)} = (\mathbf{x}_i, \mathbf{x}_j)$ according to

$$w(e^{(n)}) = \frac{|\mathbf{x}_i - \mathbf{x}_j|}{v^{(n)}}. \tag{7.8}$$

Here, w is the time taken to traverse the edge, and will provide the building block for all other useful system indicators.

To allocate flows on the network, rather than considering a dynamical system which acts to minimize a global quantity—such as electrical networks, where the dissipated power is minimized—a straightforward choice is to once again follow a transportation analogy. This means that the source-sink distribution of a general system of flows, can be replaced by an origin-destination (OD) matrix T_{ij}. Indeed, as before, this approach is also representative of the Internet, i.e., it is necessary to not only receive a packet of information, but it must be a particular packet sent from a particular server. Since each entry in the OD matrix specifies the proportion of the total flow that goes from node i to node j, this type of representation has the benefit that flows are completely specified by combining an OD matrix and a method of route choice.

The most obvious candidate for a method route choice, is to take the journey that minimizes the travel time. That is, the weighted shortest path, where the weights are given by Eq. (7.8). Here, the idea is that the ratio

$$\alpha = \frac{v^{(1)}}{v^{(2)}} \tag{7.9}$$

is a single parameter that controls the relative speed of travel on the two networks. Indeed, in order to simplify further, we impose the constraint that $\alpha \leq 1$. Since $\beta < 1$, this has the effect of enforcing the 'fast but sparse' versus 'slow but dense' scenario.

In terms of the OD matrix, it is impractical to consider the interplay between all possible forms for T_{ij}. Therefore it helps to choose a method that interpolates between two extremes, the monocentric case and a form of Erdős-Réyni random graph. We start with a monocentric OD matrix—i.e., all nodes travel to the origin—and then add noise by rewiring in the following way. For each node, with probability p, choose a random destination, and with probability $1 - p$, choose the origin (see Fig. 7.6).

7.3.1.3 Interdependence

Previous studies of interacting networks use the term *coupling* to describe how well two networks are linked. Typically, this is a purely topological definition i.e., the fraction of nodes from one network which link to another [15], or the probability that a particular node has an edge which connects both networks [20]. For transport processes, a better measure of interaction must include details of how the flows are distributed. For the system outlined above, we then specify a new quantity which we coin *interdependence* and is defined in a similar vein to the betweenness centrality

$$\lambda \equiv \sum_{i \neq j} T_{ij} \frac{\sigma_{ij}^{\text{coupled}}}{\sigma_{ij}}, \tag{7.10}$$
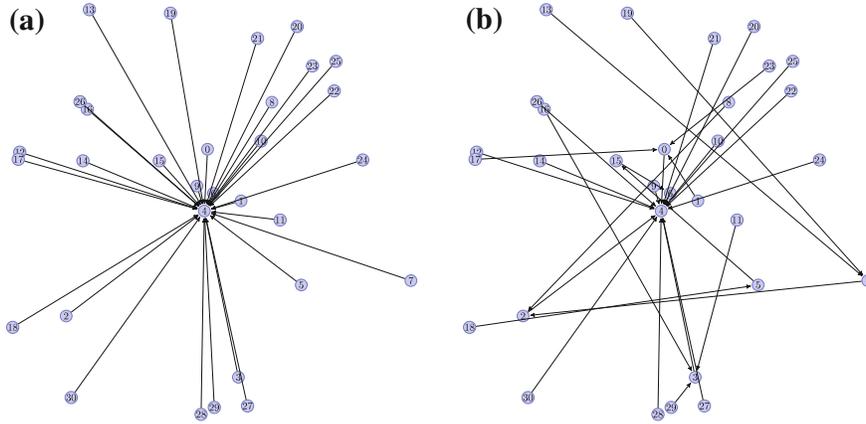
Fig. 7.6 Representations of OD matrices where each arrow corresponds to an entry in T_{ij} and which relates to the underlying geometry of Fig. 7.5. **a** A monocentric OD matrix. **b** A monocentric OD matrix randomly rewired with probability $p = 0.5$

where $\sigma_{ij}^{\text{coupled}}$ is the number of weighted shortest paths between nodes i and j, which include edges from *both* networks. In the transportation analogy, the interdependence is a way to quantify the importance of different transportation modes in order to achieve a fast journey. Here, the entries of the origin-destination matrix T_{ij} are normalized i.e., $\sum_{ij} T_{ij} = 1$, and it is clear from Eq. (7.10) that $\lambda \in [0, 1]$ is dependent on the method by which the flows are allocated and not just the system topology. The larger λ, the more one network is relying on the other to ensure efficient shortest paths (note that there is usually a maximum value of λ strictly less than one, since not all shortest paths can be multimodal). It is also clear that, by virtue of influencing the shortest paths, the number α can control the interdependence between the two networks.

With Eq. (7.10) in mind, instead of investigating the likelihood of catastrophic cascade failures, we consider more general measures of how well the system is operating. For example, one such measure is the average travel time

$$\tau = \sum_{i \neq j} T_{ij} w_{ij}, \qquad (7.11)$$

where w_{ij} is defined from Eq. (7.8) as follows: $w_p = \sum_{e \in p} w(e)$ is the cumulative weight of path p, and $w_{ij} = \min_{p \in P} w_p$ is the minimum weight of all paths P between nodes i and j. For most practical transport processes, a well designed system reduces the average time travelled (i.e., water/food supply, the Internet, transportation, etc.).

Another important quantity, which can be used as a simple proxy for traffic flow, is the edge betweenness centrality (BC). For the system at hand, the definition of the edge BC is

$$b(e) = \sum_{i \neq j} T_{ij} \frac{\sigma_{ij}(e)}{\sigma_{ij}}, \qquad (7.12)$$
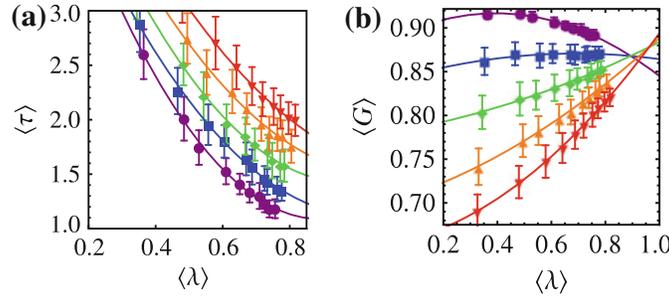
Fig. 7.7 Simulation results for the average shortest path and the Gini coefficient ($N^{(1)} = 100$, $N^{(2)} = 20$, and p values: 0 (*purple*), 0.2 (*blue*), 0.4 (*green*), 0.6 (*orange*), and 0.8 (*red*)). When the interdependence increases, the average shortest path decreases and the Gini coefficient can increase for large enough disorder (*lines are polynomial fits*)

where the sum is weighted by the proportion of trips T_{ij}, and $\sigma_{ij}(e)$ is the number of weighted shortest paths between nodes i and j, which use edge e. The betweenness centrality allows the introduction of a second measure, the Gini coefficient G. A number between zero and one, G is typically used in economics for the purpose of describing the concentration of wealth within a nation. Here it is used to characterize the disparity in the assignment of flows to the edges of a network, something that has been done before for transportation systems such as the air traffic network [27]. For example, if all flows were concentrated onto one edge, G would be one, whilst if the flows were spread evenly across all edges, G would be zero. We use the definition according to Ref. [28]

$$G \equiv \frac{1}{2|E|^2 \bar{b}} \sum_{p,q \in E} |b(p) - b(q)|, \tag{7.13}$$

where subscripts p and q label edges, E is the total number of edges, and $\bar{b} = \sum_{p \in E} b(p)/|E|$ is the average 'flow' on the system. In this picture, the Gini coefficient can now be thought of as a measure of road use. A low value indicates that the system uses all roads to a similar extent, whilst a high value indicates that only a handful of roads carry all the traffic.

7.3.1.4 Existence of Optimal Interdependence

The set of numbers p, β and α, now define an ensemble of systems that are statistically equivalent (with respect to λ, τ, and G). Therefore one may calculate the quantities $\langle \lambda \rangle$, $\langle \tau \rangle$, and $\langle G \rangle$ for different values of p and α, where angle brackets $\langle \ldots \rangle$ represent an ensemble average.

Simulation results are shown in Fig. 7.7, where each data point corresponds to an average over fifty instances of the OD matrix for each of fifty instances of the coupled network geometry. As the interdependence λ increases, the average journey
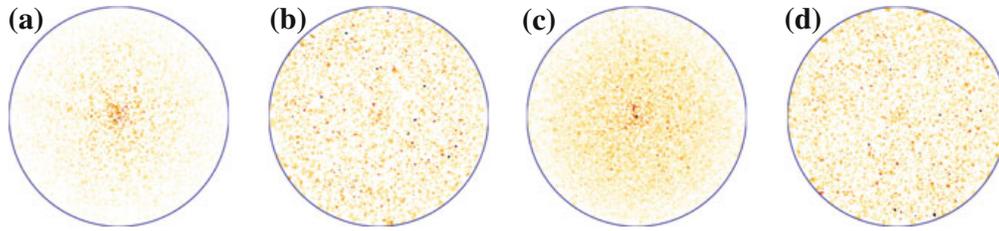
Fig. 7.8 Colormaps showing normalized edge flows—plotted at the midpoint of each edge—over many instances of the system. Colors are assigned starting from white (*for zero flow*) and moving through *yellow*, *orange* and *red* for higher values of flow, until reaching black (*maximum flow*). Each Subfigure corresponds to the following parameter values: **a** $p = 0.2$, $\alpha = 0.9$; **b** $p = 0.2$, $\alpha = 0.1$; **c** $p = 0.8$, $\alpha = 0.9$; **d** $p = 0.8$, $\alpha = 0.1$

time decreases (Fig. 7.7a). This is straightforward to understand since the increased interdependence is simply a result of reducing the velocity ratio α. Furthermore it is clear that increasing randomness in the origin destination matrix increases the length of the average shortest path by an almost constant value, irrespective of the interdependence. By contrast, the behaviour of the Gini coefficient at different interdependencies (Fig. 7.7b) is less easily explained. Consider instead Fig. 7.8. Here, each colormap shows the distribution of flows resulting from many instances of the system.

The first two plots, Figs. 7.8a and 7.8b, were generated from OD matrices rewired with low probability ($p = 0.2$) i.e., almost monocentric. The ratios of edge weights per unit distance between the two networks are $\alpha = 0.9$ and $\alpha = 0.1$ respectively. Therefore each diagram corresponds to a point on the blue line in Fig. 7.7b. For $\alpha = 0.9$, there is minimal independence between the networks and a high concentration of flows are seen around the origin. Since the flows are disproportionately clustered, this configuration is described by a high Gini coefficient. By contrast, for $\alpha = 0.1$, the difference in the edge weights means that it can be beneficial to first move away from the origin in order to switch to the 'fast' (low α) network. We therefore see a broader distribution of flows with small areas of high concentration around coupled nodes. The emergence of these *hotspots* away from the center also corresponds to a high Gini coefficient—and therefore the blue line in Fig. 7.7b is relatively flat. Figs. 7.8c and 7.8d correspond to the red line of Fig. 7.7b: generated from OD matrices rewired with high probability ($p = 0.8$). We observe that even for α close to one, the distribution of flows is broader than for $p = 0.2$—resulting in a lower Gini coefficient. As α is decreased, the second network becomes more favourable and interdependence *hotspots* can be seen once again—resulting in a high Gini coefficient and a positive gradient for the red line of Fig. 7.7b. This result points to the general idea that randomness in the source-sink distribution leads to local congestion and more generally to a higher sensitivity to interdependence.

At this stage, it is natural to combine the effects observed above into a single measure. We assert that it is likely a designer or administrator of a real system would
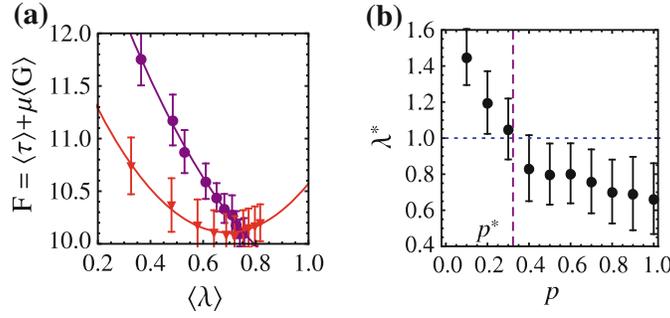
Fig. 7.9 Existence of an optimal interdependence: **a** Simulation results for $\mu = 10$, $N^{(1)} = 100$, $N^{(2)} = 20$, and p values: 0 (*purple*), 0.4 (*green*), and 0.8 (*red*) (three values only of p are shown to ensure the lines of best-fit can be seen clearly). **b** Minima of quadratic best-fit curves for different values of p. We obtain $\lambda = \lambda^*$ for $p^* \simeq 0.34$ (The error bars shown are those of the closest data point to the minimum of the best-fit curve)

wish to simultaneously reduce the average travel time and minimize the disparity in road utilization. To serve this purpose, a 'utility' function

$$F = \langle \tau \rangle + \mu \langle G \rangle \tag{7.14}$$

can be defined, where it is immediately apparent from Fig. 7.7 that, for certain values of μ, the function F will have a minimum. That is, a non-trivial (i.e., non-maximal) optimum λ will emerge. Figure 7.9a shows that, whether a non-trivial optimum interdependence exists depends on the origin-destination matrix. For OD matrices rewired with a high probability, increasing the speed of the rail network reduces the road utilization as flows become concentrated around nodes where it is possible to change modes. Dependent on the value of μ, the effect of reduced utilization can outweigh the increased journey time, leading to a minimum in F. Monocentric OD matrices, by contrast, have inherently inefficient road utilization when applied to planar triangulations, regardless of the speed of the rail network. Therefore no minimum is observed, and hence no (non-trivial) optimum λ. More systematically, one may plot the minima λ^* of best-fit curves corresponding to different values of p (Fig. 7.9b). Defining p^*, the value of p for which $\lambda^* = 1$, it is then possible to categorize the system into one of two regimes. We observe that: if $p < p^*$, then the optimal interdependence is trivially the maximum; otherwise if $p \geq p^*$, a non-trivial optimal interdependence exists.

7.4 Discussion and Perspectives

In this Chapter, we have highlighted the importance of spatial effects in coupled networks by focussing on problems of transportation and routing. In contrast to studies of failure mitigation—that often use either percolation or cascading-sandpile

techniques—models of transport are better described by measures of utility and efficiency. By using such quantities, it is possible to identify an optimal *interdependence* between the two networks. Below the optimum, the system is inefficient and travel times can be large, whilst above the optimum, system utilization is poor due to congestion arising around 'link nodes' that connect the two networks. The existence and behavior of this optimal value turns out to be very sensitive to the randomness of the individual trajectories that make up the system.

Even though the model is very simplified, it possesses the advantage of highlighting dominant mechanisms, and can serve as a basis for more sophisticated modeling such as the ones used and developed by civil engineers. The broader interpretation being that systems that rely on routing like transportation networks, or the Internet, may be inherently fragile to certain changes in supply and demand. Furthermore, if such observations can be generalized, this could have serious ramifications in other areas, such as the transition from centralized to de-centralized power generation [29].

Finally, we note that most studies have so far considered that the dynamical processes on the different interacting networks were the same. In many cases, this is not a realistic assumption, and it seems to us that an important future direction of research is understanding and classifying coupled systems where the dynamics are different. An example of this type of system are so-called *supervisory* control systems. Here, an underlying network such as the electrical distribution grid is coupled to an Information Communications Technology (ICT) network for the purposes of monitoring and control. For this case, failure spreading rules are different in each layer, and therefore the stability of the system is very difficult to predict [30].

References

1. S. M. Rinaldi, J. P. Peerenboom, T. K. Kelly. Critical infrastructure interdependencies, IEEE Control. Syst. Magn. **21**,11 (2001).
2. D. Asprone, M. Cavallaro, V. Latora, G. Manfredi, V. Nicosia (2013).
3. R. H. Samet, Futures **47** 49–58 (2013).
4. C. Barrett, K. Channakeshava, F. Huang, J. Kim, A. Marathe, M. V. Marathe, G. Pei, S. Saha, B. S. P. Subbiah, A. K. S. Vullikanti, PLoS One **7**, e45406 (2012).
5. M. Barthelemy, Phys. Rep. **499**, 1 (2011).
6. D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins, Proc. Natl. Acad. Sci. (USA) **102**, 11623–11628 (2005).
7. L. Guibas and J. Stolfi, ACM T. Graphic. **4**, 74 (1985).
8. W. L. Garrison, Regional Science **6**, 121–137 (1960).
9. K. Kansky, *Structure of transportation networks: relationships between network geometry and regional characteristics* (University of Chicago Press, Chicago, 1969).
10. P. Haggett and R. J. Chorley, *Network analysis in geography* (Edward Arnold, London, 1969).
11. E. J. Taaffe and H. L. Gauthier Jr., *Geography of transportation*, (Prenctice Hall, Englewood Cliffs, NJ, 1973).

12. J.-P. Rodrigue and C. Comtois and B. Slack, *The geography of transport systems*, (Routledge, New York, NY, 2006).
13. F. Xie and D. Levinson, Geographical analysis, **39**, 336–356 (2007).
14. E. Strano, V. Nicosia, V. Latora, S. Porta, M. Barthelemy, Nature Scientific Reports **2**:296 (2012).
15. R. Parshani, S. Buldyrev, and S. Havlin, Phys. Rev. Lett. **105**, (2010).
16. S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, Nature (London) **464**, 1025 (2010).
17. X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, Phys. Rev. E **83**, 065101(R) (2011).
18. C.-G. Gu, S.-R. Zou, X.-L. Xu, Y.-Q. Qu, Y.-M. Jiang, D. R. He, H.-K. Liu, and T. Zhou, Phys. Rev. E **84**, 026101 (2011).
19. J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, Nature Phys. **8**, 40 (2011).
20. C. D. Brummitt, R. M. D'Souza, and E. A. Leicht, Proc. Natl. Acad. Sci. USA **109**(12) (2012).
21. A. Saumell-Mendiola, M. A. Serrano, M. Boguna, arXiv:1202.4087, (2012).
22. S. Carmi, Z. Wu, S. Havlin, and H. E. Stanley, Europhys. Lett. **84**, 28005 (2008).
23. W. Li, A. Bashan, S. V. Buldyrev, H. E. Stanley, and S. Havlin, Phys. Rev. Lett. **108**, 228702 (2012).
24. A. Bashan, Y. Berezin, S. V. Buldyrev, and S. Havlin, arXiv:1206.2062, (2012).
25. A. Bunde and S. Havlin, *Fractals and Disordered Systems* (Springer, New York, 1991).
26. R. G. Morris and M. Barthelemy, Phys. Rev. Lett. **109**, 128703 (2012).
27. A. Reynolds-Feighan, J. Air Transp. Manag. **7**, 265 (2001).
28. P. M. Dixon, J. Weiner, T. Mitchell-Olds, and R. Woodley, Ecology **68**, 1548 (1987).
29. I. Lampropoulos, G. M. A Vanalme, W. L Kling, ISGT Europe 2010 IEEE PES (IEEE, New York, 2010).
30. R. G. Morris and M. Barthelemy. Sci. Rep. **3**, 2764 (2013).

Chapter 8
Electrical Networks: An Introduction

S. Pahwa, M. Youssef and C. Scoglio

8.1 Introduction to Electrical Networks

A world without electricity is beyond our imagination. Starting from the prehistoric times, man has made much progress in every walk of life. We have become accustomed to getting everything at the flick of a switch, touch of a button, or turn of a knob. While we have become so used to enjoying the benefits of electricity, it is not easy to imagine how electricity travels from its source to our homes and offices. It sometimes has to cover large distances through a complex network of transmission lines and power substations to provide us the facilities and entertainment that we take for granted. This network which transports electricity from the source to the consumers is called the electrical network. The electrical network is a collective term for different components such as transformers, transmission lines, substations, and different stages and sub-networks devoted to generation, transmission, and distribution. Sometimes, there may be sub-transmission and secondary distribution networks too. A simple schematic of an electric network is shown in Fig. 8.1. In the past decade, analysis of the electrical power system as a complex network has been an evolving and challenging topic of research.

S. Pahwa (✉)
Electrical and Computer Engineering, Kansas State University, 2082 Rathbone Hall,
Manhattan, KS 66506, USA
e-mail: sakship@ksu.edu

M. Youssef
Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute,
1880 Pratt Drive, Research Building XV, Blacksburg, VA 24060, USA
e-mail: myousssef@vbi.vt.edu

C. Scoglio
Electrical and Computer Engineering, Kansas State University, 2069 Rathbone Hall, Manhattan
66506, USA
e-mail: caterina@ksu.edu
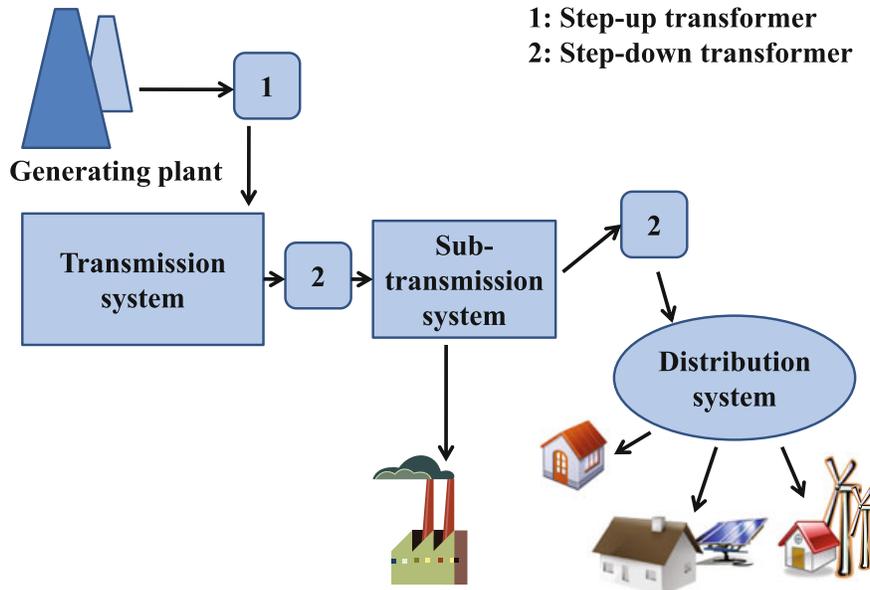
Fig. 8.1 A block schematic of an electrical network. The *transmission system* operates at the highest voltage. The *sub-transmission* operates on medium voltage levels, while the *distribution system* operates on low voltage

8.1.1 Overview

As mentioned, a general electrical network consists of three main parts: Generation, Transmission, and Distribution. The aim of the electrical network is to transport electricity from the source to the consumers. The transmission sub-system is the backbone of the complete electrical network and connects all the main load centers in the system to the main generating plants, while operating at the highest voltage level. Sometimes, there is no clear distinction between the transmission and sub-transmission networks and sometimes they are distinctly separated from each other. The generation and transmission sub-systems together constitute the bulk power system. The distribution sub-system consists of the final stage of power transfer to the individual consumers [1].

The transmission sub-system has been the most studied of the sub-networks of the electrical network. The mesh structure of the transmission network makes it particularly interesting to study different problems on this network. The distribution sub-network usually has simple topologies such as a tree or a ring. Most radial networks are meshed networks initially. However, for a better analysis of their protection schemes and losses, they are always represented and analyzed as an equivalent radial or ring network [2].

All real systems can be modeled into graphs with the individual entities of the system as the nodes and the connections between these entities as the links. The types of connections help us to classify these graphs as random, scale-free, hybrid, or some

other kind of topology. When dynamics are considered on these graphs, either on nodes or links, they are referred to as networks. It is the particular dynamic and the models considered for the analysis of these dynamics which distinguish different networks from each other.

The study of power grid as a complex network started a little before the beginning of the last decade, and gained more importance and momentum after the North American blackout of 2003 [3] and other European blackouts that followed in the same year [4–6]. A basic electrical network can be regarded as a connection of buses through transmission lines, where every bus carries a load or demand that must be satisfied by the power flowing through these lines. Every electrical network must follow the basic laws of Physics called the Kirchoff's equations. These are equalities that deal with the conservation of charge and energy in electrical circuits.

Most of the complex network analysis has been carried out on high voltage transmission grids because their structure is mesh-like and it projects a complexity that is very interesting in the study of different characteristics of electrical networks. In general, power grid networks tend to be sparse networks, as indicated by the real IEEE power grid data available at [7]. The distribution grid tends to be a ring or radial structure but the average node degree of the high-voltage grid is also small. There is a common agreement among the researchers that the average degree of the transmission grid is between 2.5 and 5 (for both, American and European grids), even though the node degree distribution has been a topic of discussion. It seems intuitive to note from a security point of view that a power grid should not have topological hubs. There are some nodes which are "critical" since they carry a very high load, yet their degree does not seem to be too large relative to the size of the grid.

Every node in the electrical network is characterized by a finite capacity, described by the maximum load that the node can carry. Similarly, every link is characterized by a capacity which indicates the maximum amount of power flow that the link can carry, without overheating, melting or discharging. The capacity of the link is not the only factor governing the flow of power on the link. It is important to incorporate the flow dynamics into the topological model of the electrical network through some power flow models. A simple approximation of the complete AC power flow model is the linearized DC model [8]. The set of equations of this model not only incorporate the Kirchoff's laws but also give a rule for the flow of power through each link based on its reactance and the phase angles of the nodes at the two ends of the link. The model is discussed in more details in Sect. 8.2.

This field has constantly evolved and continues to evolve. There are many more challenges that the electrical network of the future offers and the answers are hidden, at least partly, in the structure and design of the networks. The field of complex networks has a huge responsibility of understanding these challenges and bringing forth the answers.

8.1.2 Vulnerability Analysis

Major work has been done in the area of robustness studies of electrical networks, considering the increasing occurrences of power grid blackouts all over the world, in the last decade. Cascading failures used to be a rare phenomenon, but with the unprecedented dependence on the electricity infrastructure, bad practices, and the lack of restructuring of the system, it is now becoming an occurrence which can be heard of several times a year [3, 9–20]. Several times, these cascading failures are triggered by small local disturbances which spread throughout the network due to the complex flow dynamics of the electrical system.

At the very beginning, a study of the cascade spread models, as adapted to power grids was emerging. One of the first studies in this area was presented in [21] in which a very simple model for spread of disturbances in power transmission grid was proposed. This model considered a network of nodes, each representing a power generation or transmission element and the connections between these nodes were not transmission lines, but coupling between the nodes, which was a way to simulate the circuit equations in a real power network. Every node was characterized by a load and a threshold of the load that it could handle. Whenever a node reached its load threshold, load was randomly transferred to the neighboring nodes. The power grid networks considered for evaluation were either well-defined simple ring networks or ring-like structures with some randomness to add a paradigm of complexity as compared to the simple ring structures. The ring-type random power grid networks considered in this work were characterized by their path lengths and clustering coefficient.

Another model, known as the "capacity" model, was proposed in [22] and was supposedly the first "dynamic" model suggested for the power grid. This model considered the flow of a quantity between two nodes through a shortest path and the load on a node was the total number of shortest paths through that node, or in other words, the betweenness of the node, as mentioned in [23]. The capacity of the node was the maximum load it could handle and it was assigned to each node in proportion to the initial load carried by the node. If there was an overload failure in a node, there was a load redistribution among the neighboring nodes of the failed node, or in other words, the redistribution of the shortest paths in the neighboring nodes. The results indicated that this redistribution can lead to a cascade of overload failures in networks with a heterogeneous distribution of loads.

In the "efficiency" model for dynamic complex networks, presented in [24], the focus was also on cascades caused by overloading of nodes and the subsequent redistribution of the flow on the network but this model was different from the other models because it did not remove the overloaded nodes from the network but simply reduced the efficiency of the flow through this node. As a result, it caused a redistribution of flows through other nodes, indirectly redistributing the shortest paths and the damage to the network was quantified as the decrease in global efficiency, as described in [25]. This work mentioned that the degree distribution of an electrical

power grid is exponential but it is heterogeneous in the distribution of loads in the network.

The same model has also been used, specifically for the Italian electric power grid in [26] with the result that the grid is very vulnerable to those failures that occur on the nodes with the highest betweenness. This model distinguished between the nodes as generators and distribution stations, similar to [27] in which the structural vulnerabilities of the North American power grid were evaluated. The authors of this work also agree with [28] in noting that the degree distribution of the North American grid, similar to the Western power grid, is exponential. It is a single-scale network and there is a cost involved with addition of each edge. They also showed the vulnerability of the North American grid based on the edge range approach, discussed in [29], which was one of the first few works to consider attacks on links. They further go on to discuss a connectivity loss measure to find the number of generator nodes that are connected to any given substation node. The authors of [30] formulate a bi-level mixed integer nonlinear programming problem to identify the small groups of lines, which if removed, can cause a severe blackout.

In [31], two real power grid networks, the Nordic grid and the Western States US grid have been studied, their topological characteristics with respect to cascading failures have been compared and these results are further compared with networks from two theoretical models, the Erdos-Renyi random network model [32] and the Barabasi-Albert scale-free network model [33]. These comparisons show clearly the similarities and differences of the two real power grids with respect to the theoretical models, as well as with respect to each other. Some important topological characteristics of the two real grids are highlighted in this study. The robustness of the European power grids under intentional attack has been tested in [34] by selective node removal process. A mean field analysis of the fragility is also presented.

In [35], the authors have presented an initial evidence of the electrical network possessing a self-organized criticality and have studied the global dynamics related to the cascading failures using time-series correlation data of power system blackout sizes. Two types of transitions in the cascading failure blackouts were suggested in [36]. They show that the probability distribution of the blackout size of the North American blackout data has a power tail. This work was followed up in [37], where it was established that the power system is indeed a self-organized critical system. The total number of transmission lines tripped and the total amount of load shed were the measures used to quantify the size of the cascade in [38]. Load shed is the amount of load intentionally removed from the system to bring the system back to a stable state from the disturbed state. When there are failures, especially those which lead to the loss of the system elements causing a large redistribution of load, load shedding becomes necessary to curtail the excess load in the disturbed system, which can be restored after system stability is achieved. Load shedding, although a last resort measure, can be useful to prevent a total blackout of the system. In the above work, the authors use a Galton–Watson branching process to approximate the cascading process of load shed in blackouts.

8.1.3 Mitigation Strategies

Proposing mitigation strategies for preventing the spread of cascades has become the need of the hour and is a way to suggest solutions to the problem at hand. The work in [39] discusses a method to reduce the size of a cascade in complex networks with a heterogeneous degree distribution, after the initial failure has taken place, but before it begins to spread throughout the network. This method has been applied to the electric power network and involves making costless modifications to the network in a time less than that would take the initial failure to spread. It talks about strategies for intentional removal of nodes or links that would significantly reduce the size of cascades.

The probabilistic hidden failure model, which throws light on the protective system failures, was proposed in [40]. Hidden failures in the elements of protection systems were considered to be one of the leading causes of cascading failures in electrical power grids, after the 1996 blackout of the Western grid of the United States. More work on such reliability study was undertaken soon after, as seen in [41–43]. This hidden failure model was further adopted in blackout propagation and mitigation studies in [44–46]. These works included the linearized DC power flow model, to account for the underlying dynamics.

In [47], three mitigation strategies have been discussed for mitigating cascading failures in power grids. Two of the proposed strategies are load shedding strategies while the third one is intentional islanding using distributed sources. Intentional islanding is the intentional splitting of the power grid into sub-parts with their own generation so that these sub-parts can sustain on their own when separated from the remaining network. In this work, islanding is performed using modularity. If all the islands do not have a generator after the first step of islanding, a second step called super-islanding is performed. A polynomial time optimal load shedding algorithm is presented in [48] to control cascading failures occurring due to deterministic failures. The author also proposes another algorithm for stochastic failures. All these models justify the use of DC power flows for the reason that during emergency situations, a faster and always converging solution is needed, especially if the network size is large.

8.1.4 Vulnerability Indices or Robustness Metrics

Several metrics and vulnerability indices have been suggested as a way to identify nodes and links which play an important role in the spread of the cascade. A vulnerability index based on identification of vulnerable links by weighted betweenness of the links is proposed in [49]. The weights on the links are represented by the reactances of the links and the shortest electric distance is represented as the sum of the weights along the shortest electric path, where the shortest electric path between two nodes is the path whose sum of the weights is the smallest among all possible paths

between the two nodes. They also do time domain simulations which verify that their vulnerability metric can not only identify the most critical lines in the system but also those lines which may be vulnerable due to their position in the system, even though they are lightly loaded. They tested their results on the IEEE 39-bus test system and the Huazhong–Chuanyu power grid.

The concept of random-walk betweenness was introduced in [50] where an example of an electric circuit was used to show the effectiveness of the method. It is based on random walks, counting how often a node is traversed by a random walk between two other nodes. It is a generalized technique that may be used for the analysis of power grid networks.

Attention shifted towards the use of power flow model along with the topological models since it was being realized that all the information about power grids was not being captured by purely topological models, although they provided useful information about the structure of the system. The work in [51] talks about the electrical centrality measure for power networks considering the electrical topology rather than the physical topology. They mention the use of the standard AC power flow model for this work [8], without going into the details of the model. The flow propagates through the path of least resistance, and this flow distribution is governed by the relative complex impedance of each path. Also, there can be several paths through which power can flow between two nodes. They use the bus impedance matrix or the inverse of the admittance matrix to define electrical distance between nodes and use this information to represent an electrical topology. They present a conclusion that electrically, the power grid is a scale-free network, although a lot of topological studies indicate a single-scale structure [27, 28, 52–54], while a few show a scale-free structure [33, 55]. Similar work has been done in [56], without the use of any power flow model. Other centrality measures, based on not only the topology but also the electrical parameters of the grid are investigated in [57]. "Efficiency" of the network, as mentioned in previous works was replaced by "net-ability" in [58]. The results obtained using efficiency and net-ability were compared with the reference DC power flow model and net-ability emerged to be a better metric than efficiency. Another metric called "entropic degree" was presented in [59], along with net-ability. These findings were further strengthened by the work presented in [60], in which the authors use the DC power flow model with the IEEE 300-bus system [7]. The results of this work indicate that although topological models can provide the general vulnerability trend, they may not be realistic to suggest any risk mitigation resources without the help of physics-based models.

A metric, η, to measure the robustness of a power grid network with respect to cascading failures was discussed in [61], based on probability of link survival as well as the average rank of the link. Probability of link survival is calculated as the ratio of the number of times a particular link failed due to the removal of an initial link to the total number of links in the network, while the average link rank is calculated depending on the stage of cascade that the link fails at, considering different initial failures. The average depth of cascade is then the product of the link survival probability and the average link rank and it is used to determine η.

The long-term reliability effects for an electric transmission grid, evolving over time, are explored in [62]. The authors take into account policies such as N-1 criterion as well as direct response policy to quantify the reliability of the evolving transmission grid with respect to cascading line overloads and outages as well as slow load growth. The N-1 policy is the standard policy which ensures the upgrade of the transmission lines to satisfy the requirement that a single outage does not lead to overloading of the other transmission lines. The direct response policy leads to an upgrade of the transmission lines involved in the cascading outage that led to some load-shedding. The authors have compared the long-term effect of these policies on the probability distribution of outage size with different number of contingencies.

An electrical power system can be considered as robust only if it can operate in a state of equilibrium, not just in normal but also in perturbed conditions. This was a possibility until a few years ago, but in the current state of affairs, an upgrade in the electrical infrastructure is definitely called for. The stability study of electrical networks, in general, is a multi-disciplinary piece of work, involving fields like electrical and computer engineering, physics, networks, controls, and others. It depends on many natural and human factors which lead to one or more of the events such as load or generation change, short circuit of transmission lines, which is regarded as a link failure in network theory, and other behavioral changes.

8.1.5 Network Generation Models

The real power grid structure does not exactly fit any of the existing network models. Also, a power grid being critical infrastructure, very limited information and data of the real grids are available easily for analysis. With this motivation, a few researchers presented new models such as those mentioned in [47, 63, 64] for generation of synthetic power grid networks. Most of these models considered some of the important characteristics of the real power grids as a part of their network generation algorithm so that the synthetic grids are not completely unrealistic. In [47], the authors designed a first approximation network generator to produce networks having characteristics similar to the real power networks. They have developed a variation of the Generalized Random Graph model to generate power grid networks with realistic degree distributions. The original Generalized Random Graph model [65, 66], which is also known as the Configuration model [67], produces graphs with random connections but with predefined degree sequences. The network generation model in [47] does not impose a degree distribution but imposes certain constraints on the distribution such as the maximum possible degree of a node and the average node degree. The degree distribution of the generated networks using the variation of the Generalized Random Graph model follows the degree distributions of the real power grid networks closely. The impedance and load data is generated from the distributions obtained from the real power grid networks available at [7].

The random topology generation model discussed in [57] uses a variation of the Small World network model, named as the RT-nested-SmallWorld model. It

functions in three steps—(i) creating connected sub-networks, (ii) introducing lattice connections to connect these sub-networks, and (iii) generating line impedances from a specific distribution to assign to the links in the network. The first step varies from the Small World model in the way the links are selected and rewired. A number of links k are randomly selected from a local neighborhood and then a Markov transition is used to rewire a group of links to form sub-networks. The second step is executed by randomly selecting lattice connections to connect the sub-networks into a single large network, with the number of lattice connections roughly around $< k >$. The third step uses a heavy-tailed distribution, to generate line impedances for each link in the network and depending on whether the link is a local link or a lattice connection, a low or high value of impedance is assigned.

Finally, in [64], the authors propose a minimum distance graph generation model to generate synthetic power grid topologies. According to this method, the links are connected such that the Euclidean distance between the two nodes is minimum. This is incorporated in this model from the fact that the link costs increase with an increase in geographical distance in a power grid. A variation of the minimum distance graph model, known as the minimum distance graph with bisection is also proposed to allow for addition of new nodes in the network. This model introduces a new cost called the "bisection cost"; if the cost for creation of a new link is lower, the algorithm creates a new link, else the new node bisects an existing link. The power grid topologies created by this model had many properties similar to the real power grids such as fairly high clustering co-efficient, exponential degree distribution, and characteristic path lengths scaling linearly with number of nodes, n. These graphs had positive assortativity which is the only characteristic different from real grids.

8.1.6 Interconnected Networks and Grid of the Future

Restructuring the electrical network would be an important step in the planning and design of the future electrical network. However, implementation and operation also need to be changed and this realization has driven us towards the Smart Grid. The Smart Grid would lead to many changes in the current infrastructure of the electrical power system, including heavy incorporation of decentralized distributed generators (DGs), renewable energy resources, energy storage, bidirectional flows, improved communications, higher security, climate change mitigation, an increased degree of interconnections [68] and above all, the need for systemic governance.

Interconnections are an inherent part of electrical systems, whether it is the interconnection of several electrical grids or the interconnection of electrical grids with other complex systems such as communication networks. A recent report by the World Energy Council discusses the importance of interconnecting different grids, even across the borders, to fulfill the increasing energy demands of the world [69]. At the same time, it also talks about the challenges that would be faced for such interconnections. Several approaches to model the interdependence between the telecommunication network and the electrical network are discussed in [70], including the

use of Bayesian networks and "precedence graphs". A simulative approach has been used to evaluate the interdependence between the communication and power grid networks in [71] using MPLS [see Chap. 11 in this book]. The results show how a fault in the communication network may propagate to the connected power grid and lead to failures in the latter.

As studied in [72], interconnected networks behave very differently with respect to failures in comparison to single networks. When there is a failure in one network, the dependent nodes in the other network also fail and this may result in a cascade of failures in the interdependent networks. They study the percolation threshold for interconnected networks which is much larger than that of a single network. This study is continued in [73, 74]. In order to understand how interdependence among systems affect the cascading behaviors, the authors in [75] study a sandpile model on modular random graphs as well as graphs based on interdependent power grids. They show both, the advantages and disadvantages of interdependent systems and conclude that some interdependence is beneficial but too much interdependence opens up new possibilities of cascading failures.

In general, electrical networks usually do not operate in isolation. There is usually some kind of loose tie between the electrical systems owned by different utilities, within a region, a country or even between neighboring countries. Whenever failures take place, there is always a risk that the initial failure that occurred in one part of the interconnected power grid might spread to the other parts. This is what happened in the very recent blackout in India, where the Northern, Eastern, and the North Eastern grids were affected due to the failure that occurred in one location [20]. Power grid intentional islanding is gaining a lot of importance as a mitigation strategy for cascading failures in interconnected power grids. However, it is also necessary that the island creation does not lead to further failures in the system and cause excessive load-shedding. Multiple approaches to intentional islanding have been suggested to find the optimal set of lines to be disconnected, including modularity, mixed-integer non-linear programming, spectral matrix methods, simulated annealing, slow-coherency based methods and many others [47, 76–83].

Some of the basic quantities that are usually monitored in case of an electrical network using an AC model are voltages, currents, power, and phase angles. In case of a DC model, the number of quantities to be monitored reduces to real power and phase angles, which are closely related to each other. Some of these basics, along with the two power flow models are explained in Appendix.

8.2 Concluding Remarks

Electrical power grids as critical infrastructures continue to evolve and pose newer challenges. While topological models give important information about the structure of the grids, the electrical models add information about the complex flow dynamics. It is very important that the topological and electrical models are incorporated into each other and work hand in hand for the planning and restructuring of the grid, and

for the implementation of proper measures to make it robust to all kinds of failures. Also, further investigation into the design of interconnected networks, such that the pros are higher than the cons, is essential. In the present times when the demand for electricity is ever increasing, a proper restructuring could be the key to more robust and stable interconnected electrical grids.

Appendix: Electrical Network Terminology and Models for Analysis

Phasor Representation

The voltage and current in electrical power systems are sinusoidal quantities that vary with time at the same frequency. A sinusoidal voltage $v(t)$ and a sinusoidal current $i(t)$ are expressed as follows:

$$v(t) = V_m cos(\omega t + \delta) \tag{8.1}$$

$$i(t) = I_m cos(\omega t + \beta) \tag{8.2}$$

where V_m and I_m are the maximum voltage and current, ω is the angular speed, and δ and β are the phase shift of the voltage and current, respectively.
The voltage is expressed in Volt. However, power systems operate on voltages that range from several 1,000s to 100,000s of Volt. Consequently, it is more convenient to express the voltage in KiloVolt (KV). Electrical current is measured in Ampere (A), the angular speed in radian per second (rad/sec), and the phase shift in radian (rad). The angular speed is proportional to the electrical frequency f, which is the number of cycles per second, expressed in Hertz (Hz). The relationship between the angular speed and the frequency is

$$\omega = 2\pi f. \tag{8.3}$$

As the voltage and current have the sinusoidal form at steady state, it is convenient to express the magnitude and phase angle of the voltage in a complex number form called a phasor. A phasor is developed using the Euler's identity as follows:

$$e^{\pm j\phi} = cos\phi \pm jsin\phi. \tag{8.4}$$

The terms $cos\phi$ and $sin\phi$ are the real and imaginary parts and they are denoted by $Re\{e^{j\phi}\}$ and $Im\{e^{j\phi}\}$, respectively. Therefore, the voltage and current can be written in the phasor form as follows

$$v(t) = Re\{V_m e^{j(\omega t+\delta)}\} = Re\{V_m e^{j\omega t}e^{j\delta}\} \tag{8.5}$$

$$i(t) = Re\{I_m e^{j(\omega t+\beta)}\} = Re\{I_m e^{j\omega t}e^{j\beta}\}. \tag{8.6}$$

Since both the voltage and current have the same frequency, the component $e^{j\omega t}$ becomes less important, and for convenience it is enough to express the voltage and current in terms of their magnitude and phase shift using the following form

$$V = V_m e^{j\delta} = V_m \angle \delta \tag{8.7}$$

$$I = I_m e^{j\beta} = I_m \angle \beta. \tag{8.8}$$

The voltage and current are usually represented through their effective values, called the root-mean-square (rms) values. The effective phasor representations of the voltage and current are

$$V = \frac{V_m}{\sqrt{2}} e^{j\beta} = |V| e^{j\beta} \tag{8.9}$$

$$I = \frac{I_m}{\sqrt{2}} e^{j\delta} = |I| e^{j\delta} \tag{8.10}$$

where $|V| = \frac{V_m}{\sqrt{2}}$ and $|I| = \frac{I_m}{\sqrt{2}}$ are the rms values for the sinusoidal form of the voltage and current. The rms voltage phasor and rms current phasor can be written in the rectangular form as follows

$$V = |V|(cos\delta + jsin\delta) \tag{8.11}$$

$$I = |I|(cos\beta + jsin\beta). \tag{8.12}$$

Instantaneous Power

The electrical power is the work done by the electrical system in unit time. It is a function of both the voltage and current. The unit of electrical power is Watt, however it is convenient to use MegaWatt (MW) when dealing with large amounts of power generation and loads. In a closed circuit with a voltage source and a load, the instantaneous power $p(t)$ that is absorbed by the load is the product of the instantaneous voltage difference across the load and the instantaneous current passing through the load. Mathematically, the instantaneous power is evaluated as follows

$$
\begin{aligned}
p(t) &= v(t)i(t) \tag{8.13}\\
&= V_m cos(\omega t + \delta) I_m cos(\omega t + \beta)\\
&= \frac{V_m I_m}{2}[cos(\delta - \beta) + cos(2(\omega t + \delta) - (\delta - \beta))]\\
&= \frac{V_m I_m}{2} cos(\delta - \beta)[1 + cos(2(\omega t + \delta))]\\
&+ \frac{V_m I_m}{2} sin(\delta - \beta)[sin(2(\omega t + \delta))]
\end{aligned}
$$

The instantaneous power is composed of two components as shown in Eq. (8.13). Assume that the phase angle difference $\delta - \beta$ is constant. The first component is sinusoidal function with a frequency that is twice the frequency of the voltage and current. The maximum value equals $V_m I_m cos(\delta - \beta)$ and the minimum value equals zero. The constant term $\frac{V_m I_m}{2} cos(\delta - \beta)$ represents the average power, while the time-varying sinusoidal function has zero average. We refer to the first component as the instantaneous active power. The second component is time-varying sinusoidal function with zero mean value, twice the frequency of the voltage and current, and maximum value of $\frac{V_m I_m}{2} sin(\delta - \beta)$. This component is called the reactive power, which represents the power that oscillates with twice the frequency of the voltage and current between the reactive components in the power systems that stores the electrical energy and the power generation. In other words, the component of complex power, that averaged over a complete cycle of the AC waveform, causes a net transfer of energy in one direction is known as real power. The component of complex power due to stored energy, which returns to the source in each cycle, is known as reactive power.

Using the rms values for the voltage and current, the active power P and the reactive power Q are as follows

$$P = |V||I|cos(\delta - \beta) \tag{8.14}$$
$$Q = |V||I|sin(\delta - \beta). \tag{8.15}$$

The cosine of the phase angle difference $cos(\delta - \beta)$ is called the power factor. The unit of the active power is Watt, while the unit of the reactive power is Volt-Ampere Reactive (VAR). Let us assume that there are three cases for loads:

- Resistive load: There is no phase angle difference between the voltage and current. Therefore, the power factor is 1, and the active power is $|V||I|$, while the reactive power equals zero because there is no reactive load elements that can store the electrical energy.
- Inductive load: The voltage phase angle leads the current phase angle by $90°$ i.e. $\delta - \beta = \frac{\pi}{2}$. The reactive power is $|V||I|$, while the active power is zero because there is not resistive load elements that can consume the active power.
- Capacitive load: The voltage phase angle lags the current phase angle by $90°$ i.e. $\delta - \beta = -\frac{\pi}{2}$. The reactive power $-|V||I|$, while the active power is zero.

The complex power is defined as a complex number with a real part representing the active power and an imaginary part representing the reactive power as follows

$$\mathbf{S} = P + jQ \tag{8.16}$$
$$\mathbf{S} = \mathbf{V}\mathbf{I}^* \tag{8.17}$$

where I^* is the complex conjugate of the current $\mathbf{I}$. The apparent power S is the magnitude of the complex power

$$S = \sqrt{P^2 + Q^2} \tag{8.18}$$
$$= |V||I|(cos^2(\delta - \beta) + sin^2(\delta - \beta)) \tag{8.19}$$
$$= |V||I|. \tag{8.20}$$

The complex power and the apparent power are related through the following equation

$$\mathbf{S} = S(cos(\delta - \beta) + sin(\delta - \beta)) \tag{8.21}$$

The unit of both the complex power and the apparent power is Volt-Ampere (VA).

Per Unit System

Any power grid is composed of 100s of electrical elements such as transmission lines, transformers, circuit breakers and shunt impedances. Every element can be represented using the ideal form in which it is lossless; however, the ideal form hides many details that influence the performance of a power system. On the other hand, a detailed representation of each element will account for the amount of electrical power loss. Computationally, analysis of detailed representation of power grids is not trivial. Therefore, voltage, current, and power are normalized with respect to their base values, and they become "per unit values". The "per unit" method is a very powerful method for analyzing the power grid because (1) it can be applied to a detailed representation of a power grid, thus reducing the error, and (2) it can be systematically applied to different circuits throughout the power grid, and each circuit has its voltage value close to the normal value. The per unit value is defined as follows

$$\text{Per unit value} = \frac{\text{Actual value}}{\text{Base value}}. \tag{8.22}$$

Both the actual value and the base value have the same dimension, while the per unit value is dimensionless. Traditionally, the base value of the complex power is arbitrarily chosen, and the per unit value becomes as follows

$$\frac{\mathbf{S}}{S_{base}} = \frac{\mathbf{VI}^*}{S_{base}} \tag{8.23}$$

$$\frac{S\angle\theta}{S_{base}} = \frac{V\angle\delta I\angle - \beta}{S_{base}}. \tag{8.24}$$

The base complex power is defined as

$$S_{base} = V_{base}I_{base}. \tag{8.25}$$

In addition to the base complex power, either the base voltage or the base current is arbitrarily chosen. Because a power grid is composed of multiple circuits, each has a voltage level, the base voltage is usually proposed, and the based current is evaluated using Eq. (8.25). Using the base values for the complex power, voltage, and current, Eq. (8.24) becomes as follows

$$\mathbf{S_{pu}} = \mathbf{V_{pu}} \mathbf{I_{pu}^*}. \tag{8.26}$$

We notice that the phase angles do not change using the per unit system, showing that the per unit system is only applied to the magnitude values. The base impedance becomes as follows

$$Z_{base} = \frac{V_{base}}{I_{base}} \tag{8.27}$$

$$= \frac{V_{base}^2}{S_{base}}. \tag{8.28}$$

We further obtain the per unit impedance as follows

$$\mathbf{Z} = \frac{\mathbf{V}}{\mathbf{I}} \tag{8.29}$$

$$\frac{\mathbf{Z}}{Z_{base}} = \frac{\mathbf{V}/V_{base}}{\mathbf{I}/I_{base}} \tag{8.30}$$

$$\mathbf{Z_{pu}} = \frac{\mathbf{V_{pu}}}{\mathbf{I_{pu}}} \tag{8.31}$$

$$\mathbf{Z_{pu}} = \frac{R + jX}{Z_{base}} \tag{8.32}$$

$$\mathbf{Z_{pu}} = R_{pu} + jX_{pu}. \tag{8.33}$$

We notice that the resistance and the reactance have the same base value, which is base impedance

$$Z_{base} = R_{base} = X_{base}. \tag{8.34}$$

Similarly, the active power and the reactive power have the same base value as follows

$$P_{base} = Q_{base} = S_{base}. \tag{8.35}$$

The base complex power is usually expressed in MVA, while the base voltage is expressed in KV. Therefore, it is worth noticing that the base current is in KA, and the base impedance is in Ohm.

Transformers and Transmission Lines

Electrical power is generated at low voltage level leading to increase in the power loss which is proportional to I^2 in the transmission systems. On the other hand, loads do not require high voltages for operation. Transformers are used to step up the voltage from the generation side to the transmission side. Similarly, transformers step down the voltage from the transmission side to the distribution side. Below, we discuss the operation and the representation of the transformers and the transmission lines in more details.

Transformers

A transformer is composed of a primary side and a secondary side. Each side is connected with a winding coil that generates magnetic field, which in turn creates electric current and voltage across the secondary coil. The equivalent circuit of a practical transformer is composed of winding resistance and leakage reactance on each side in which the reactance is added in series with the resistance. In addition, there is power loss in the magnetizing equivalent circuit due to hysteresis current losses. In an ideal transformer, the internal resistances, reactances, and the magnetization circuits are neglected, and the transformer becomes lossless. A practical representation of the transformer is to neglect the magnetization circuit because the magnetizing current is very small compared to the rated current, and to consider the resistances and the reactances of the primary and secondary sides. For transformers that handle large power, the internal resistances become very small compared to the reactance. Thus the internal resistance can be neglected.

Denote the voltages across the primary and secondary coils as $\mathbf{E_1}$ and $\mathbf{E_2}$, respectively. In addition, denote the currents in the primary and secondary sides as $\mathbf{I_1}$ and $\mathbf{I_2}$, respectively. Let the ratio between number of turns in the primary side to number of turns in the secondary side be n. The fact that the complex power at each side of the transformer is preserved, the voltages and currents at both sides are related as follows

$$\frac{\mathbf{E_1}}{\mathbf{E_2}} = \frac{\mathbf{I_2}}{\mathbf{I_1}} = n. \tag{8.36}$$

The reactance of the secondary side x_2 seen from the primary side is $n^2 x_2$. Therefore, the equivalent reactance of the transformer seen at the primary side is the sum of the reactance at the primary side and $n^2 x_2$. The transformer can be represented in terms of per unit as follows

$$\frac{V_{base1}}{V_{base2}} = n \tag{8.37}$$

$$\mathbf{E_{1pu}} = \frac{\mathbf{E_1}}{V_{base1}} \tag{8.38}$$

$$E_{2\mathbf{pu}} = \frac{E_2}{V_{base2}} \tag{8.39}$$

$$E_{2\mathbf{pu}} = \frac{E_1/n}{V_{base1}/n} \tag{8.40}$$

$$E_{2\mathbf{pu}} = E_{1\mathbf{pu}}. \tag{8.41}$$

Similarly, the per unit currents at each side of the transformers are equal. To study the per unit representation of the reactance in the primary side, we have

$$x_2 = \frac{x_1}{n^2} \tag{8.42}$$

$$x_{1pu} = \frac{x_1}{Z_{base1}} \tag{8.43}$$

$$x_{1pu} = \frac{x_1}{V_{base1}^2/S_{base}} \tag{8.44}$$

$$x_{1pu} = \frac{x_2 n^2}{V_{base1}^2/S_{base}} \tag{8.45}$$

$$x_{1pu} = \frac{x_2}{Z_{base2}} = x_{2pu}. \tag{8.46}$$

Therefore, the per unit value of the reactance on one side of the transformer is used when studying the integration of the transformer in the single phase diagram.

Transmission lines

Transmission lines are responsible for transferring the generated power from the generation side to the loads. Depending on the length of the transmission line, the operating voltage is set to reduce the amount of power loss in the lines. Transmission lines with short length require lower voltages than long transmission lines. Transmission lines are classified to short-length, medium-length and long transmission line. A transmission line has an equivalent resistance, inductance, and capacitance. The equivalent π-model is used to represent the transmission lines in the grid. In π-model, the resistance and the inductance are connected in series, and the equivalent capacitance is connected in parallel at the sending and receiving ends of the lines. In short-length transmission lines, the capacitance is neglected, and the transmission line is represented using the series connection of the resistance and reactance. In medium-length transmission lines, half of the total equivalent capacitance is represented at each end of the line, while the series resistance and reactance connection exists between the two ends of the line.

The analysis of the electrical power grid has to be done through a power flow model which can solve the optimal load flow problem considering all the elements described above. In the next two subsections, we discuss two such models for solving the power flow problem.

AC Power Flow Model

To study the power flow in the power grid, assume that the generators, transmission lines and loads locations are given. First we would like to classify the buses into three groups:

- Slack bus: A slack bus produces enough active and reactive power to match the system needs. The voltage and angle at the slack bus are 1 p.u. and zero, respectively, while the generated power P and Q are unknown.
- Load bus: Load bus connects load(s) with the grid. There is no generator connected with the load bus. The amount of active and reactive power needed at the loads are given.
- Voltage controlled bus: Bus that connects a generator with the power grid. Load can be connected on the same bus. The bus voltage and generated active power are known, while the voltage phase shift angle and the reactive power are unknown.

To find the power flow in each transmission line, we first apply Kirchhoff's current law (KCL) at each bus by assuming that the algebraic sum of the currents at any bus is equal to zero. We obtain a group of equations representing the relationship between the voltages and currents, which can be written in a matrix form as follows

$$\begin{bmatrix} \mathbf{Y_{11}} & \mathbf{Y_{12}} & \ldots & \mathbf{Y_{1N}} \\ \mathbf{Y_{21}} & \mathbf{Y_{22}} & \ldots & \mathbf{Y_{2N}} \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{Y_{N1}} & \ldots & \mathbf{Y_{N(N-1)}} & \mathbf{Y_{NN}} \end{bmatrix} \begin{bmatrix} \mathbf{V_1} \\ \mathbf{V_2} \\ . \\ . \\ \mathbf{V_N} \end{bmatrix} = \begin{bmatrix} \mathbf{I_1} \\ \mathbf{I_2} \\ . \\ . \\ \mathbf{I_N} \end{bmatrix} \tag{8.47}$$

Where I_k is the current that enters the bus from the generator/load side. The first matrix is called the admittance matrix or the $\mathbf{Y_{bus}}$ matrix. Each diagonal element $\mathbf{Y_{kk}}$ equals the sum of the admittances of all branches connected to bus k. Every off-diagonal element $\mathbf{Y_{jk}}$ where $j \neq k$ is the sum of admittances of all branches between bus j and bus k multiplied by -1. Using Eq. (8.47), we obtain the following equation at bus k

$$\mathbf{V_1 Y_{k1}} + \mathbf{V_2 Y_{k2}} + \cdots + \mathbf{V_k Y_{kk}} + \cdots + \mathbf{V_N Y_{kN}} = \mathbf{I_k} = \frac{P_k - jQ_k}{\mathbf{V_k^*}}. \tag{8.48}$$

To find all unknown active power, reactive power, voltages, voltage angles, a famous method called Gauss-Seidel iterative approach is used by assuming flat initial solutions for all voltages and voltage angles equal 1 p.u. and zero, respectively. For bus k at iteration $i + 1$, the following iterative equation is used to find the solution of the unknown variables

$$\mathbf{V_k^{i+1}} = \frac{1}{\mathbf{Y_{kk}}}\left[\mathbf{I_k^i} - \sum_{n=1}^{k-1} \mathbf{V_n^{i+1}Y_{kn}} - \sum_{n=k+1}^{N} \mathbf{V_n^i Y_{kn}}\right] \tag{8.49}$$

DC power flow model

A power grid can be considered as a complex network with N nodes and L links. Nodes represent the generation and transmission substations, and links represents the transmission lines. To simplify the power flow analysis in the power grids, the DC Power Flow model has been originally introduced as *DC Power Flow* in the DC network analyzer [84] as suggested by [85]. In the original work, the network branch is represented by a resistance and the resistance value is proportional to the reactance that is connected in series with the resistance and each DC current is proportional to the power flow. The DC power flow model represents a linearization of the full AC model. In the AC model, let V_i and V_j represent the voltage at the buses i and j, respectively. In addition, let Y_{ij} represent the admittance of the transmission line between buses i and j. The relation between real power, complex voltages and line impedance is expressed through the following equation which describes the amount of real power flowing through a transmission line

$$P_{ij} = |V_i||V_j||Y_{ij}|cos(\delta_i - \delta_j + \theta_{ij}) \tag{8.50}$$

where θ_{ij} is the phasor angle of the admittance Y_{ij}. To obtain the DC power flow model, the following assumptions are applied to Eq. (8.50) as follows

- Voltage angle differences are small, i.e. $sin(\delta_{ij}) \approx \delta_{ij}$.
- Flat Voltage profile: All voltages are considered 1 $p.u.$.
- Line resistance is negligible i.e. $R \ll X$.

Applying Taylor expansion on Eq. (8.50) around the operating voltage, and neglect the coupling between the power flow and the voltage, we obtain

$$P_{ij} = \frac{\delta_{ij}}{x_{ij}} \tag{8.51}$$

where δ_{ij} is the difference in phase shift angle between the voltages at the sending and receiving buses, and x_{ij} is the reactance of the transmission line. The DC power flow Eq. (8.51) can be written in matrix form where P is the $N \times N$ matrix of power flows between each node i and j in the network, δ is the $N \times 1$ vector of phase angles and X is the $N \times N$ weighted adjacency matrix, each element of which represents the reactance of a transmission line. It is a real number if a line is present between two nodes, and zero otherwise. In matrix form,

$$[P] = [b][\delta] \tag{8.52}$$

The matrix $[b]$ represents the imaginary part of the Y_{bus} matrix of the power grid, where $b_{ij} = -\frac{1}{x_{ij}}$ and $b_{ii} = \sum_{i \in N} -b_{ij}$ for $i \neq j$. We usually assume that there is a reference node with voltage angle equals 0. The power handled by each node is the net sum of all the ingoing and outgoing power flows at that node as follows:

$$P_i = \sum_{j=1}^{N} P_{ij} = \sum_{j=1}^{N} (-b_{ij} \delta_{ij}) \tag{8.53}$$

The total load at each node is given, while the phase angles are computed using the following equation:

$$[\delta] = [b]^{-1}[P]. \tag{8.54}$$

References

1. KUNDUR, P.: *Power System Stability and Control, Chapter 1: General Characteristics of Modern Power Systems*, McGraw-Hill, Inc.
2. SWARANKAR, A., GUPTA, N., NIAZI, K. R.: *Efficient Reconfiguration of Distribution Systems Using Ant Colony Optimization Adapted by Graph Theory*, Proceedings of the IEEE Power and Energy Society General Meeting, Detroit, MI, (2011).
3. U.S.- CANADA POWER SYSTEM OUTAGE TASK FORCE: *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*,http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/BlackoutFinal-Web.pdf, (2004).
4. LARSSON, S., AND EK, E.: *The Blackout in Southern Sweden and Eastern Denmark, September 23, 2003*, Proceedings of the IEEE Power Engineering Society General Meeting, Denver, CO, (2004).
5. JOHNSON, C. W.: *Analysing the Causes of the Italian and Swiss Blackout, 28th September, 2003*, Proceedings of the 12th Australian Workshop on Safety Critical Systems and Software-Related Programmable Systems, Adelaide, Australia, (2007).
6. WORLD ENERGY COUNCIL, http://www.worldenergy.org/focus/blackouts/390.asp.
7. CHRISTIE, R. D.: *University of Washington Power Systems Test Case Archive*, http://www.ee.washington.edu/research/pstca/.
8. GUNGOR, B. R.: *Power Systems - Chapter 6*, Technology Publications, (1988).
9. NATIONAL GRID TRANSCO: *Investigation Report into the Loss of Supply Incident Affecting Parts of South London at 18:20 on Thursday, 28 August 2003*, http://www2.tech.purdue.edu/eet/courses/eet331/blackout/London03BlackoutExecutiveSummary.pdf, (2003).
10. UNION FOR THE COORDINATION OF THE TRANSMISSION OF ELECTRICITY: *Final Report of the Investigation Committee on the 28th September 2003 Blackout in Italy*, http://www.rae.gr/old/cases/C13/italy/UCTE_rept.pdf, (2004).
11. *Massive Blackout Hits Java, Bali*, The Jakarta Post, http://www.thejakartapost.com/news/2005/08/19/massive-blackout-hits-java-bali.html, (2005).
12. EUROPEAN REGULATORS' GROUP FOR ELECTRICITY AND GAS: *ERGEG Final Report - The Lessons to be Learned from the Large Disturbance in the European Power System on the 4th of November 2006*, http://www.energy-regulators.eu/portal/page/portal/EER_HOME/EER_PUBLICATIONS/CEER_PAPERS/Electricity/2007/E06-BAG-01-06_Blackout-FinalReport_2007-02-06.pdf, (2007).
13. *Melting in Zanzibar's Blackout*, BBC News, http://news.bbc.co.uk/2/hi/africa/7427957.stm, (2008).

14. LYONS, J.: *Brazil Blackout Sparks Infrastructure Concerns*, The Wall Street Journal, http://online.wsj.com/article/SB125798817743744475.html, (2009).
15. *Widespread Power Blackout Hits Chile*, CNN World, http://www.cnn.com/2010/WORLD/americas/03/14/chile.blackout/index.html, (2010).
16. *Fire Cuts Power to Thousands of Portsmouth Homes*, BBC News, http://www.bbc.co.uk/news/10423764, (2010).
17. *Tripping of Taramani Substation Causes Blackout*, The Hindu, http://www.thehindu.com/news/cities/chennai/tripping-of-taramani-substation-causes-blackout/article423673.ece, (2010).
18. *Heatwave Causes Massive Power Outages*, The Huffington Post, http://www.huffingtonpost.com/2011/09/15/south-korea-heat-wave_n_963922.html, (2011).
19. WATSON, I., AND COMERT, Y.: *Istanbul Blackout Leaves Millions in Dark*, CNN, http://edition.cnn.com/2012/01/14/world/meast/turkey-blackout/, (2012).
20. MINISTRY OF POWER: *Report of the Enquiry Committee on Grid Disturbance in Northern Region on 30th July, 2012 and in Northern, Eastern and North-Eastern Region on 31st July, 2012*, http://www.powermin.nic.in/pdf/GRID_ENQ_REP_16_8_12.pdf, (2012).
21. SACHTJEN, M. L., CARRERAS, B. A., AND LYNCH, V. E.: *Disturbances in a Power Transmission System*, Physical Review E, **61**(5), (2000), 4877–4882.
22. MOTTER, A. E., AND LAI, Y- C.: *Cascade-based Attacks on Complex Networks*, Physical Review E, **66**, (2002), 065102.
23. HOLME, P., AND KIM, B. J.: *Vertex Overload Breakdown in Evolving Networks*, Physical Review E, **65** (2002), 066109.
24. CRUCITTI, P., LATORA, V., AND MARCHIORI, M.: *Model for Cascading Failures in Complex Networks*, Physical Review E, **69** (2004), 045104.
25. LATORA, V., AND MARCHIORI, M.: *Efficient Behavior of Small-World Networks*, Physical Review Letters, **87** (2001), 198701.
26. CRUCITTI, P., LATORA, V., AND MARCHIORI, M.: *A Topological Analysis of the Italian Electric Power Grid*, Physica A, **338** (2004), 92–97.
27. ALBERT, R., ALBERT, I., AND NAKARADO, G. L.: *Structural Vulnerability of the North American Power Grid*, Physical Review E, textbf69 (2004), 025103.
28. AMARAL, L. A. N., SCALA, A., BARTHELEMY, M., AND STANLEY, H. E.: *Classes of Small-World Networks*, Proceedings of the National Academy of Science USA, **97** (2000), 11149.
29. MOTTER, A. E., NISHIKAWA, T., AND LAI, Y. C.: *Range-Based Attacks on Links in Scale-Free Networks: Are Long-Range Links Responsible for the Small-World Phenomenon?*, Physical Review E, **66** (2002), 065103.
30. PINAR, A., MEZA, J., DONDE, V., LESIEUTRE, B.: *Optimization Strategies for the Vulnerability Analysis of the Electric Power Grid*, SIAM Journal on Optimization, **20** (2011), 1786–1810.
31. HOLMGREN, A. J.: *Using Graph Models to Analyze the Vulnerability of Electric Power Networks*, Risk Analysis, **26** (2006), 955–969.
32. ERDOS, P., AND RENYI, A.: *On the Evolution of Random Graphs*, Publications of the Mathematical Institute of the Hungarian Academy of Sciences, **5** (1960), 17–61.
33. BARABASI, A. L., AND ALBERT, R.: *Emergence of Scaling in Random Networks*, Science, **286** (1999), 509–512.
34. SOLE, R. V., ROSAS- CALAS, M., COROMINAS- MURTRA, B., VALVERDE, S.: *Robustness of the European Power Grids Under Intentional Attack*, Physical Review E, **77** (2008), 026102.
35. CARRERAS, B. A., NEWMAN, D. E., DOBSON, I., AND POOLE, A. B.: *Initial Evidence for Self-Organized Criticality in Electric Power System Blackouts*, Proceedings of the 33rd Hawaii International Conference on System Sciences, Maui, Hawaii, (2000).
36. CARRERAS, B. A., LYNCH, V. E., DOBSON, I., NEWMAN, D. E.: *Critical Points and Transitions in an Electric Power Transmission Model for Cascading Failure Blackouts*, CHAOS, **12** (2002), 985–994.
37. CARRERAS, B. A., NEWMAN, D. E., DOBSON, I., AND POOLE, A. B.: *Evidence for Self-Organized Criticality in a Time Series of Electric Power System Blackouts*, IEEE Transactions on Circuits and Systems - I : Regular Papers, **51**, (2004), 1733–1740.

38. KIM, J., DOBSON, I.: *Propagation of Load Shed in Cascading Line Outages Simulated by OPA*, Proceedings of the IEEE Workshop on Complexity in Engineering, Rome, Italy, (2010).

39. MOTTER, A. E.: *Cascade Control and Defense in Complex Networks*, Physical Review Letters, **93** (2004), 098701.

40. TAMRONGLAK, S., HOROWITZ, S. H., PHADKE, A. G., AND THORP, J. S.: *Anatomy of Power System Disturbances: Preventive Relaying Strategies*, IEEE Transactions on Power Delivery, **11** (1996), 708–715.

41. BAE, K., AND THORP, J. S.: *A Stochastic Study of Hidden Failures in Power System Protection*, Decision Support Systems, **24** (1999), 259–268.

42. WANG, H., AND THORP, J. S.: *Enhancing Reliability of Power Protection Systems Economically in the Post-Restructuring Era*, Proceedings of the 32nd North American Power Symposium, Ontario, Canada, (2000).

43. CHEN, J., AND THORP, J. S.: *A Reliability Study of Transmission System Protection via a Hidden Failure DC Load Flow Model*, Proceedings of the IEEE 5th International Conference on Power System Management and Control, (2002).

44. CHEN, J., THORP, J. S., AND DOBSON, I.: *Cascading Dynamics and Mitigation Assessment in Power System Disturbances via a Hidden Failure Model*, Electrical Power and Energy Systems, **27** (2005), 318–326.

45. CHEN, G., DONG Z. Y., HILL, D. J., ZHANG, G. H., AND HUA, K. Q.: *Attack Structural Vulnerability of Power Grids: A Hybrid Approach Based on Complex Networks*, Physica A, **389** (2010), 595–603.

46. CARRERAS, B. A., LYNCH, V. E., NEWMAN, D. E., DOBSON, I.: *Blackout Mitigation Assessment in Power Transmission Systems*, Proceedings of the Hawaii International Conference in System Sciences, Big Island, Hawaii, (2003).

47. PAHWA, S., SCOGLIO, C., SCHULZ, N.: *Topological Analysis and Mitigation Strategies for Cascading Failures in Power Grid Networks*, http://arxiv.org/submit/619427/view.

48. BEINSTOCK, D.: *Optimal Control of Cascading Power Grid Failures*, Proceedings of the 50th IEEE Conference on Decision and Control, Orlando, FLA, (2011).

49. CHEN, X., SUN, K., CAO, Y., AND WANG, S.: *Identification of Vulnerable Lines in Power Grid Based on Complex Network Theory*, IEEE Power Engineering Society General Meeting, Tampa, FLA, (2007).

50. NEWMAN, M. E. J.: *A Measure of Betweenness Centrality Based on Random Walks*, Social Networks, **27** (2005), 39–54.

51. HINES, P., AND BLUMSACK, S.: *A Centrality Measure for Electrical Networks*, Proceedings of the 41st Annual Hawaii International Conference on System Sciences, Big Island, Hawaii, (2008).

52. NEWMAN, M. E. J: *The Structure and Function of Complex Networks*, SIAM Review, **45:2** (2003), 167–256.

53. ALBERT, R., AND BARABASI, A.: *Statistical Mechanics of Complex Networks*, Reviews of Modern Pysics, **74** (2002), 47–97.

54. ROSATO, V., BOLOGNA, S., AND TIRITICCO, F.: *Topological Properties of High-Voltage Electrical Transmission Networks*, Electric Power Systems Research, **77** (2007), 99–105.

55. CHASSIN, D. P., POSSE, C.: *Evaluating North American Electric Grid Reliability Using the Barabasi-Albert Network Model*, Physica A, **355** (2005), 667–677.

56. WANG, Y., ZHAO, J., ZHANG, F., AND LEI, B.: *Study of Structural Vulnerabilities of Power Grids Based on the Electrical Distance*, Proceedings of the IEEE PES Innovative Smart Grid Technologies - Asia, Tianjin, China, (2012).

57. WANG, Z., SCAGLIONE, A., AND THOMAS, R. J.: *Electrical Centrality Measures for Electric Power Grid Vulnerability Analysis*, Proceedings of the 49th IEEE Conference on Decision and Control, Atlanta, GA, (2010).

58. ARIANOS, S., BOMPARD, E., CARBONE, A., AND XUE, F.: *Power Grid Vulnerability: A Complex Network Approach*, CHAOS, **19** (2009), 013119.

59. BOMPARD, E., NAPOLI, R., AND XUE, F.: *Analysis of Structural Vulnerabilities in Power Transmission Grids*, International Journal of Critical Infrastructure Protection, **2** (2009), 5–12.

60. HINES, P., COTILLA- SANCHEZ, E., BLUMSACK, S.: *Do Topological Models Provide Good Information About Electricity Infrastructure Vulnerability?*, CHAOS, **20** (2010), 033122.

61. YOUSSEF, M., SCOGLIO, C., AND PAHWA, S.: *Robustness Measure for Power Grids with Respect to Cascading Failures*, Proceedings of the International Workshop on Modeling, Analysis, and Control of Complex Networks - 23rd International Teletraffic Congress, San Fransisco, CA, (2011).

62. REN, H., DOBSON, I., AND CARRERAS, B. A.: *Long-Term Effect of the N-1 Criterion on Cascading Line Outages in an Evolving Power Transmission Grid*, IEEE Transactions on Power Systems, **23** (2008), 1217–1225.

63. WANG, Z., SCAGLIONE, A., AND THOMAS, R. J.: *Generating Statistically Correct Random Topologies for Testing Smart Grid Communication and Control Networks*, IEEE Transactions on Smart Grid, **1** (2012), 28–39.

64. HINES, P., BLUMSACK, S., COTILLA- SANCHEZ, E., AND BARROWS, C.: *The Topological and Electrical Structure of Power Grids*, Proceedings of the 43rd IEEE Hawaii International Conference on System Sciences, Kauai, Hawaii, (2010).

65. BARRAT, A., BARTHELEMY, M., AND VESPIGNANI, A.: *Dynamical Processes on Complex Networks*, Cambridge University Press, (2008).

66. JACKSON, M. O.: *Social and Economic Networks*, Princeton University Press, (2008).

67. BENDER, E. A., AND CANFIELD, E. R.: *The Asymptotic Number of Labeled Graphs with Given Degree Sequences*, Journal of Combinatorial Theory, Series A, **24** (1978), 296–307.

68. WORLD ENERGY COUNCIL: *Smart Grids: Best Practice Fundamentals for a Modern Energy Society*, http://www.worldenergy.org/documents/20121006_smart_grids_best_practice_ fundamentals_for_a_modern_energy_system.pdf, (2012).

69. WORLD ENERGY COUNCIL: *Interconnectivity: Benefits and Challenges*, http://www. worldenergy.org/documents/interconexsum.pdf, (2010).

70. HADJSAID, N., TRANCHITA, C., ROZEL, B., VIZITEU, M., AND CAIRE, R.: *Modeling Cyber and Physical Interdependencies - Application in ICT and Power Grids*, Proceedings of the IEEE PES Power Systems Conference and Exposition, (2009).

71. KUBLER, S., PAHWA, S., SCHULZ, N., AND SCOGLIO, C.: *A Simulative Analysis of the Robustness of Smart Grid Communication Networks*, Proceedings of the North American Power Symposium, Boston, MA, (2011).

72. BULDYREV, S. V., PARSHANI, R., PAUL, G., STANLEY, H. E., AND HAVLIN, S.: *Catastrophic Cascade of Failures in Interdependent Networks*, Nature, **464** (2010), 1025–1028.

73. PARSHANI, R., BULDYREV, S. V., AND HAVLIN, S.: *Interdependent Networks: Reducing the Coupling Strength Leads to a Change from a First to Second Order Percolation Transition*, Physical Review Letters, **105** (2010), 048701.

74. GAO, J., BULDYREV, S. V., STANLEY, H. E., AND HAVLIN, S.: *Networks Formed from Interdependent Networks*, Nature Physics, **8** (2012), 40–48.

75. BRUMMITT, C. D., D'SOUZA, R. M., LEICHT, E. A.: *Suppressing Cascades of Load in Interdependent Networks*, Proceedings of the National Academy of Sciences USA, **109** (2011).

76. FAN, N., IZRAELEVITZ, D., PAN, F., PARDALOS, P. M., AND WANG, J.: *A Mixed Integer Programming Approach for Optimal Power Grid Intentional Islanding*, Energy Systems, **3** (2012), 77–93.

77. TRODDEN, P. A., BUKHSH, W. A., GROTHEY, A., AND MCKINNON, K. I. M.: *MILP Formulation for Islanding of Power Networks*, IEEE Transactions on Power Systems: *In Print*.

78. HAMAD, I. A., ISRAELS, B., RIKVOLD, P. A., AND POROSEVA, S. V.: *Spectral Matrix Methods for Partitioning Power Grids: Applications to the Italian and Floridian High-Voltage Networks*, Physics Procedia, **4** (2010), 125–129.

79. HAMAD, I. A., RIKVOLD, P. A., AND POROSEVA, S. V.: *Floridian High-Voltage Power-Grid Network Partitioning and Cluster Optimization using Simulated Annealing*, Physics Procedia, **15** (2011), 2–6.

80. ROZEL, B., CAIRE, R., HADJSAID, N., ROGNON, J- P.: *Complex Network Theory and Graph Partitioning: Application to Large Interconnected Networks*, Proceedings of the IEEE Bucharest Power Tech Conference, Bucharest, Romania, (2009).

81. PEIRAVI, A., AND ILDARABADI, R.: *A Fast Algorithm for Intentional Islanding of Power Systems using the Multilevel Kernel k-means Approach*, Journal of Applied Sciences, **9** (2009), 2247–2255.

82. YANG, B., VITTAL, V., AND HEYDT, G. T.: *Slow-Coherency-Based Controlled Islanding - A Demonstration of the Approach on the August 14, 2003 Blackout Scenario*, IEEE Transactions on Power Systems, **21** (2006), 1840–1847.

83. WANG, X., AND VITTAL, V.: *System Islanding Using Minimal Cutsets with Minimal Net Flow*, Proceedings of the IEEE PES Power Systems Conference and Exposition, New York City, NY, (2004).

84. HAHN, W. C.: *Load Studies on the D-C Calculating Table*, General Electric Review, **34** (1931).

85. STOTT, B., JARDIM, J., AND ALSAC, O.: *DC Power Flow Revisited*, IEEE Transactions on Power Systems, **24** (2009), 1290–1300.

Chapter 9
Smart Grid as Multi-layer Interacting System for Complex Decision Makings

Ettore Bompard, Bei Han, Marcelo Masera and Enrico Pons

Abstract This chapter presents an approach to the analysis of Smart Grids based on a multi-layer representation of their technical, cyber, social and decision-making aspects, as well as the related environmental constraints. In the Smart Grid paradigm, self-interested active customers (prosumers), system operators and market players interact among themselves making use of an extensive cyber infrastructure. In addition, policy decision makers define regulations, incentives and constraints to drive the behavior of the competing operators and prosumers, with the objective of ensuring the global desired performance (e.g. system stability, fair prices). For these reasons, the policy decision making is more complicated than in traditional power systems, and needs proper modeling and simulation tools for assessing "in vitro" and ex-ante the possible impacts of the decisions assumed. In this chapter, we consider the smart grids as multi-layered interacting complex systems. The intricacy of the framework, characterized by several interacting layers, cannot be captured by closed-form mathematical models. Therefore, a new approach using Multi Agent Simulation is described. With case studies we provide some indications about how to develop agent-based simulation tools presenting some preliminary examples.

E. Bompard (✉) · M. Masera
Institute for Energy and Transport, Joint Research Centre, European Commission, Petten, The Netherlands
e-mail: ettore.bompard@ec.europa.eu

M. Masera
e-mail: marcelo.masera@ec.europa.eu

B. Han · E. Pons
Dipartimento Energia, Politecnico di Torino, 10129 Torino, Italy
e-mail: bei.han@polito.it

E. Pons
e-mail: enrico.pons@polito.it

9.1 Introduction

Environmental issues and electricity supply security concerns are aggravating the pressure towards a smarter electricity system. The integration of distributed generation and storage, both at the local and regional levels, is most promising. Nevertheless, it presents challenging problems regarding the design and implementation of the systems and the related markets. Promising effects (such as the mitigation of environmental problems and more reliable electricity supply) are expected from these developments.

On the other hand, complexities deriving from the use of devices based on new technologies (and mainly those intensely based on information processing and communications), the multiplicity of players and the interactions among them, can hinder the implementation of smart grids. Such complexities have to be considered by decision makers at all regulatory levels. Thus, regulatory decision-making becomes a most complex issue that has to be analyzed ex-ante, so as to ensure the system performance and anticipate any situation that can impair the system.

The general goals from the regulator's point of view are related to the various social, technical, economic and ecological aspects of the system. Those goals target specific issues such as economic efficiency, social welfare (education, health, and security), satisfaction of energy needs, and reduction of environmental impacts.

This chapter represents and studies these emerging complexities of smart grids in a multilayer framework (Sect. 9.2); proposes a MAS (Multi-Agent Simulation) tool that can implement various smart grid scenarios (Sect. 9.3), and finally assesses a simple scenario under different regulatory regimes (Sect. 9.4).

9.2 The Framework of Complex Power System

The essence of the smart grid paradigm is an upgraded efficient, clean, secured and flexible electrical system, satisfying the following seven principles:

- enabling the active participation of consumers in demand response;
- operating resiliently against physical and cyber attacks and hazards;
- providing power quality, security and adequacy;
- accommodating all generation and storage options;
- enabling new products, services, and markets;
- optimizing assets and operating efficiently;
- self-healing from power disturbance events.

Realizing these Smart Grid principles will affect all components of the electricity generation, transmission, distribution and consumption, while also enabling new types of autonomous behavior in all system and market players. A key element will be the active participation of the large population of consumers and prosumers. Traditional optimization of power systems, using mathematical functions, may only

serve for taking into account the technical level of the system. The social and decision-making aspects require a complementary approach.

A multi-layer framework of complex power system is proposed in this section to integrate all the elements of the Smart Grids. The intention is to consider all relevant fields (technical, social, economic and political), for studying the system performance in an aggregated way.

9.2.1 Multi-layer Platform

Our multi-layer platform model includes the power, cyber, social and environment layers, along with threats and factors that may affect the system and lead to unwanted situations. It is shown in Fig. 9.1.

The physical layer contains all the hardware components. It is the container of electricity flows where physical variables and indices are computed, monitored, or optimized by the system operators. Here we concentrate on MV/LV distribution networks, while including data and constraints from the upper transmission level. The arbitrary plug and play connection of distributed generation is enforced by low or zero connection charging policies.

In addition, we consider some capabilities normally proposed in Smart Grids: the network structure can be adjusted according to the real-time demand and generation; the local network can serve local demands under emergence conditions, isolated from the upper grid; the local network can assist in black starts as well as provide security reserve to the upper grid.

In contrast with the physical layer, the cyber layer is the container of information flows, where all the operations and market-related data sets are managed: prosumer generation and consumption values, market prices, physical conditions; operational commands, and so on. Also various technical innovations are required in this layer, such as smart meters, optical and power-line communications, home area communications, wide area measurement system, and so on.

The social layer aggregates the actors of the power system, i.e. users, prosumers and marketers. This layer has been identified as the main source of complexities that lead to the unpredictable performance of the overall system. The value sense of each prosumer is initially decided by factors related to their psychology, education, profession and so on, and then evolving through the interactions within their social network. Also the social network itself is evolving through random relationships and the establishment of new or interruption of existing social links.

The environment layer stands for the natural phenomena, and influence all the other layers. Most obviously, weather conditions, geographical conditions, and primary and secondary resource conditions, directly affect consumption and generation. Above than, society typically imposes sustainability goals that require the respect of several environmental targets (emission, energy efficiency, and so on), with a key role for regulators.
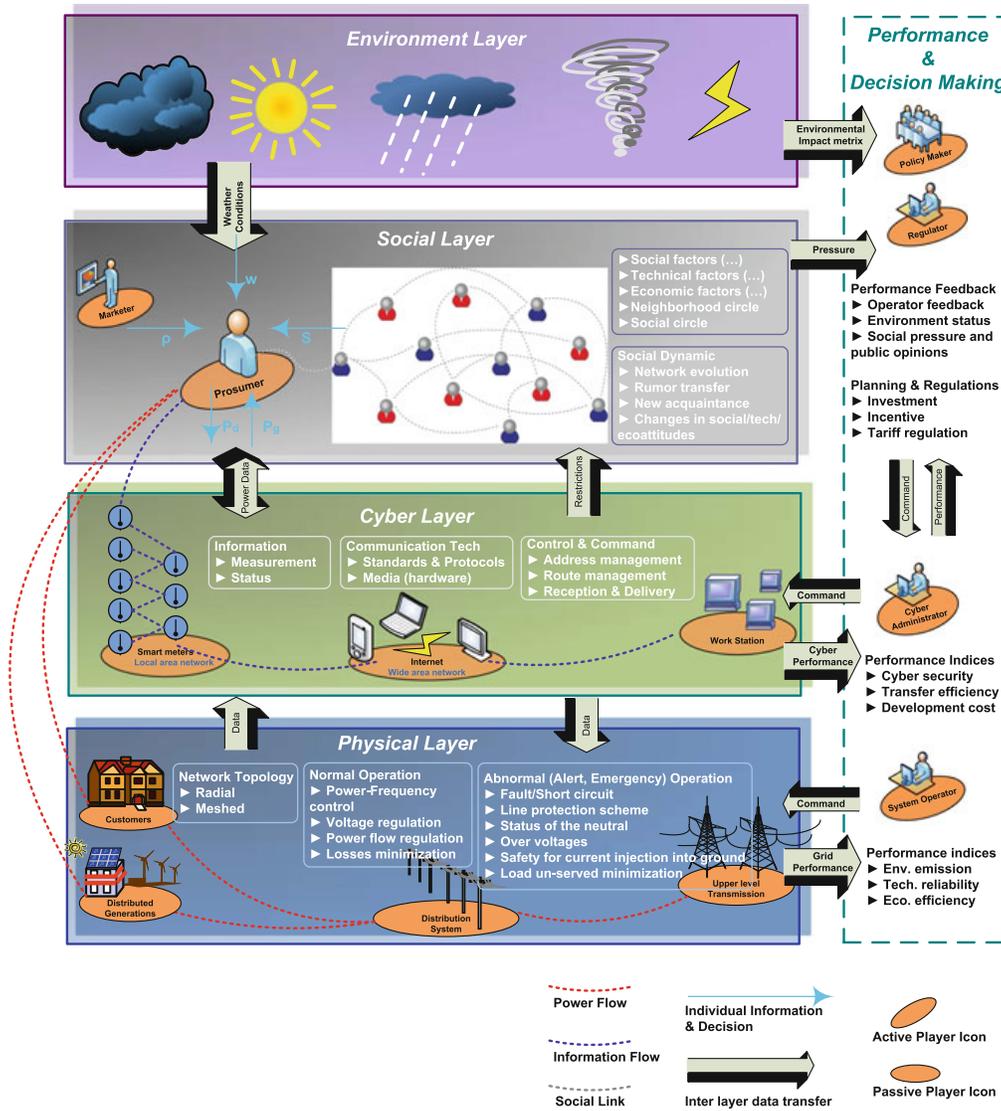
Fig. 9.1 Multi-layer platform of complex power system

Except for the environmental indices mentioned above, the overall performance of the system should include all the other dimensions: physical performance in terms of power security, power quality, reaction under emergency; technical performance in terms of technology penetration, technical reliability and efficiency; social performance in terms of satisfaction of the objectives set by regulations, individuals, and social groups; and market performance in terms of market power allocation, competitiveness, etc.

Of great importance are the interconnections between the different layers, as these connects are at the basis of the arising complexity. For example the weather conditions impact on the social layer (people behavior) and on the physical layer (e.g. distributed generation); the physical layer exchanges data with the cyber layer, which performs

measurements and provides commands, but influences also the environment layer (e.g. with pollutant emissions); the cyber layer is the mean for prosumers (social layer) to interact with the grid (physical layer).

The decision-making processes interact with the other four layers. For example people (social layer) can exercise pressure on politicians for changes in the performance objectives; on the other hand, decision makers can obtain information from the cyber layer and provide commands to it, or can act directly on the physical layer (e.g. the system operators).

9.2.2 Simulating the Social Layer: The Actors and the Stage

The actors having decision-making capabilities in their domain, according to specific roles, are divided into two classes.

The major population under study is composed of the "players", simulated and observed by the other class called "directors", who set the rules, monitor trends, and coordinate the layers. In our model, the players are the Prosumers, Distribution System Operators (DSO) and Retailers; while directors are the Regulators in charge of issuing the rules and exerting the control over the electricity (and more generally energy) markets, and Policy Decision Makers who represent the public institutions, such as parliament, governments, ministries, that fix the general goals and decide the policies.

We call ESTS (which stands for Environmental Social Technical System) the stage over which the players interact according to the rules set by the "directors". We can see an example of the stage in Sect. 3.2. The ESTS name derives from its elements: environmental (weather, emissions), social (relationships, value senses, and so on), and technical (consumption units, distributed generation units, storage units, and so on). The ESTS stage provides inputs to players, defines the possibility of interactions and gets the feedbacks from the evolving system.

9.2.3 Simulating the Interactions

For simulating the system, the interactions among the prosumers and other actors, and the interactions of all layers with the ESTS, are predefined in a common schedule.

We consider that the actions by the prosumers only refer to the next 24-h, and regard their own generation, consumption, and storage—and that these decisions are inputs to ESTS. These decisions are constrained by the available technical options and depend upon the utility choice of each player. As a result, the set of all these options, choices and interactions with the external ESTS, constitutes a complicated decision making space.

The prosumers update their decision-making strategies according to the feedback of the system in terms of the overall performance and the resulting benefits for

themselves. On their hand, with the evolving of the system conditions, the system operators or regulators (and other macro players) update their rules, justify pricing methods, add constraints, provide other technical choices for enabling fully competitive markets and physical connections of DG at the prosumers' sites.

Two approaches can be applied when designing a simulation. The first refers to the ex-ante studies of prosumer behaviors and of regulation strategies for large distributed generation capacities. For example, one can study 24-h periods. The second is a validation process, using real 24-h dispatch and strategic reactions from different market and system players, with the objective of carrying out an ex-post analysis of the interactions among the players and the system performance.

9.3 Design of MAS Models

Multi-Agent Simulation (MAS) can be employed for prototyping, implementing and simulating complex power system and their actors. Agents instantiate the various elements concerned according to the different system or market status and behaviors. Following predefined social, economic or technical rules, the agents, embedded with either simple or sophisticated decision making approaches, communicate among themselves and update information about the ESTS. As the simulation runs, system dynamics induced by the autonomous agent behaviors can be observed.

In this section, we present solutions to design issues identified as problematic in the development of MAS for smart grid simulations.

9.3.1 Issues of Implementation

When implementing a model, the following issues have to be solved:

- Which is the behavior of the individual agents (prosumers)? How will they react to changes in the ESTS and to inputs from their own neighborhood? How many different types of behaviors have to be categorized? What is the typical decision making process for each of them?
- Which are the strategies to be taken by the coordination agents (DSO or marketer) to ensure the attainment of the desired performance, at both objective levels (local, for each particular operator; global, for the overall system)? What are the best strategies for market agents to maximize their profits or other objectives?
- Which system attributes should concern the regulators or policy makers? This includes the norms and rules limiting the actions of the agents to ensure good performance of the system and to ensure meeting system-level objectives. How can they design those norms and rules and how can they study them ex-ante?
- How can one model the ESTS? Which are the interactions between the ESTS and the prosumer models? How to model the influence of the prosumers on the

ESTS environment? Conversely, how does the ESTS environment constrain the prosumers? Are the general ESTS laws fixed or are they part of the design?

- Which are the key-metrics to be used for assessing the performance of the overall system (energetic, environmental, economic, social…)?
- How to reproduce the agents' decision-making? What are the mechanisms behind their decision-making? How to consider individual and local specificities, such as personal constraints, objectives, preferences, reasoning, perception, etc.? At what level of aggregation should the agents be modeled? Is it at all possible or relevant to model the agents' decisions?
- Given a real system to model (like a real distribution system with practical or real data and GIS information as well as information of the citizens, the regulators, the marketers…), how to model agents' interactions according to the established coordination procedures? How to model the organization of agents, including the norms constraining the agents' actions and interactions? How to model the ESTS?

9.3.2 Modeling of the Prosumer Structure

One key issue is the modeling of the prosumers, including their structure and the interactions among themselves and with the environment.

Figure 9.2 shows a typical prosumer in the ESTS environment (other elements omitted), highlighting the internal blocks and variables of an individual prosumer.

We propose to characterize individual prosumers according to four elements: action set, communication, intelligence, and performance. The action set is determined by the technical options available to the prosumers and to their social status. It is composed of a finite dimension of action choices. Communication refers to the exchange of information with other agents or the environment; with each new set of data, prosumers can modify their actions for fulfilling their objective. Intelligence denotes processes of learning in the search for objectives (e.g. in efficiency or economic terms); zero intelligence prosumers behave as random or static action choosing agents. The actions taken by the prosumers feed back into ESTS, and the performance of the system can be evaluated.

Considering the same set of internal blocks for a population of prosumers, with different value added parameters representing the different properties of each one, the resulting actions of the prosumers will be thoroughly disparate. With an increasing number of prosumers, the cooperation and competition among them will interact with the other links deriving from their social lives. Thus a proper decision making for intelligent prosumers cannot be embodied by a simple optimization problem. This also applies to the other actors.
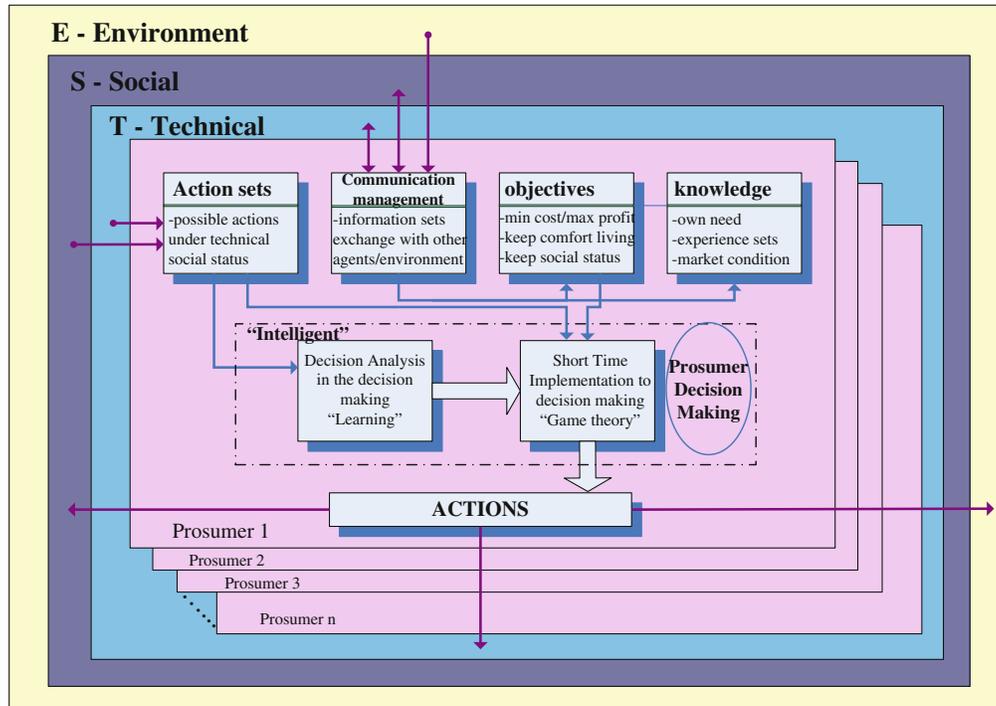
Fig. 9.2 Prosumers in ESTS

9.4 Explicative Example

In the following we describe a simulation example using MAS to implement multi-layer model of a smart grid.

This simulation is intended to capture the interactions between the prosumers, and the MV/LV distribution system operator and the regulator, taking into account the social characteristics of prosumers, the physical properties of the network, and the regulation methods of the operator. The goal is to link the social behavior to the impacts on the power network.

9.4.1 Modeling of the Prosumers

As already discussed in Sect. 9.3.2, the modeling of a large number of prosumers requires to represent the attributes of the prosumers and the relationships inside the prosumer society in a measurable way.

The social factors of the prosumer agents are typified in terms of psychological and economic attitudes regarding consumption and generation patterns. Two dimensional attitude spaces, respectively for consumption and generation, are designed as shown in Fig. 9.3. Prosumers are allocated in these coordinates, with different
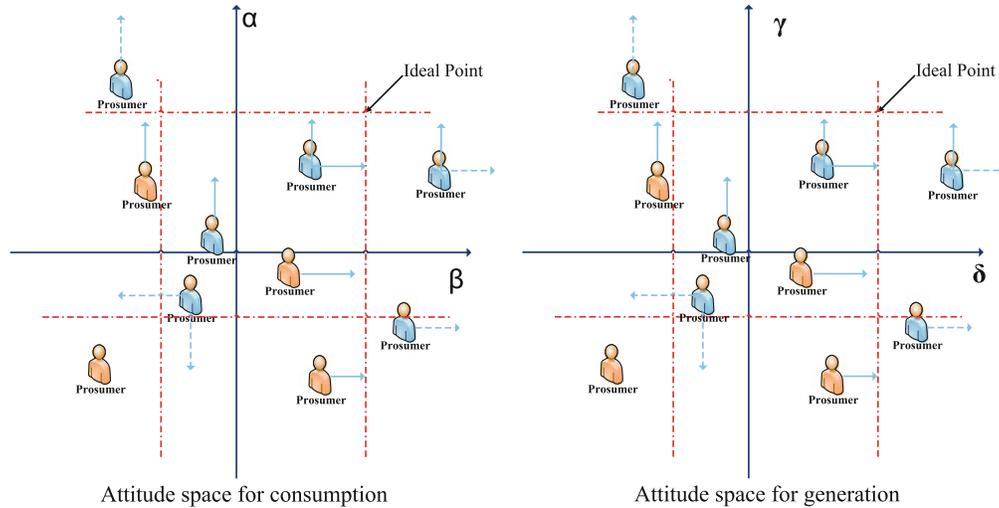
Fig. 9.3 Attitude spaces of prosumers

colors standing for different types of prosumers. According to their locations in the attitude spaces, decisions about generation and consumption at each time step are determined.

On the consumption side, considering the needs at each specific moment (appliances, devices, air conditioning, heating...), the power use is driven by two factors:

- $\alpha \rightarrow$ Comfort expectation (affecting the amount of available capacity the prosumer will effectively exploit)
- $\beta \rightarrow$ Demand price elasticity (responsiveness to price variation).

For generation, depending on the available sources (distributed generation sources, PV, Wind, FC, and so on) the power injection is driven by two factors:

- $\gamma \rightarrow$ Predisposition to technical management and attitude toward green energy (affects the amount of generation capacity effectively exploited)
- $\delta \rightarrow$ Generation price elasticity (responsiveness to price variation).

9.4.1.1 Prosumer Decision Making

The consumption capability and the generation capacity of the prosumers are determined as the basis for then computing the real time consumption and generation decisions.

The total capability of consumption is defined as ci, standing for the available level of comfort for prosumer I, and can be measured as the sum of the power of all electric appliances and devices installed (kW). ci is fixed in the short time but can change in the mid-term (i.e. the quantity and quality of the appliances and devices can change).

The total installed capacity is defined as e_i, standing for the available level of generation for prosumer I, and can be measured as the sum of the power of all generation devices installed (kW). e_i is also fixed in the short time but can change in the mid term (i.e. the generators can change).

Thus demand $d_i(t)$ and generation $g_i(t)$ by prosumer i in time t can be respectively expressed as:

$$d_i(t) = |\alpha_i(t)|^* c_i - |\beta_i(t)|^* \rho(t)$$

$$g_i(t) = |\gamma_i(t)|^* e_i - |\delta_i(t)|^* \upsilon(t)$$

where, $\alpha_i(t)$, $\beta_i(t)$, $\gamma_i(t)$, $\delta_i(t)$ are derived from the attitude spaces of prosumer i in time t; $\rho(t)$ and $\upsilon(t)$ are the prices for respectively withdrawing and injecting power at time t.

9.4.1.2 Social Dynamics and Attitude Update

The individual behavior of each prosumer can be affected by the behavior of others. This influence of one prosumer over another takes place the social network of relationships of each consumer.

In our simulation, two social networks (Neighborhood Circle and Social Circle) are defined for each prosumer. The Neighborhood Circle is "spatial" and is related to the living place of the prosumers. The Social Circle is "relational" and is related to social links of the prosumers.

At time step, prosumers will evaluate the rewards (in the form of comfort and cost savings) obtained by the other prosumers within their social networks. Those prosumers with best rewards from both neighborhood circle and social circles will be followed as role models, copying their choices.

When prosumers want to improve their rewards (comfort or savings), they will shift their behaviors towards that of their role models. In our simulation this adaptation process occurs by fixed values at each time step.

9.4.1.3 Quadrant Sensitivity

For studying the adaptation towards more comfort and less cost, we separate the attitude space in four quadrants. In Fig. 9.3, the preferable directions of adaptation are shown in blue arrows of each prosumer. Generally speaking, positive signs denote the interest of improving the savings or comfort; while negative signs indicate indifference. The sensitivities of the four quadrants are classified in Table 9.1.

As in the real world, if it were possible, all prosumers would like the maximum comfort together with the most cost efficient choices. But because the optimal behavior is unknown, and there is randomness in the behavior of prosumers and the system, there are boundaries for maximum price and comfort sensitivities (shown as red dashed lines in the attitude spaces in Fig. 9.3).

Table 9.1 Quadrant sensitivities

Quadrant	Signs (x, y)	Rules
First	$(+, +)$	Sensitive to both attitudes of x/y axis to improve through interactions
Second	$(-, +)$	Only care to improve the attitude of y axis
Third	$(-, -)$	Not influenced by others in their attitudes to consume and generate
Fourth	$(+, -)$	Only care to improve the attitude of x axis

9.4.1.4 Prosumer Intelligence

One can guess that most prosumers will value more reliable power consumptions and appropriate energy bills than steady earnings from their DGs. On the other hand, some large prosumers could offer reliable generation units, bidding in the local market or getting supply contracts.

9.4.2 Basic Case

To evaluate the impact of the autonomous prosumer behavior onto the physical grid, Fig. 9.4 shows two important layers of the multi-layer platform: the social layer and the physical layer.

The social layer in this example is composed of 1,000 prosumers. Their social status is marked with different colors, and the attitude values assigned to those status. The physical layer is a Medium and Low Voltage distribution system based on the benchmark of CIGRE task force C6.04.02, with LV (0.4 kV) and MV (20 kV) with bilateral power flow (reverse flow marked as red line). On each node, the consumption and generation values are marked as variable green arrows and orange arrows respectively. Thus consumption and generation decisions from prosumers at each time step determine the power flow in the physical layer.

As the simulation evolves, the prosumers apply their decision-making rules and adapt to the changing situation. As a result, they converge, as we can see in Fig. 9.5. Certain degree of social convergence driven by interactions among prosumer may initially help in improving the system performance; while an over convergence among the prosumers could damage the stability of the system.

Social convergence is not an unexpected outcome, similarly to what can be observed in retail markets. But power networks can be negatively affected by this extreme convergence of consumption or generation. In Fig. 9.5, almost all nodes are injecting power into the grid, as shown by the orange arrows. The opposite situation of all prosumers withdrawing power can also exist. Both situations are not what one would look for when developing Smart Grids with large penetration of distributed generations.

Social Layer

Physical Layer
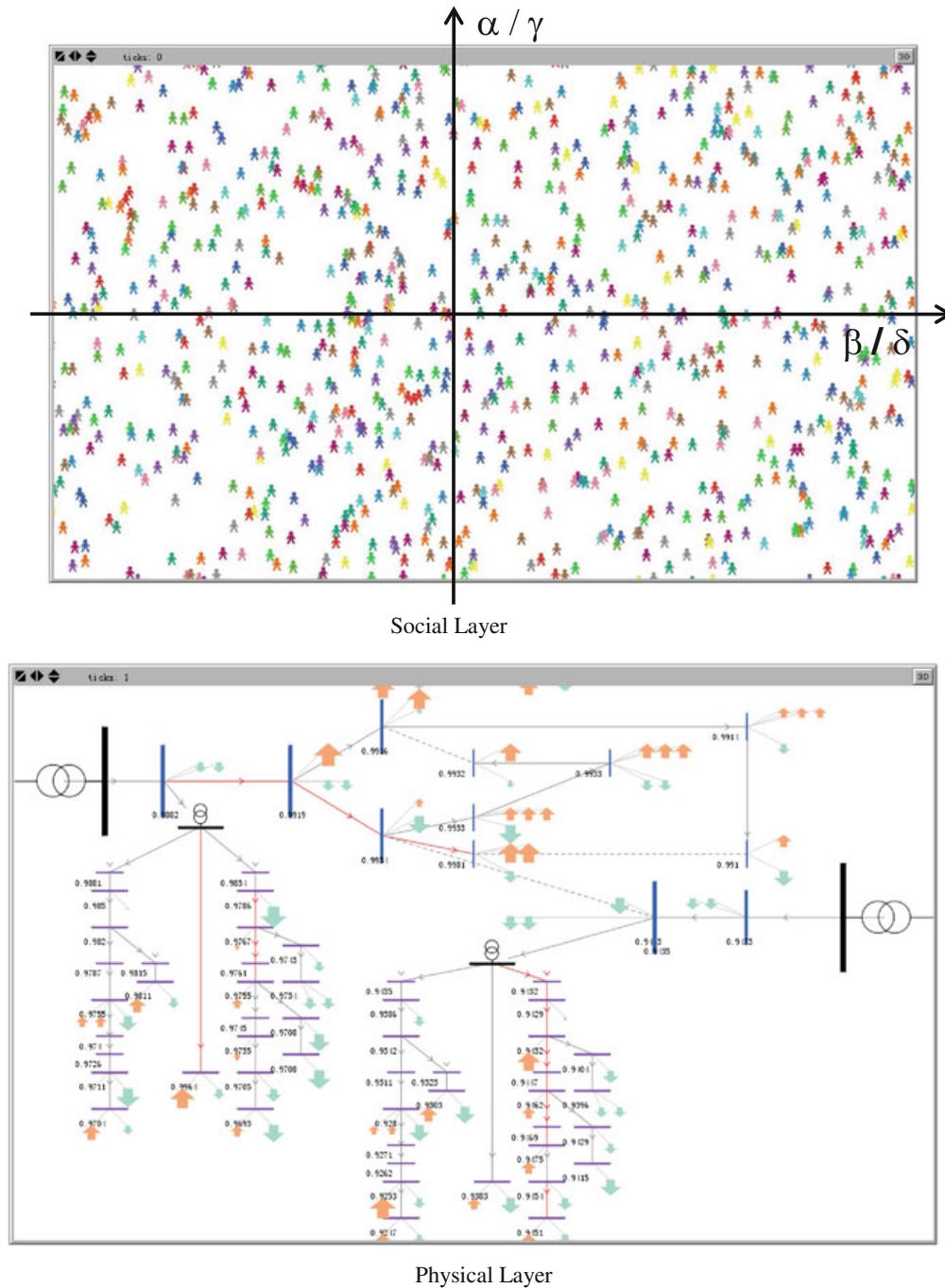
Fig. 9.4 Initial states of layers

Theferore, certain regulation rules should be introduced to guide the behaviors of prosumers for the sake of ensuring a healthy Smart Grid. Ideally the result will be an optimized system with satisfied prosumers.

Social Layer

Physical Layer

Fig. 9.5 Final states of layers

(a) *Without Regulation*

a.1   a.2   a.3

(b) *Unique Charging*

b.1   b.2   b.3
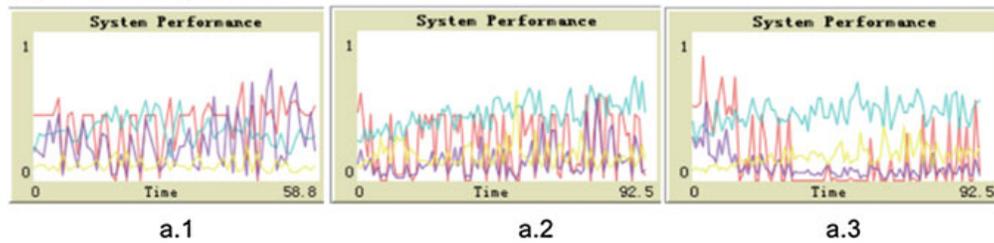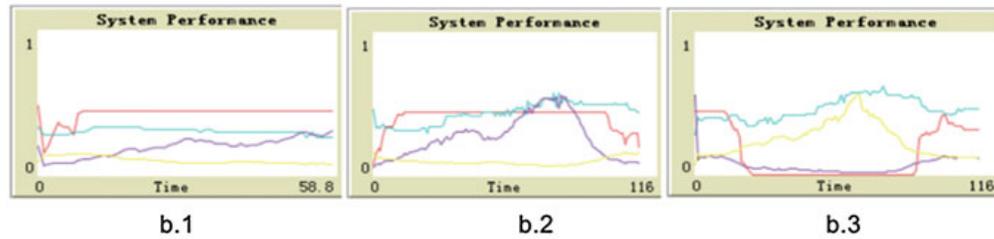
(c) *Location based Charging*

c.1   c.2   c.3
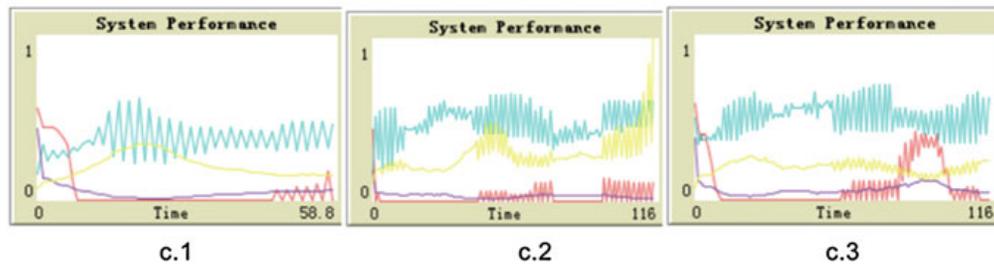
Fig. 9.6 Comparison of system performance

Outrange Voltage rate: preferred voltage range are set 0.9 to 1.1; this index measured the percentage of nodes with an voltage out of this range;

Power loss: losses in this MV/LV network are added and unified to compare;

DG rate: presents the ratio of power consumption that comes from local generatio;

Reverse line flow rate: this index measures the rate of lines with reverse line flows;

Time: marks the number of simulation run steps;

*.1 with fewer running steps show curves before first gathering of prosumers;

*.2 and *.3 with more running steps show curves under fullconvergence of prosumers.

9.4.3 Case Comparison

An effective guiding policy by the DSO that respects the freedom of the prosumers could be regulation based on alternative network charging. Here we consider two types of network charging. One policy will be unique network charging based on the global behavior of consumption and generation. The other will be location based charging, meaning that network charges are calculated based on the aggregated behavior of the prosumers connected to the same node; thus prices for consumption and generation are computed for every node in this distribution network.

Results of the simulations are shown in Fig. 9.6: though influenced by dynamic social behaviors and the convergence among prosumers, the system performance in both charging strategies is stabilized.

While delivering higher standards of system performance, the unique charging strategy fails to reduce outrange voltages and power loss with compared to the location based charging strategy, which minimizes both indices in all conditions.

Location based charging shows a characteristic efficient reaction through the convergence process of prosumers. In summary, based on these preliminary simulations, it appears that location based charging strategies can be more efficient in the management of prosumer behaviors.

9.5 Conclusion

The development of the Smart Grid paradigm will require analytic capabilities for evaluating the different aspects concerned, technical, economic, political and social. The multi-layer model introduced in this paper incorporates those various elements. The paper shows how Multi-Agent Simulation (MAS) applied to that multi-layer model can be used for studying Smart Grid, taking into consideration the complexities deriving from autonomous agents and their interactions.

The paper presents a formal approach to the representation and analysis of autonomous prosumer behaviors. Though examples, it was shown the use of simulations, with possible solutions to the problem at issue. Similarly, many other cases can be considered in this framework, with more complex interventions, regulatory settings and social structures.

Chapter 10
Network Physiology: Mapping Interactions Between Networks of Physiologic Networks

Plamen Ch. Ivanov and Ronny P. Bartsch

Abstract The human organism is an integrated network of interconnected and inter-acting organ systems, each representing a separate regulatory network. The behavior of one physiological system (network) may affect the dynamics of all other systems in the network of physiologic networks. Due to these interactions, failure of one sys-tem can trigger a cascade of failures throughout the entire network. We introduce a systematic method to identify a network of interactions between diverse physiologic organ systems, to quantify the hierarchical structure and dynamics of this network, and to track its evolution under different physiologic states. We find a robust rela-tion between network structure and physiologic states: every state is characterized by specific network topology, node connectivity and links strength. Further, we find that transitions from one physiologic state to another trigger a markedly fast reorganiza-tion in the network of physiologic interactions on time scales of just a few minutes, indicating high network flexibility in response to perturbations. This reorganization in network topology occurs simultaneously and globally in the entire network as well as at the level of individual physiological systems, while preserving a hierarchical order in the strength of network links. Our findings highlight the need of an inte-grated network approach to understand physiologic function, since the framework we develop provides new information which can not be obtained by studying indi-vidual systems. The proposed system-wide integrative approach may facilitate the development of a new field, Network Physiology.

P. Ch. Ivanov (✉)
Department of Physics and Center for Polymer Studies, Boston University, Boston, MA, USA
e-mail: plamen@buphy.bu.edu

P. Ch. Ivanov
Institute of Solid State Physics, Bulgarian Academy of Sciences, 1784 Sofia, Bulgaria

P. Ch. Ivanov · R. P. Bartsch
Harvard Medical School and Division of Sleep Medicine, Brigham and Women's Hospital, Boston, MA, USA
e-mail: bartsch.ronny@gmail.com

10.1 Introduction

In contrast to the unorthodox diagnostic approaches of the fictional character Dr. Gregory House from the acclaimed US TV-series "House" who, in a detective-like manner, considers a variety of interactions between multiple physiologic systems and variables to understand origins of symptoms in order to reach the right diagnosis, health care specialists traditionally focus on a single physiological system. Cardiologists mainly examine the heart and consider ECG signals; pulmonologists check lung structure and function and probe respiratory patterns; and brain neurologists study EEG. However, the human organism is an integrated network of interconnected and interacting physiologic organ systems, where each system is a multi-component structural and regulatory network. The complex behavior of one physiological system may be affected by changes in the dynamics of other systems in the physiologic network of organ networks. Due to these interactions, failure of one system may trigger a breakdown of the entire physiologic network.

Multiple organ failure is often the reason for fatal outcome in critical clinical care [1, 2]. In fact, multiple organ dysfunction remains a leading cause of death in most intensive care units. Clinical medicine offers support for specific organ systems that has proven necessary but often insufficient to promote recovery. If the links between physiological organ systems remain substantially altered, recovery is unlikely even when the structure and function of a specific failed system is restored after treatment. Indeed, autopsy findings in patients who succumb to multiple organ failure usually show that tissue architecture is preserved, cells do not appear abnormal and there is no widespread thrombosis. Nor does organ function appear to be irretrievably lost for patients who survived multiple organ failure [3]. This underscores the importance of identifying and quantifying the interactions between physiological organ systems, and how these interactions change under different physiologic states, pathologic conditions and with medical treatment. Further, medications developed to treat one physiological system often influence the function and have side effects on other systems. While some of the interactions between organ systems are partially known at the qualitative level, more precise quantitative estimates are important especially in the context of evaluating the proper medication dosage. Thus, the framework we propose to investigate a network of physiologic interactions between organ networks may help (i) to uncover new hitherto unknown links between organ systems, and (ii) to quantify the degree and strength of physiologic coupling and interactions, and how they change under various physiologic states and pathologic conditions.

A defining feature of physiological organ systems is their complexity. Decoding the remarkable range of behaviors of living systems in health and disease has emerged as a major focus of contemporary medicine. Physiological systems under neural regulation exhibit nonstationary, intermittent, scale-invariant and nonlinear behaviors [4, 5]. Moreover, physiologic dynamics transiently change in time with different physiologic states and under pathologic conditions [6–8], in response to changes in the underlying control mechanisms. The structural and neuronal control networks that underlie each physiologic organ system lead to the the high degree of

complexity in the output signals of physiological systems. This complexity is further compounded by various coupling [9] and feedback interactions [10–12] among different systems, the nature of which is not well-understood. Quantifying these physiologic interactions is a challenge as one system may exhibit multiple simultaneous interactions with other systems where the strength of the couplings may vary in time.

Therefore, to understand physiologic function it is critical to identify the network of physiologic interactions, and to track its evolution under different physiologic states and pathological conditions. This enterprise requires collaboration among scientists with different backgrounds, and the need to foster multidisciplinary approaches to problems at the interface of physics and physiology. Modern methods of statistical physics and recent advances in the theory of complex networks have great potential to uncover and quantify the structural and dynamical characteristics of the physiologic network of organ networks. Here, we introduce a method to identify interactions between physiologic systems, and we propose an integrative approach to study the dynamical evolution of an entire network of physiologic interactions in relation to changes under different physiologic states.

The central task of statistical physics is to understand macroscopic phenomena that result from microscopic interactions among many individual components often driven by competing forces and nonlinear feedback mechanisms. This problem is akin to many investigations undertaken in physiology. In particular, physiological systems under neural regulation and their complex nonlinear interactions among each other are good candidates for a statistical physics approach, since (i) physiological systems include many individual components (nodes) connected through a network of nonlinear feedback interactions, as observed in certain physical systems, and (ii) each physiologic system has multiple simultaneous interactions with other systems, thus forming a network of physiologic networks.

Complex networks have attracted enormous attention in the past decade in various fields of application. However, despite the importance to physiology and medicine, the network of interactions between diverse vertically- and horizontally-integrated organ systems is not known. Dynamical networks of physiologic interactions are particularly challenging because most physiological systems are multiple component complex systems with their own regulatory mechanism, and their function is affected by various interactions with other systems and by their integration in the human organism. Furthermore, physiologic dynamics and interactions continuously change in time due to changes in physiologic conditions. Thus, most of the complexities encountered in many of the networks studied so far are simultaneously present in physiological networks.

The interdisciplinary field of Network Physiology bridges two active fields of modern science: (A) the physics of complex networks, and (B) the organization and control of integrated physiologic organ systems. There are several fundamental questions that are critical for the development of both fields:

(A) In the field of complex networks: (A.1) it remains an unsolved problem how to identify and quantify networks comprised of *diverse* systems with very different types of interactions; (A.2) the relation between network topology and function is

hypothesized but has not been demonstrated yet on real systems; (A.3) there are no studies on real networks evolving in time and undergoing topological phase transitions from one state to another, and (A.4) the relation of network topology to network robustness and to the propagation of cascades of failure. These questions are even more challenging for networks of networks, where each subnetwork is characterized by different topology and dynamics of interactions with other subnetworks.

(B) In the field of integrated physiology: (B.1) it is not known how different physiologic organ systems simultaneously interact as a network in the human body; (B.2) whether different physiologic states are characterized by distinct networks of physiologic interactions; (B.3) how transitions across physiologic and pathologic states lead to transitions in the strength of physiologic interactions and in physiologic network topology, and (B.4) quantitative knowledge of the critical zone of physiologic coupling between multiple organ systems is essential to predict disintegration of the physiologic network leading to multiple organ failure and other pathologies.

10.2 Complex Networks Approach to Physiologic Interactions

Research in statistical physics of networks has identified networks with complex topologies [13, 14], and has focused on the role of topology for network function and robustness [15–17], on the evolution of network topology under varied conditions [18], emergence of self-organization and complex network behavior out of simple interactions [19], and more recently on critical transitions due to failure in the coupling of interdependent networks [20]. Recent advances in complex networks theory are of relevance to a broad range of real systems including industrial [20, 21], transportation [22, 23] and communication networks [24], food and ecological webs [19], financial systems and social interactions [21, 25–29] as well as biological systems at the microscopic level such as genetic and protein-interaction networks [30], biochemical [31], metabolic [32] and cell signaling networks [33]. However, understanding the relation between topology and dynamics of complex networks remains a challenge, especially when (i) the network evolves with links created or lost in time, (ii) links between different nodes have different functional form and strength/weight which change over time and (iii) network nodes are of different kind with different dynamical properties and types of links [34, 35]. A further challenge to the contemporary theoretical framework of complex networks is posed by real-world systems where each network node represents a multicomponent complex system, a network on its own, with its own topology and regulatory mechanism that can vary in time, and where the transient output dynamics of individual networks affect the entire "network of networks" by reinforcing (or weakening) the coupling between individual networks and changing network topology. A prime example of a network of networks is the human organism, where integrated physiologic systems, each representing a complex network, form a network of interactions that in turn affect physiologic function of individual systems or of the entire organism, and

where breakdown in the interaction between physiological systems under certain conditions may lead to a cascade of system failures [1, 2].

Physiological systems exhibit remarkable dynamic complexity where transient changes in the underlying control mechanisms associated with different physiologic states and conditions lead both to changes in their individual output characteristics as well as in their interactions [6, 36–46]. Here, we introduce a framework to study the network of interactions between physiological systems, and we focus on the topology and dynamics of this network and their relevance to physiologic function. We hypothesize that during a given physiologic state the physiologic network of organ networks may be characterized by a specific topology. Further, we hypothesize that even for networks with relatively stable topology associated with specific physiologic states, the coupling strength between physiologic systems may change in time due to the inherent variability in the regulation and output of these systems. Moreover, coupling strength and physiologic network topology may change with transition from one physiologic state to another, where physiologic interactions (network links) are established or lost leading to a completely new network configuration. Such transitions may also be associated with changes in the connectivity of specific network nodes, i.e., the number of systems to which a given physiologic system is connected can change, forming sub-networks of physiologic interactions. Thus, probing physiologic network connectivity and the stability of physiologic coupling may provide new insights on integrated physiologic function.

10.3 Time Delay Stability and Network of Physiologic Interactions

We introduce the concept of time delay stability (TDS) to identify and quantify dynamic links among physiological systems. The framework we propose allows (i) to quantify the topology and global dynamics of physiologic networks, taking into account the output of individual physiologic systems as well as the interactions among them, and (ii) to track the dynamical evolution of multiple interconnected systems undergoing transitions from one physiologic state to another (Fig. 10.1). We construct a network of interactions for an ensemble of key integrated physiologic systems (cerebral, cardiac, respiratory, ocular and muscle activity). We consider different sleep stages (deep, light, REM sleep and quiet wake) as examples of physiologic states. We demonstrate that sleep stages are associated with markedly different networks of physiologic interactions (Fig. 10.2) characterized by different number and strength of links (Fig. 10.3), and by specific node connectivity (Fig. 10.6). In particular, during deep sleep we find a much lower number of links in the physiologic network compared to light sleep (Figs. 10.2 and 10.3)—individual physiologic systems, such as the cardiac, are highly connected to other systems during light sleep while there are practically no TDS links during deep sleep (Fig. 10.6). Furthermore, the network links are much weaker during deep compared to light sleep (Figs. 10.3d
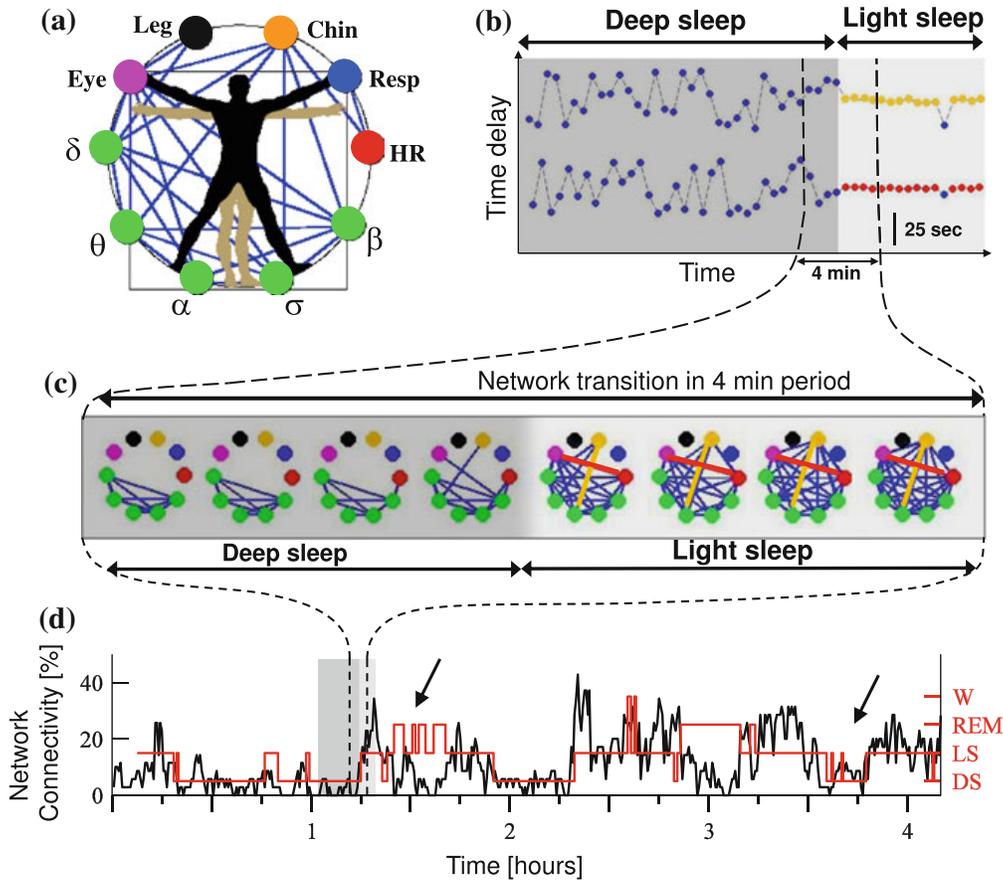
Fig. 10.1 Transitions in the network of physiologic interactions. **a** Dynamical network of interactions between physiological systems where ten network nodes represent six physiologic systems—brain activity (EEG waves: δ, θ, α, σ, β), cardiac (HR), respiratory (Resp), chin muscle tone, leg and eye movements. **b** Transition in the interactions between physiological systems across sleep stages. The time delay between two pairs of signals, (*top*) α-brain waves and chin muscle tone, and (*bottom*) HR and eye movement, quantifies their physiologic interaction: highly irregular behavior (*blue dots*) during deep sleep is followed by a period of time delay stability during light sleep indicating a stable physiologic interaction (*red dots* for the HR-eye and *orange dots* for the α-chin interaction). **c** Transitions between physiologic states are associated with changes in network topology: snapshots over 30 s windows during a transition from deep sleep (*dark gray*) to light sleep (*light gray*). During deep sleep the network consists mainly of brain-brain links. With transition to light sleep links between other physiologic systems (network nodes) emerge and the network becomes highly connected. The stable α-chin and HR-eye interactions during light sleep in (**b**) are shown by an *orange* and a *red* network link respectively. **d** Physiologic network connectivity for one subject during night sleep calculated in 30 s windows as the fraction (%) of present links out of all possible links. *Red line* marks sleep stages as independently scored in a sleep lab. Low connectivity is consistently observed during deep sleep (0:30–1:15 and 1:50–2:20 h) and REM sleep (1:30–1:45 and 2:50–3:10 h), while transitions to light sleep and wake are associated with a significant increase in connectivity [47]
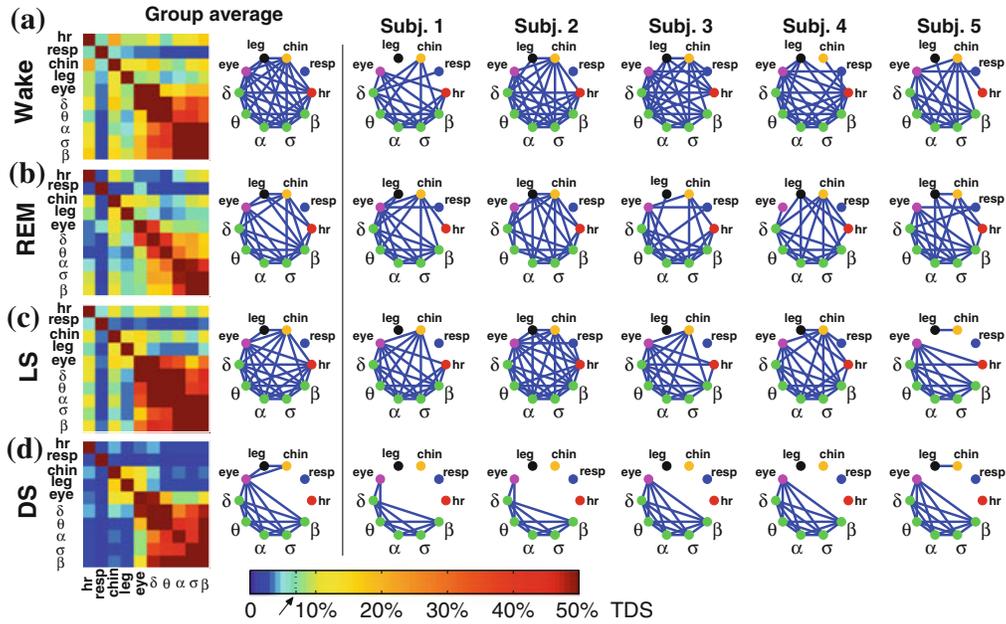
Fig. 10.2 Network connectivity across sleep stages. Group-averaged time delay stability (*TDS*) matrices and related networks of physiologic interactions during different sleep stages. Matrix elements are obtained by quantifying the TDS for each pair of physiologic systems after obtaining the weighted average of all subjects in the group. *Color code* represents the average strength of interaction between systems quantified as the fraction of time (out of the total duration of a given sleep-stage throughout the night) when TDS is observed. The physiologic network exhibits transitions across sleep stages—lowest number of links during deep sleep, higher during REM and highest during light sleep and quiet wake—a behavior observed in the group-averaged network as well as for each subject. Network topology also changes with sleep-stage transitions: from predominantly brain-brain links during deep sleep to a high number of brain-periphery and periphery-periphery links during light sleep and wake

and 10.5a). Traditionally, differences between sleep stages are attributed to modulation in the sympatho-vagal balance with dominant sympathetic tone during wake and REM [48]: spectral, scale-invariant and nonlinear characteristics of the dynamics of individual physiologic systems indicate higher degree of temporal correlations and nonlinearity during wake and REM compared to NREM (light and deep sleep) where physiologic dynamics during exhibit weaker correlations and loss of nonlinearity [6, 45]. In contrast, the network of physiologic interactions shows a completely different picture: the network characteristics during light sleep are much closer to those during wake and very different from deep sleep (Figs. 10.2 and 10.3). Specifically, network connectivity and overall strength of physiologic interactions are significantly higher during wake and light sleep, intermediate during REM and much lower during deep sleep. Thus, our empirical observations indicate that while sleep-stage related modulation in sympatho-vagal balance plays a key role in regulating individual physiologic systems, it does not fully account for the physiologic network topology and dynamics across sleep stages, showing that the proposed framework captures principally new information.

Entire Network

Brain-Periphery & Periphery-Periphery Subnetwork

Brain-Brain Subnetwork
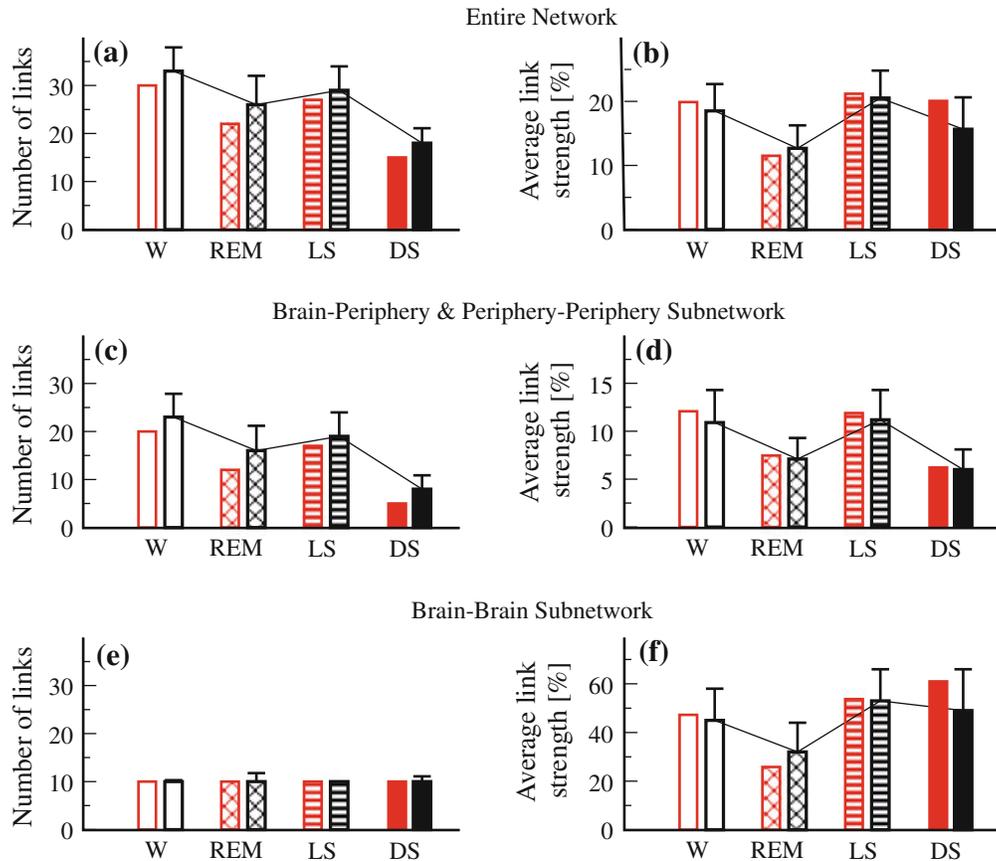
Fig. 10.3 Sleep-stage stratification pattern in network connectivity and network link strength. Group-averaged number of links (**a**) and averaged link strength (**b**) are significantly higher during wake and light sleep compared to REM and deep sleep. There is no significant difference between wake and light sleep. This pattern is even more pronounced for the subnetwork formed by the brain-periphery and periphery-periphery links shown in (**c**) and (**d**). In contrast, the number of brain-brain links remains practically unchanged with sleep-stage transitions (**e**), and the average brain-brain link is ≈ 5 times stronger in all sleep stages compared to the other network links (**f**). The group-averaged patterns in the number of network links and in the average link strength across sleep stages (*black bars*) are consistent with the behavior observed for individual subjects (*red bars* in all panels represent the same subject). The average link strength represents the average strength of all links in a network obtained from a given subject during a specific sleep stage which then is averaged over all subjects. Error bars indicate standard deviation

To quantify the interaction between physiologic systems and to probe how this interaction changes in time under different physiologic conditions we study the time delay with which modulations in the output dynamics of a given physiologic system are consistently followed by corresponding modulations in the signal output of another system. Periods of time with approximately constant time delay indicate a stable physiologic interaction, and stronger coupling between physiologic systems results in longer periods of time delay stability (TDS). The TDS method is general, and can be applied to diverse systems. It is more reliable in identifying physio-

logic coupling compared to traditional cross-correlation and cross-coherence analyses (Fig. 10.7) which are not suitable for heterogeneous and nonstationary signals, and are affected by the degree of auto-correlations in these signals [49]. Utilizing the TDS method we build a dynamical network of physiologic interactions, where network links between physiological systems (considered as network nodes) are established when the time delay stability representing the coupling of these systems exceeds a significance threshold level, and where the strength of the links is proportional to the percentage of time when time delay stability is observed. This dynamic network approach provides an integrated view of the simultaneous interactions of multiple physiologic systems, where transient changes in physiologic conditions of the human organism are reflected in continuous fluctuations in the strength of network links, variations in the connectivity of individual network nodes, and emergence or loss of specific links in response to changes in physiologic function—all leading to transitions in network topology.

10.4 Transitions in Network Topology with Physiologic Function

We apply this new approach to a group of young subjects with continuously recorded multi-channel physiologic data during sleep which allows us to track the dynamics and evolution of the network of physiologic interactions during different sleep stages and sleep-stage transitions (Fig. 10.1). We focus on physiologic dynamics during sleep since sleep stages are well-defined physiological states, and external influences due to physical activity or sensory inputs are reduced during sleep. While earlier studies have identified how sleep regulation influences aspects of the specific control mechanism of individual physiologic systems (e.g., cardiac or respiratory [6, 7, 45, 48]), the dynamics and topology of an entire physiologic network have not been studied so far. Utilizing sleep data as an example we demonstrate that a network approach to physiologic interactions is necessary to understand how modulations in the regulatory mechanism of individual systems translate into reorganization of physiologic interactions across the human organism.

   We find that the network of interactions between physiologic systems is very sensitive to sleep-stage transitions. In a short time window of just a few minutes the network topology can dramatically change—from only a few links to a multitude of links (Fig. 10.1c)—indicating transitions in the global interconnectivity between physiological systems. These network transitions are not associated with random occurrence or loss of links but are characterized by certain organization in network topology where given links between physiological systems remain stable during the transition while others do not—e.g., brain-brain links persist during the transition from deep to light sleep while brain-periphery links significantly change (Fig. 10.1c). Further, we find that sleep-stage transitions are paralleled by abrupt jumps in the total number of links leading to higher or lower network connectivity (Fig. 10.1c, d). However, even during stable physiologic conditions within a specific sleep stage, the network of physiologic interactions does not remain static and undergoes continuous dynamical changes with small fluctuations in the number of network links. These
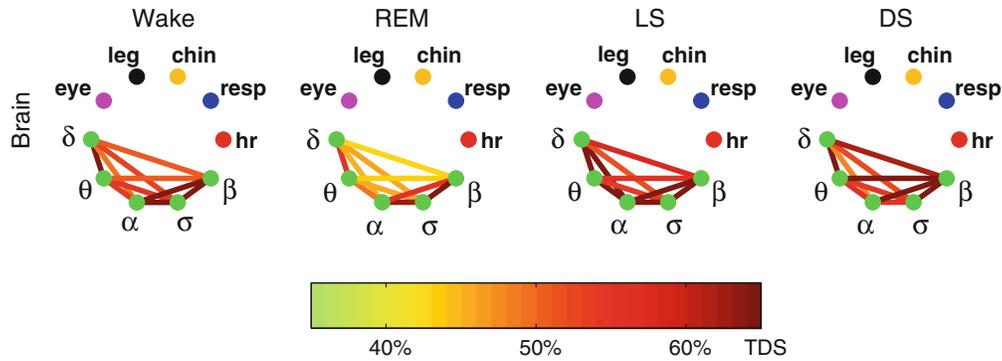
Fig. 10.4 Network connectivity and link strength of the brain-brain subnetwork for different sleep stages. While the topology of the brain subnetwork does not change, the strength of network links significantly changes with strongest links during light sleep and deep sleep (*brown* and *dark red color*), intermediate during wake (*red* and *orange color*) and weakest links during REM sleep (*yellow color*)

network dynamics are observed for each subject in the database, where consecutive episodes of sleep stages (scored from standard polysomnographic recordings) are paralleled by a level of connectivity specific for each sleep stage, and where sleep-stage transitions are consistently followed by transitions in network connectivity throughout the course of the night. Indeed, the network of physiologic interactions exhibits a remarkable responsiveness as network connectivity changes even for short sleep-stage episodes (Fig. 10.1d).

To identify the characteristic network topology for each sleep stage we obtain group-averaged time delay stability matrices, where each matrix element represents the percentage of time with stable time delay between two physiological systems, estimated over all episodes of a given sleep stage throughout the night. Matrix elements with values above a threshold of statistical significance determined by surrogate analysis, indicate stable interactions between physiologic systems represented by network links (Fig. 10.2). We find that matrix elements greatly vary for different sleep stages with much higher values for wake and light sleep, lower values for REM and lowest for deep sleep. This is correspondingly reflected in higher network connectivity for wake and light sleep, lower connectivity for REM and significantly reduced number of links during deep sleep (Fig. 10.3a). Further, the time delay stability matrices indicate separate subgroups of interactions between physiologic systems—brain-periphery, periphery-periphery and brain-brain interactions—which are affected differently during sleep stages and form different sub-networks. Specifically, matrix elements representing interactions between peripheral systems (cardiac, respiratory, chin, eye, leg) and the brain as well as interactions among the peripheral systems are very sensitive to sleep-stage transitions, leading to networks of very different topology for different sleep stages (Fig. 10.2). We find sub-networks with high number of brain-periphery and periphery-periphery links during wake and light sleep, lower number of links during REM and a significant reduction of links at

deep sleep (Fig. 10.3c). In contrast, matrix elements representing brain-brain inter-
actions form a subnetwork with the same number of brain-brain links (Fig. 10.3e),
and stable topology consistently present in the physiologic network during all sleep
stages (Fig. 10.2). These sleep-stage related transitions in network connectivity and
topology are not only present in the group-averaged data but also in the physiologic
networks of individual subjects, suggesting universal behavior (Fig. 10.2). Notably,
we find a higher number of brain-periphery links during REM compared to deep
sleep despite inhibition of motoneurons in the brain leading to muscle atonia during
REM [50]. Further, the empirical observations of significant difference in network
connectivity and topology during light sleep compared to deep sleep are surprising,
given the similarity in the output dynamics of physiologic systems during light and
deep sleep [6, 7, 45, 48] (both stages traditionally classified as NREM), and indi-
cate that previously unrecognized aspects of sleep regulation may be involved in the
control of physiologic network interactions.

10.4.1 Physiologic States and Network Link Strength Stratification

Networks with identical connectivity and topology can exhibit very different strength
of their links. We find that not only network connectivity but also the average strength
of network links changes with sleep-stage transitions: network links are significantly
stronger during wake and light sleep compared to REM and deep sleep—a pattern
similar to the behavior of the network connectivity across sleep stages (Fig. 10.3a, b).
Further, subgroups of physiologic interactions exhibit different relationship between
their respective subnetwork connectivity and the average link strength. Specifically,
the subnetwork of brain-periphery and periphery-periphery interactions is charac-
terized by significantly stronger links (and also higher connectivity) during wake
and light sleep, and much weaker links (with lower network connectivity) during
deep sleep and REM (Fig. 10.3c, d). In contrast, the subnetwork of brain-brain inter-
actions exhibits very different patterns for the connectivity and the average link
strength—while the group average subnetwork connectivity remains constant across
sleep stages, the average link strength varies with highest values during light and deep
sleep and a dramatic ≈40 % decline during REM. The observation of significantly
stronger links in the brain-brain subnetwork during NREM compared to REM sleep
is consistent with the characteristic of NREM as EEG-synchronized sleep and REM
as EEG-desynchronized sleep [50]. During NREM sleep adjacent cortical neurons
fire synchronously with a relatively low frequency rhythm [51] leading to coherence
between frequency bands in the EEG signal, and thus to stable time delays and strong
network links (Fig. 10.3f and 10.4). In contrast, during REM sleep cortical neurons
are highly active but fire asynchronously [51] resulting in weaker links (Fig. 10.3f
and 10.4). Our findings of stronger links in the brain-brain subnetwork during NREM
indicate that bursts in the spectral power of one EEG-frequency band are consistently
synchronized in time with bursts in a different EEG-frequency band, thus leading to
periods of longer time delay stability. This can explain some seemingly surprising

network links—for example, we find a strong link between α and δ brain activity during NREM sleep (Fig. 10.2) although α waves are greatly diminished and δ waves are dominant [50]. Since the spectral densities of both waves are normalized before the TDS analysis, the presence of a stable α–δ link indicates that a relative increase in the spectral density in one wave is followed with a stable time delay by a corresponding increase in the density of the other wave—an intriguing physiologic interaction which persists not only during deep sleep but is also present in light sleep, REM and quiet wake (Fig. 10.2). Notably, the average link strength of the brain-brain subnetwork is by a factor of 5 higher compared to all other links in the physiologic network (Fig. 10.3d, f).

Our finding that after averaging over all links in the physiologic network the resulting average link strength exhibits a specific stratification pattern across sleep stages, with strongest links during light sleep and wake, and weaker links during deep sleep and REM (Fig. 10.3), raises the question whether the underlying distribution of the network links strength is also sleep-stage dependent. To this end and to probe the relative strength of individual links we obtain the rank distribution of the strength of the brain-periphery and periphery-periphery network links for each sleep stage averaged over all subjects in the group (Fig. 10.5a). The link strength shown in the rank plots in Fig. 10.5a is determined by the degree of time delay stability, quantified as the fraction of time when TDS is observed. We find that the rank distribution corresponding to deep sleep is vertically shifted to much lower values for the strength of the network links, while the rank distribution for light sleep and wake is for all links consistently higher than the distribution for REM. Thus, the sleep-stage stratification pattern we find for the average strength of the network links (Fig. 10.3d) originates from the systematic change in the strength of individual network links with sleep-stage transitions as demonstrated by the rank analysis. Notably, although the strength of individual network links changes significantly with sleep stages, the rank order of the links does not significantly change. Remarkably, after rescaling the rank distributions for all sleep stages, we find that they collapse to two distinct functional forms: (i) a slow and smoothly decaying rank distribution for REM and wake, and (ii) a much faster decaying rank distribution for deep sleep and light sleep with a characteristic plateau in the mid rank range indicating a cluster of links with similar strength (Fig. 10.5b). Despite the similarity in the functional form of the distributions and in the rank order in the strength of individual network links, our analyses show a significant difference in the average strength of network links during deep sleep compared to light sleep and REM compared to wake (Fig. 10.3d).

10.4.2 Local Topology and Connectivity of the Physiologic Network

Our observations that physiologic networks undergo dynamic transitions where key global properties such as network connectivity and average link strength significantly change with sleep-stage transitions following a robust stratification pattern, raise the question whether local topology and connectivity of individual network nodes also
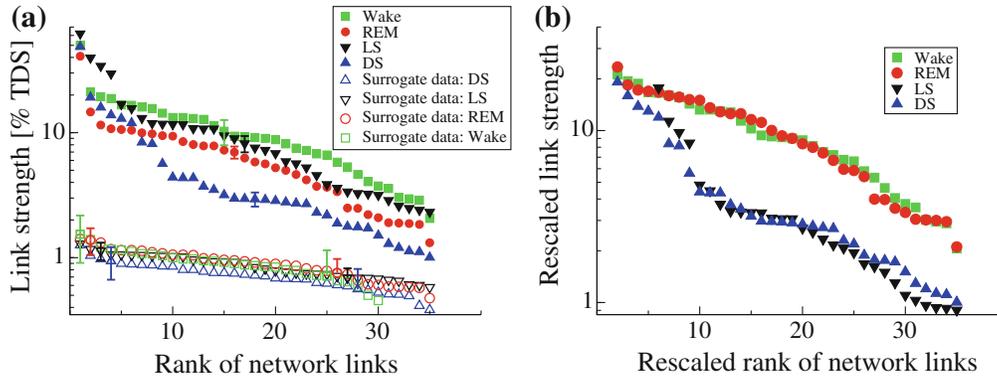
Fig. 10.5 Rank distributions of the strength of network links. Group-averaged strength of individual physiologic network links for different sleep stages. Rank 1 corresponds to the strongest link in the network, i.e., highest degree of time delay stability (*TDS*) (shown are all periphery-periphery and brain-periphery links). **a** The rank distributions for different sleep stages are characterized by different strength of the network links—consistently lower values for most links during deep sleep, higher values during REM and highest during light sleep and wake, indicating that the stratification pattern in Fig. 10.3d is present not only for the average link strength (when averaging over different types of links in the network) but also for the strength of individual links. Indeed, links from all ranks are consistently stronger in light sleep compared to deep sleep and REM: such rank-by-rank comparison of links across sleep stages is possible because the rank order of the links does not change significantly from one sleep stage to another. A surrogate test based on TDS analysis of signals paired from different subjects, which eliminates endogenous physiologic coupling, leads to significantly reduced link strength ($p < 10^{-3}$) and close to uniform rank distributions with no difference between sleep stages (*open symbols*), indicating that the TDS method uncovers physiologically-relevant information. Error bars indicate standard error. **b** Rescaling the plots reveals two distinct forms of rank distributions: a slow decaying distribution for wake and REM, and a fast decaying distribution for light sleep and deep sleep with a pronounced plateau in the middle rank range corresponding to a cluster of links with similar strength, most of which related to the cardiac system

change during these transitions. Considering each physiologic system (network node) separately, we examine the number and strength of all links connecting the system with the rest of the network. For example, we find that the cardiac system is highly connected to other physiologic systems in the network during wake and light sleep (Fig. 10.6). In contrast, during deep sleep we do not find statistically significant time delay stability in the interactions of the cardiac system, which is reflected by absence of cardiac links (Fig. 10.6). Further, we find that the average strength of the links connected to the cardiac system also changes with sleep stages: stronger interactions (high % TDS) during wake and light sleep and significantly weaker interactions below the significance threshold during deep sleep (Fig. 10.6). Such 'isolation' of the cardiac node from the rest of the network indicates a more autonomous cardiac function during deep sleep—also supported by earlier observations of breakdown of long-range correlations and close to random and more linear behavior in heartbeat intervals in this sleep stage [6]. With transition to light sleep, REM and wake where the average link strength and connectivity of the cardiac system is significantly higher, indicating increased interactions with the rest of the network that lead to cor-
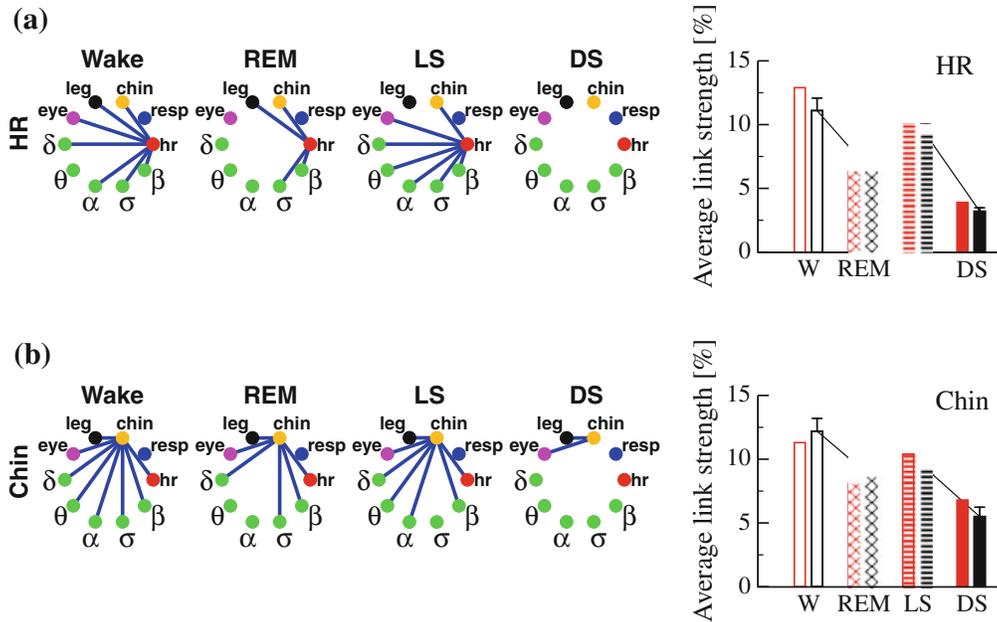
Fig. 10.6 Transitions in connectivity and link strength of individual network nodes across sleep stages. The number of links to specific physiologic systems (network nodes) significantly changes, with practically no links during deep sleep, a few links during REM and much higher connectivity during light sleep and wake. Notably, the average strength of the links connecting a given network node is also lowest during deep sleep and highest during light sleep and wake. Shown are connectivity and average link strength for two network nodes, heart and chin. This sleep-stage stratification pattern in individual physiologic system (node) connectivity and in the average strength of the links connecting a specific network node is consistent with the transitions of the entire network across sleep stages shown in Fig. 10.3. Networks for heart and chin are obtained by averaging the corresponding networks for all subjects. During deep sleep no links to the heart are shown since the strength of each link averaged over all subjects is below the significance threshold. *Right bars* in the panels represent for different sleep stages the group mean of the average strength of network links connecting heart and chin respectively, and error bars show the standard deviation. *Left bars* represent an individual subject

respondingly higher degree of correlations and nonlinearity in cardiac dynamics [6]. Similarly, respiratory dynamics also exhibit high degree of correlations during REM and wake, lower during light sleep and close to random behavior during deep sleep [45]. Such transitions in the number and strength of links across sleep stages we also find for other network nodes (for example chin, Fig. 10.6). Moreover, the sleep-stage stratification pattern in connectivity and average link strength for individual network nodes (Fig. 10.6) is consistent with the pattern we observe for the entire network (Fig. 10.3). Our findings of significant reduction in the number and strength of brain-periphery and periphery-periphery links in the corresponding sub-networks during deep sleep indicate that breakdown of cortical interactions, previously reported during deep sleep [52], may also extend to other physiologic systems under neural regulation. Indeed, the low connectivity in the physiologic network we find in deep sleep may explain why people awakened during deep sleep do not adjust immediately
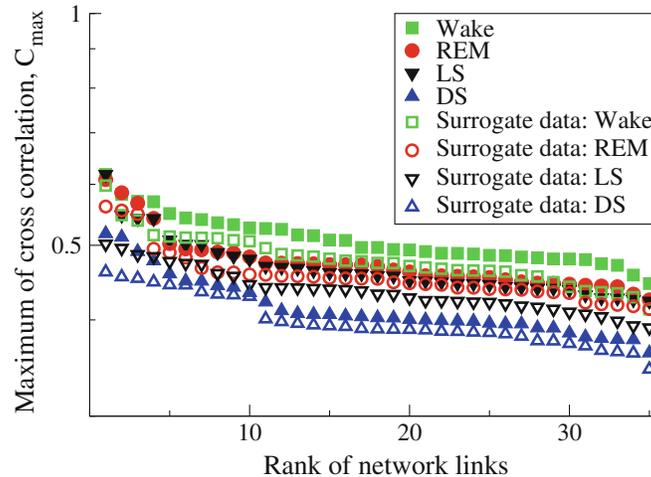
Fig. 10.7 Cross-correlation and surrogate analysis. Rank plots obtained from cross-correlation analysis show no statistically significant differences between real and surrogate data, indicating that cross-correlation is not a reliable measure to identify physiologic interactions

and often feel groggy and disoriented for a few minutes. This effect is not observed if subjects are awakened from light sleep when we find the physiologic network to be highly connected (Fig. 10.2). Further, the fact that deep sleep in primates dominates at the beginning of the night and not close to dawn, when many large predators preferably hunt, may have been evolutionarily advantageous.

Introducing a framework based on the concept of TDS we identify a robust network of interactions between physiologic systems, which remains stable across subjects during a given physiologic state. Further, changes in the physiologic state lead to complex network transitions associated with a remarkably structured reorganization of network connectivity and topology that simultaneously occurs in the entire network as well as at the level of individual network nodes, while preserving the hierarchical order in the strength of individual network links. Such network transitions lead to the formation of sub-networks of physiologic interactions with different topology and dynamical characteristics. In the context of sleep stages, network transitions are characterized by a specific stratification pattern where network connectivity and link strength are significantly higher during light compared to deep sleep and during wake compared to REM. This can not be explained by the dynamical characteristics of the output signals from individual physiologic systems which are similar during light and deep sleep as well as during wake and REM. The observed stability in network topology and rank order of links strength during sleep stages, and the transitions in network organization across sleep stages provide new insight into the role which individual physiologic systems as well as their interactions play during specific physiologic states. We note that traditional methods based on cross-correlation or cross-coherence analysis lead to spurious detection of interrelations and coupling in signals of different origin and with different autocorrelations, and fail to identify and quantify a the network of physiologic interactions (Fig. 10.7). While we

demonstrate one specific application, the framework we develop can be applied to a broad range of complex systems where the TDS method can serve as a tool to characterize and understand the dynamics and function of real-world heterogeneous and interdependent networks.

10.5 Summary

We introduce a new framework to investigate a network of interactions between complex physiological systems, each representing a separate regulatory network. This proves useful to uncover key aspects of physiologic dynamics and coupling in the context of the integrated function of diverse physiological systems in the human organism, and may facilitate novel theoretical approaches to study dynamical processes on networks of networks. These investigations constitute a first step in the development of a new field we call Network Physiology.

Specifically:

1. This is the first study of a network comprised of diverse complex systems.
   Earlier studies have focused on networks where (i) all nodes are of the same type, and (ii) network links are static and do not change in time. This is not the case in many real networks. Further, such "idealized" networks can not exhibit transitions in topology, and thus do not allow investigation of key questions such as the relation between network topology and function.
   Quantifying networks comprised of different types of nodes, where the nodes are not identical and simple units, but represent complex multi-component dynamical systems with their own regulatory mechanisms, is a major challenge which has not been addressed so far. The reason that network interactions between such complex systems have not been studied is that different types of systems have output signals with very different characteristics, which can also change in time. Thus, current methods tailored to probe the interaction/coupling between two similar systems do not work for a pair of different systems. This is a strong limitation when studying real-world networks.
   To overcome this limitation we developed a framework, based on a novel concept of time delay stability (TDS), to probe interactions among diverse systems by quantifying interrelations between their transient signal outputs. Utilizing this new approach we identify a dynamic network that represents the global behavior of a group of complex systems (networks) even when the links between these systems are not a-priori known. Our approach is general, and can be applied to many real dynamical systems and networks.
2. We present the first physiologic network. Specifically, we identify and quantify a network of interactions between key integrated physiologic systems: cerebral, cardiac, respiratory, ocular and motor system. These are diverse and complex systems, with their own regulatory neuronal networks, and with very different types of output signals.

This discovery provides a first dynamical map of the human organism as an integrated network of interacting physiological systems. Utilizing the physiologic network we can track how the behavior of one organ system can be affected by changes in the dynamics of other systems. Further, this approach allows to estimate whether, under certain conditions, failure of one system may trigger a breakdown of the entire network of physiologic systems. This network information is critical to understand physiologic function and uncovers new aspects of the mechanisms of physiologic regulation, and cannot be obtained by traditional studies focusing on individual systems.

The new physiologic information we obtain is relevant and may be utilized for clinical applications in critical care units, in situations of multiple organ failure, or in assessing side effects of pharmacological treatment when targeting a specific system may also affect other systems via the network of physiologic interactions.

3. Of importance to complex networks, we show that there is a robust interplay between network topology and function. In network theory it is hypothesized that network function is influenced by network structure, however, examples on real networks did not exist prior to these investigations.

   We demonstrate that each physiologic state is associated with a specific network of physiologic interactions that is characterized by a given topology, node connectivity, number and strength of network links. A similar network topology and strength of network links is consistently observed for individual subjects in the same physiologic state, indicating *universal* behavior.

   In particular, relating physiologic function to network topology we show that during deep sleep several integrated systems (e.g., cardiac, respiratory and brain) act as if disconnected from each other. This is a principally new information, which can explain (i) why earlier studies have found that correlations and scaling properties in heartbeat intervals break down and exhibit close to random behavior during deep sleep (as it would be the case of a denervated heart), and (ii) why people awakened during deep sleep do not adjust immediately and often feel groggy and disoriented for a few minutes.

   Since specific mechanisms regulate physiologic function during each physiologic state, our observations provide a first empirical evidence on a real network of a robust relation between network structure and function.

4. We identify phase transitions in network topology. There is no precedent of such behavior. We quantify the process of transition by tracking the network evolution in time.

   We find that with transition from one physiologic state to another (for example across sleep stages), network topology dramatically changes within a short time window of 2–3 min—from only a few links to a multitude of links—indicating a remarkable flexibility in the interaction between physiologic systems in response to change in physiologic regulation. Such change in network structure in response to change in the mechanisms of control during different physiologic states indicates that our findings reflect *intrinsic* features of physiologic interaction.

   Further, we find that transitions from one physiologic state to another trigger a remarkably-structured reorganization of physiologic interactions. This

reorganization occurs simultaneously and globally in the entire network as well as at the level of individual network nodes (physiologic systems), while preserving a hierarchical order in the strength of network links.

Although our study is limited to a data-driven approach the empirical findings may facilitate future efforts on developing and testing network models of physiologic interactions. In relation to critical clinical care, where multiple organ failure is often the reason for fatal outcome [1, 2], our framework may have practical utility in assessing whether dynamical links between physiological systems remain substantially altered even when the function of specific systems is restored after treatment [3]. While we demonstrate one specific application, the framework we developed can be applied to a broad range of complex systems where the TDS method can serve as a tool to characterize and understand the dynamics and function of real-world heterogeneous and interdependent networks. The established relation between dynamical network topology and network function has not only significant medical and clinical implications, but is also of relevance for the general theory of complex networks, including the emerging field of networks of networks.

References

1. Buchman, T. G. *Complex Systems Science in BioMedicine,* chap. Physiologic failure: multiple organ dysfunction syndrome, 631–640 (New York: Kluwer Academic/Plenum Publishers, 2006).
2. Deitch, E. A. Multiple organ failure. Pathophysiology and potential future therapy. *Ann Surg* **216**, 117–134 (1992).
3. Lizana, F. G., Alonso, J. L. M., Santana, B. G., Esteban, J. F. & Santana, P. S. Survival and quality of life of patients with multiple organ failure one year after leaving an intensive care unit. *Med Clin (Barc)* **114 Suppl 3**, 99–103 (2000).
4. Bassingthwaighte, J. B., Liebovitch, L. & West, B. J. *Fractal Physiology*. (Oxford University Press, 1994).
5. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
6. Bunde, A. *et al.* Correlated and uncorrelated regions in heart-rate fluctuations during sleep. *Phys. Rev. Lett.* **85**, 3736–3739 (2000).
7. Dvir, I., Adler, Y., Freimark, D. & Lavie, P. Evidence for fractal correlation properties in variations of peripheral arterial tone during REM sleep. *Am. J. Physiol. Heart Circ. Physiol.* **283**, H434–H439 (2002).
8. Ivanov, P. Ch. et al. Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis. *Nature* **383**, 323–327 (1996).
9. Schäfer, C., Rosenblum, M. G., Kurths, J. & Abel, H. H. Heartbeat synchronized with ventilation. *Nature* **392**, 239–240 (1998).

10. Shlesinger, M. F., Zaslavsky, G. M. & Klafter, J. Strange kinetics. *Nature* **363**, 31–37 (1993).
11. Collins, J. J., Imhoff, T. T. & Grigg, P. Noise-enhanced tactile sensation. *Nature* **383**, 770 (1996).
12. Hegger, R., Bunner, M. J., Kantz, H. & Giaquinta, A. Identifying and modeling delay feedback systems. *Phys. Rev. Lett.* **81**, 558–561 (1998).
13. Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Rev. Mod. Physics* **74**, 47–97 (2002).
14. Song, C., Havlin, S. & Makse, H. A. Self-similarity of complex networks. *Nature* **433**, 392–395 (2005).
15. Albert, R., Jeong, H. & Barabási, A. L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
16. Cohen, R., Havlin, S. & Ben-Avraham, D. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.* **91**, 247901 (2003).
17. Newman, M. E. J. The structure and function of complex networks. *Siam Rev.* **45**, 167–256 (2003).
18. Dorogovtsev, S. N. & Mendes, J. F. F. Evolution of networks. *Advances in Physics* **51**, 1079–1187 (2002).
19. Williams, R. J. & Martinez, N. D. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
20. Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028 (2010).
21. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
22. Banavar, J. R., Maritan, A. & Rinaldo, A. Size and form in efficient transportation networks. *Nature* **399**, 130–132 (1999).
23. Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci. USA* **103**, 2015–2020 (2006).
24. de Menezes, M. A. & Barabási, A.-L. Fluctuations in network dynamics. *Phys. Rev. Lett.* **92**, 028701 (2004).
25. Mantegna, R. Hierarchical structure in financial markets. *Eur. Phys. J. B* **11**, 193–197 (1999).
26. Newman, M. E. J. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98**, 404–409 (2001).
27. Tumminello, M., Di Matteo, T., Aste, T. & Mantegna, R. N. Correlation based networks of equity returns sampled at different time horizons. *Eur. Phys. J. B* **55**, 209–217 (2007).
28. Tumminello, M. *et al.* Happy aged people are all alike, while every unhappy aged person is unhappy in its own way. *PLoS One* **6**, e23377 (2011).
29. Tumminello, M., Lillo, F., Piilo, J. & Mantegna, R. N. Identification of clusters of investors from their real trading activity in a financial market. *New Journal of Physics* **14** (2012).
30. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
31. Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).
32. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. The large-scale organization of metabolic networks. *Nature* **407**, 651–653 (2000).
33. Bhalla, U. S. & Iyengar, R. Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387 (1999).
34. Strogatz, S. H. Exploring complex networks. *Nature* **410**, 268–276 (2001).
35. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D. U. Complex networks: Structure and dynamics. *Phys. Rep.* **424**, 175–308 (2006).
36. Ivanov, P. Ch. *et al.* Sleep-wake differences in scaling behavior of the human heartbeat: analysis of terrestrial and long-term space flight data. *Europhys. Lett.* **48**, 594–600 (1999).
37. Karasik, R. *et al.* Correlation differences in heartbeat fluctuations during rest and exercise. *Phys. Rev. E* **66**, 062902 (2002).

38. Ivanov, P. Ch., Hu, K., Hilton, M. F., Shea, S. A. & Stanley, H. E. Endogenous circadian rhythm in human motor activity uncoupled from circadian influences on cardiac dynamics. *Proc. Natl. Acad. Sci. USA* **104**, 20702–20707 (2007).

39. Angelone, A. & Coulter, N. A. Respiratory sinus arrhythmia: a frequency dependent phenomenon. *J. Appl. Physiol.* **19**, 479–482 (1964).

40. Song, H.-S. & Lehrer, P. M. The effects of specific respiratory rates on heart rate and heart rate variability. *Appl. Psychophysiol. Biofeedback* **28**, 13–23 (2003).

41. Tass, P. et al. Detection of n : m phase locking from noisy data: Application to magnetoencephalography. *Phys. Rev. Lett.* **81**, 3291–3294 (1998).

42. Bartsch, R., Kantelhardt, J. W., Penzel, T. & Havlin, S. Experimental evidence for phase synchronization transitions in the human cardiorespiratory system. *Phys. Rev. Lett.* **98**, 054102 (2007).

43. Leeuwen, P. V. *et al.* Influence of paced maternal breathing on fetal-maternal heart rate coordination. *Proc. Natl. Acad. Sci. USA* **106**, 13661–13666 (2009).

44. Ivanov, P. Ch., Ma, Q. D. Y. & Bartsch, R. P. Maternal-fetal heartbeat phase synchronization. *Proc. Natl. Acad. Sci. USA* **106**, 13641–13642 (2009).

45. Schumann, A. Y., Bartsch, R. P., Penzel, T., Ivanov, P. Ch. & Kantelhardt, J. W. Aging effects on cardiac and respiratory dynamics in healthy subjects across sleep stages. *Sleep* **33**, 943–955 (2010).

46. Bartsch, R. P., Schumann, A. Y., Kantelhardt, J. W., Penzel, T. & Ivanov, P. Ch. Phase transitions in physiologic coupling. *Proc. Natl. Acad. Sci. USA* **109**, 10181–10186 (2012).

47. Bashan, A., Bartsch, R. P., Kantelhardt, J. W., Havlin, S. & Ivanov, P. Ch. Network physiology reveals relations between network topology and physiological function. *Nat. Commun.* **3**, 702 (2012).

48. Otzenberger, H. *et al.* Dynamic heart rate variability: a tool for exploring sympathovagal balance continuously during sleep in men. *Am. J. Physiol.* **275**, H946–H950 (1998).

49. Podobnik, B., Fu, D. F., Stanley, H. E. & Ivanov, P. Ch. Power-law autocorrelated stochastic processes with long-range cross-correlations. *Eur. Phys. J. B* **56**, 47–52 (2007).

50. Kryger, M. H., Roth, T. & Dement, W. C. (eds.) *Principles and Practice of Sleep Medicine* (Elsevier Saunders, Philadelphia, 1994).

51. Siegel, J. M. Clues to the functions of mammalian sleep. *Nature* **437**, 1264–1271 (2005).

52. Massimini, M. et al. Breakdown of cortical effective connectivity during sleep. *Science* **309**, 2228–2232 (2005).

Part III
Phenomenological Models

Under the definition of 'phenomenological models' we have included different attempts to reproduce the real behaviour of the critical infrastructure. These models are able to provide detailed—and realistic—information on the behaviour of critical infrastructures and the fault-propagation mechanisms at different levels of abstraction. The higher is the accuracy of the model, the larger is the amount of data and time required for the calculation. Moreover, an extremely detailed output can represent a challenge in order to understand the emergent behaviour of the whole system.

Chapter 11 presents the state of the art for the so-called federated simulations. Such approach consists in merging in a common software the simulation tools normally employed by the asset owners of the single critical infrastructures. By this means it provides extremely detailed predictions.

Chapter 12 represents a compromise between the need of details and the urge of including the effects of the behaviour of the human actors (both decision-makers and users). To this purpose, agent-based models are widely employed.

Chapter 13 compares the different approaches under the perspective of the evaluation of the systemic risk.

Chapter 14 specialises the systemic analysis to the interdependent components of the financial system. This allows the definition and implementations of metrics and tools that are currently applied by financial institutions to assess the systemic part of the risk.

Chapter 6 focuses on the spatial and temporal structure of the inter-dependencies among infrastructural networks. Such approach enables an adaptive optimization of countermeasures both to contain ongoing damage and to speed-up the recovery process.

Chapter 11
Federated Modelling and Simulation for Critical Infrastructure Protection

Erich Rome, Peter Langeslag and Andrij Usov

Abstract   Modelling and simulation is an important tool for Critical Infrastructure (CI) dependency analysis, for testing methods for risk reduction, and as well for the evaluation of past failures. Moreover, interaction of such simulations with external threat models, e.g., a river flood model, or economic models enable consequence analysis and thus may assist in what-if decision-making processes. The simulation of complex scenarios involving several different CI sectors requires the usage of heterogeneous federated simulations of CIs. However, common standards for modelling and interoperability of such federated CI simulations are missing. Also, creating the required abstract models from CIs and other data, setting up the individual federate simulators and integrating all subsystems is a time-consuming and complicated task that requires substantial know-how and resources. In this chapter, we outline applications and benefit of federated modelling, simulation and analysis (MS&A) for Critical Infrastructure Protection (CIP). We review the state of the art in federated MS&A for CIP and categorise common approaches and interoperability concepts like central and lateral coupling of simulators. As examples for the latter two concepts, we will present in more detail an interoperability standard from the military domain, HLA, and an approach developed in the DIESIS project. Special emphasis will also be put on describing the problem of synchronising systems with different time models. Also, we will briefly assess the state of transferring MS&A for CIP

E. Rome (✉) · A. Usov
Fraunhofer IAIS, Schloss Birlinghoven, 53754 Sankt Augustin, Germany
e-mail: erich.rome@iais.fraunhofer.de
http://www.iais.fraunhofer.de

A. Usov
e-mail: andrij.usov@iais.fraunhofer.de
http://www.iais.fraunhofer.de

P. Langeslag
TNO Defence, Security and Safety, Oude Waalsdorperweg 63, The Hague, The Netherlands
e-mail: peter.langeslag@tno.nl
http://www.tno.nl

research results to practical application by comparing the situations in the USA and in Europe.

11.1 Introduction

Infrastructures operate globally and are increasingly dependent and interdependent: a breakdown or disruption of functions may have serious national or even multi-national consequences [1]. The disruptions of the power grids in 11 European states and Morocco on November 4, 2007 affecting 15 million people are a case in point. It is for this reason that such infrastructures can be called Critical Infrastructures (CI). A CI is defined by [2] as an asset, system or part thereof that is essential for the maintenance of vital societal functions, health, safety, security, economy or social well-being of people, and the disruption or destruction of which would have a significant impact as a result of the failure to maintain those functions. Clearly, it is mandatory to protect these assets. In this respect, modelling, federated simulation, and analysis are of vital importance [3]. They are required for the investigation of CI and their dependencies, for training CI operators and crisis managers, as well as for the development of methods for Critical Infrastructure Protection (CIP).

Simulation is particularly well-suited for capturing dynamic effects within the complex system of interconnected critical infrastructure systems, like cascading effects. A failure or loss of service in one infrastructure, like power transmission, may cause a loss in a dependent infrastructure, like railway transport. The simulation and investigation of large scenarios with cascading CI failures affecting critical services in multiple CI requires the use of **federated** simulations consisting of simulators and suitable models[1] for each of the involved CI [3, 4]. In addition, models that may generate threats to the CI and models that analyse the consequences (e.g., economic loss, number of casualties, affected area), and visualisation as well as other real-time tools may need to take part in the federation. These simulators and components need to be coupled by means of a suitable middleware that allows the synchronisation of events and simulation times, the exchange of data, and the exertion of control functions like starting and stopping simulations (Fig. 11.1). That is, components of a federated simulation need to be **interoperable**. Different from the situation in most military simulations, current CI simulators are in general not interoperable and often lack proper interfaces.

Some federated simulations need to be realised as distributed systems, for various reasons. Depending on the computational demands of individual simulators that

---

[1] In this publication, we refer to 'simulation' as the dynamic part of a computer model, and we refer to 'model' as the static part of a computer model.
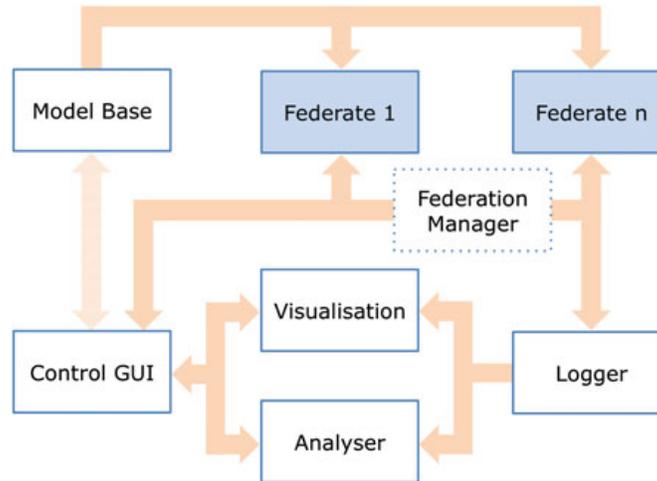
Fig. 11.1 A generic schema for modelling, federated simulation and analysis for CI. A model base contains abstract (conceptual) models for each of the CI simulators (federates). Simulation is started by a control GUI. Federates exchange data with each other and report status to control GUI. Intermediate states and results of the federates are logged by a data logger. A visualisation module can be employed for visualising results and simulation states online (while the simulation is running) or offline (after the simulation terminated). Analysis tools can provide additional information by evaluating the results online or offline. Sometimes, one of the federate components serves as an orchestrator or manager that controls the simulation steps of all individual simulators

comprise the federation, it is sometimes not commendable or possible to run the entire federation on a single computer, e.g. when federates require different operating systems. And lastly, in a collaborative research project individual simulators composing a federation may need to run on different partners' computers for license or financial reasons. **Distributed (federated) simulation** includes running a federation on a local computer grid or in a locally distributed fashion via the Internet or other communication infrastructure.

After the above brief characterisation of federated simulation systems, we will describe applications, technologies, state of the art, challenges, and standardisation of federated modelling, simulation and analysis (MS&A) in more detail in the remainder of this chapter. It is organised as follows. First, we describe application areas of federated MS&A. Then we take a look at the basic technical properties of federated simulations and describe current interoperability approaches and the synchronisation problem. Then we review the state of the art in federated MS&A for Critical Infrastructure Protection and describe in detail two different interoperability approaches, HLA and DIESIS. We conclude with summarising the main insights.

11.2 Applications of Federated Modelling, Simulation and Analysis

Federated modelling, simulation and analysis has a wide range of applications, with two foci, research—as aid for and subject of—on one hand and applications in security on the other hand. Both foci are closely related. The US American facility NISAC employs federated MS&A for homeland security. It emerged from a cooperation between two research institutes, the Sandia and Los Alamos National Laboratories in the year 2000 and is now part of the Department of Homeland Security. The research institutes successfully managed transferring their CIP expertise and federated MS&A technology to practical applications. In Europe, federated MS&A for CIP is still rather a subject of research, but there is a small community working towards a European facility, comparable to NISAC, that shall provide expertise and technology to offices, institutions, and people responsible for Critical Infrastructure Protection and Civil Security.

In this section, we will present both typical applications of federated MS&A and outline emerging applications. It should be noted here that developments of federated MS&A were and are strongly driven from applications in the military domain, for instance, the HLA middleware standard. That is, some of the state of the art presented here stems from the military domain, but is relevant also for CIP.

11.2.1 Investigating Dependencies Between Critical Infrastructures

For this chapter, we adopt the definition of dependent and interdependent CI given by Luiijf et al. [5, p. 304], "A CI dependency is the relationship between two CI products or services in which one product or service is required for the generation of the other product or service; a CI interdependency is a mutual CI dependency." A loss of service or reduction of quality of a service in one infrastructure may lead to loss of service in a dependent second infrastructure, this again may lead to loss of service in a dependent third infrastructure, and so on. This is called a cascading effect. Since CI form a very complex system of dependent systems at various scales (local, regional, national, continental, global), it is difficult to understand the nature and effects of these dependencies. A prerequisite of developing methods and measures for protecting CI or making them more resilient was an improved understanding of the dependencies. Consequently, this type of investigation was a major focus in the CIP research area over the last 10 years. While some dependencies are of static nature (by construction or local neighbourhood of CI elements), the more difficult cases are dynamic dependencies, occurring while CI are operating. Since it is prohibitive to use real CI for investigation, researchers started using conceptual models and computer simulation of CI for investigating dependencies and interdependencies of CI (Examples: [3, 4, 6–17]). An example of a dynamic effect is a delayed cascading effect, like a loss of power supply for a hospital that has a diesel generator as a backup

power supply, which fails several hours later after it runs out of fuel. Depending on the duration of the power outage, this cascading effect may or may not happen.

11.2.2 Soft Exercises and Training

One of the essential elements for protecting Critical Infrastructures is maintaining or achieving a high level of preparedness of staff responsible for security, like crisis and emergency managers. This requires practice in real emergencies and crises as well as training of simulated emergencies and crises in exercises. Typically, national or regional exercises take place once a year, while some enterprises do monthly exercises. Such exercises are necessary and useful. However, given the wide range of potential scenarios, annual practical exercises seem not sufficient for being prepared for even the most likely scenarios in an ever-changing world. Computer simulation would be suited as an additional means for exercising mitigation of crises and emergencies, and raising awareness of the role of CI in crises and emergency situation [18]. Similarly, federated M&S of CI could be used to train operators of CI for mitigating crisis and emergency situations. Here, scenarios and scripts of the simulations could be altered to cover a wide range of possible situations.

11.2.3 Decision Support

In cases of crises, crisis and emergency managers may encounter situations in which different courses of action are possible. Decision Support Systems (DSS) provide methods for assessing the consequences of certain decisions and thus may aid crisis and emergency managers in taking the right decisions. By using simulations, DSS can be enhanced to perform 'what if' analyses, that is, dynamically explore the different courses of action and their different consequences. In this way, these end users are enabled to plan the most effective use of resources in an emergency and to explore a variety of scenarios, for example:

- which region to evacuate first, which infrastructures to reinforce best/first,
- which transport or traffic infrastructures required for a mitigation plan will be affected by a disaster and what contingency planning is required,
- which infrastructures outside a region affected by a disaster need to be operational in order to supply that region and thus need to be protected too.

Examples of this type of application are I2Sim [19], described later in this chapter, and the work of Tolone et al. [15].

11.2.4 Environment for Testing and Benchmarking New Methods

The IRRIIS project ([20], cf. also below) developed a federated simulation of the interdependent power distribution and telecommunication networks of two

infrastructures operators. The simulation was orchestrated by means of an agent-based simulator called SimCIP [21]. During the simulation, early indicators for reductions of quality of service or loss of service were computed independently for each of the four simulated infrastructure topologies. The project investigated whether a risk reduction could be achieved by communicating the early risk indicators between the two infrastructure operators [22]. A potential future extension of this type of application, proposed in [23], is using federated MS&A for benchmarking competing new methods for risk reduction in or protection of CI.

11.3 Basic Technical Aspects of Federated Simulation

11.3.1 Interoperability Approaches

The assessment of simulator interoperability provided by Usov et al. still holds today: "In recent years, a large number of projects have investigated and tested methods for coupling simulators. As a general result it can be stated that the technological task of coupling simulators is highly demanding and that there are no ideal general purpose solutions for the coupling task, but the applied methods are strongly determined by the general requirements and the application task at hand" [24]. In this section, we will review some basic interoperability approaches.

A connection creates a communication link between two or more systems. Interoperability between these systems could be considered from a double point of view [25]: *technical connectivity* and *semantic connectivity*.

Technical connectivity considers in which way systems are able to solve the problem of sharing and exchanging data across multiple platforms. It is strongly related to the capability of systems to implement a common data structure and syntax in order to achieve a connection among them. This aspect of interoperability implies that the exchanged information is understandable by any other system not initially developed for the cooperation. So a common language is an essential requirement. It enables the description of the structure and syntax of the underlying data.

To achieve a meaningful connection among systems it is necessary to establish how they can exchange information or in which way they combine other information about resources and subsequently the way to process them, in a significant manner. Thus, **semantic connectivity** requires agreement on a wide variety of issues relating to the context within information is created and used. The aim is not only to allow information resources to be linked up, but also to give context to information in a scenario in which different systems have their own perspective on that. Only in this way information can be automatically understandable and, consequently, reusable by systems that were not involved in its creation. Thus, the semantic connectivity concerns the need to agree on common definitions and to understand information that is necessary to exchange.
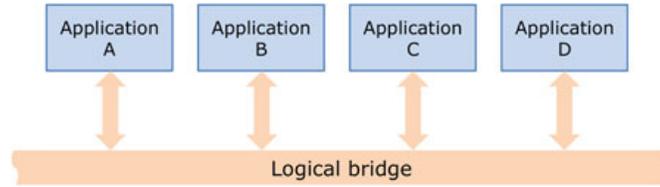
Fig. 11.2 Central coupling: Connected systems exchange information via a common logical bridge, using a standardised exchange format
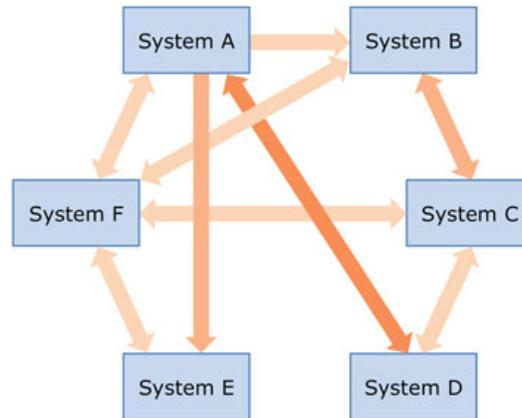

Fig. 11.3 Lateral coupling: Connected systems exchange information bi- or multi-laterally. Exchange formats may vary

For connecting simulations, two technologies can be considered: central coupling and lateral coupling. The **central coupling** topology is of a typical architecture oriented to the distribution of services, where the applications communicate each other through a logical channel or bridge (Fig. 11.2). From an architectural point of view, such logical channel is based on a software layer centralising functionalities. It supports synchronous and asynchronous communication based on messages and intelligent routing as well as data transformation and connectivity towards heterogeneous applications. Typically, such a logical channel is implemented by means of middleware, usually based on standards. It supplies the fundamental services for more complex architectures through event-driven and message passing mechanisms. Central coupling approaches with standardised exchange formats, such as HLA [26–29] or OpenMI [30, 31], are particularly suited when all federates support the exchange format standard. If this is the case, the integration of a new simulator is relatively easy and is limited to the implementation of an interface between this simulator and interoperability middleware which already contains ready-made solutions for communication, time management, etc.

Despite of its convenience, central coupling approaches are generally not applicable if federation members do not use one and the same interoperability standard. Especially the time management is often a challenging problem that has only one-off solutions for particular federations but is unsolvable in general, for arbitrary

combinations of models and technologies (see Sect. 11.3.2). Other than simulators in the military area, federations of CI simulations are often quite heterogeneous in terms of interfaces, modelling approaches and even time scales. For realisation of such a federation, a **lateral coupling** topology [24] is recommended. This architecture foresees the development of dedicated links between pairs of federates, according to their logical interconnections (Fig. 11.3). Besides of pairwise couplings, the resulting federation may also contain centrally coupled clusters of simulators that all support a certain interoperability standard. A drawback of the lateral coupling approach is obviously the large number of links that has to be developed for creating a federation as well as for adding a new member to an existing one. Creating potentially reusable links and storing them in a kind of repository is a possible solution for this problem. Application of lateral coupling architecture for federations of CI simulators and corresponding strategies are discussed in Sect. 11.4.5

11.3.2 Time Models and Synchronisation

The semantic connectivity described in the previous section has two essential aspects: regulation of data exchange between federates and correct interpretation of the exchanged data. Here, we will discuss the former aspect, and the latter one will be handled in Sect. 11.3.3. Regulation of data exchange means, in the first line, the preservation of causality. In other words, it is necessary to ensure that particular *events* (i.e., discrete messages that usually represent state changes) are processed by all federates in a logically correct order. This is an essential requirement for the reproducibility of simulation results for the same scenario, which is required for many evaluation techniques. For training applications, minor deviations from the correct event order can often be tolerated as long as perceived realism of simulation results is not violated [32].

   The problem of time management for federated simulations is as old as the idea of federated simulation itself. First formulations of this problem and corresponding first solutions were published in the late 1970s (e.g., by Chandy and Misra [33]). However, since then, many new approaches or variations of already existing approaches have been developed and published. The reason for this continuing interest in the time management problem is that there is no universal optimal solution for all possible simulators with their different time and execution models. Irrespective of their *internal* time representations, the simulators may offer interaction capabilities and provide simulation results according to the following schemes:

- **Steady state:** a simulation runs until a steady state of the internal model (or some state where reasonable simulation results can be produced) is reached. This kind of computation is often employed for relatively fast processes and sometimes they even have no notion of time in their models (e.g., switching actions in a power distribution network).

- **Discrete events:** simulation results are produced after variable simulation time intervals that are determined by occurrence of some internal event or by reaching some state in which the simulator assumes that interaction with its environment is required.
- **Constant time steps:** simulators of this type are clocked and offer interaction after every constant simulation time period. Simulation steps are usually configurable: shorter steps increase the accuracy of results, longer steps improve the simulation performance.
- **Real time:** those simulations are usually employed for training with human in the loop. They often work with small constant time steps that are synchronised with the real time. Hence, their capability to wait for simulation results produced by other federates is very limited. Interaction with real time simulators places high demands on performance of other federates and of the middleware.

The choice of an appropriate time management approach depends on the combination of different time models within a federation. Another factor that determines the synchronisation strategy is the desired usage of simulation results: for exact results evaluation the synchronisation algorithm has to work exactly too. Finally, the ability of simulators to return to a time point in the past by rolling back recent state changes allows to employ advanced synchronisation techniques. Most time management implementations are based either on *conservative* or on *optimistic* synchronisation approaches. In this section, we will provide only a brief introduction into the core ideas of these techniques. More detailed description and an overview over related algorithms can be found, for example, in works by Fujimoto [32] and Riley et al. [34].

Conservative time management algorithms prevent causality violations by restricting event processing to "safe" events. An event can be safely processed by a particular federate only if it is absolutely certain that no other events assigned to an earlier simulation time point will be received by this simulator afterwards. An essential prerequisite for the application of this approach is the ability to compute a *lookahead* (i.e., remaining simulation time until the next "public" event) for all federates. Several solutions were developed in order to detect and avoid deadlocks as well as to minimise the communication overhead [35, 36] of conservative synchronisation algorithms.

Optimistic synchronisation approaches use another strategy for keeping the event order correct. Federates are explicitly allowed to process potentially "unsafe" events. However, if an event with smaller simulation time stamps arrives later, simulators must be prepared to roll back all effects of recently processed events and to reprocess events in a correct order. Optimistic synchronisation provides a higher degree of parallelism and potentially better simulation performance. However, its major drawback is a high demand on memory that is required for storing checkpoints for possible rollbacks. Furthermore, simulations may be significantly slowed down by an increasing number of rollbacks which is highly scenario-dependent. This problem is even exacerbated by the fact that processing an event may produce new events that have to be sent to other simulators. In this case, the rollback process includes the cancellation

of sent events and, hence, it initiates rollbacks at other federates. A "too optimistic" event processing strategy may unleash extremely time-consuming rollback cascades.

Unlike conservative approaches, the optimistic ones do not rely on the computation of lookaheads. On the other hand, the ability of federates to maintain checkpoints and to perform rollbacks is required. The Time Warp algorithm published by Jefferson in 1985 [37] was probably the first optimistic synchronisation approach. It foresees "anti-messages" for cancelling already sent events in case of a non-local rollback. In the early 90s, several modifications of the Time Warp algorithm as well as completely new ideas were developed in order to decrease memory consumption [38] and to avoid costly distributed rollbacks [39, 40].

Neither conservative nor optimistic time management algorithms can provide a universally optimal solution for arbitrary federations due to the fact that the efficiency in both approaches highly depends on scenarios (global event order), on federation topology (logical links among simulators) as well as on specific features supported by particular simulators (lookahead computation and rollback functionality). In the area of CIP, it is often required to simulate interactions of different infrastructures that are described by quite different physical and logical laws. The resulting federation can be extremely heterogeneous and may contain simulators that internally work with *completely* different time scales. Typical temporal intervals between events are milliseconds for power network and communication domains, seconds for urban traffic, minutes for evacuation and smoke propagation and hours for flooding simulations. Some simulators may not support rollbacks, since others may be unable to forecast their lookaheads. Hence, the choice of an appropriate time management solution for a federation of CI simulators is determined by the composition of this federation. Furthermore, there is no guarantee that a globally applicable solution exists. In this case, according to the idea of lateral interoperability, different time management approaches have to be employed for particular pairs or clusters of simulators within the federation (see Sect. 11.3.1).

11.3.3 Modelling for Federated Simulation

As stated in the previous section, the correct interpretation of the exchanged data is of utmost importance for a federation to become more than the sum of its federates. Also, this data has to be interpreted in the right manner. Another issue that rises is that the federates are created and validated for a special purpose. Using them as part of a federation may cause the validation to become insufficient and thus it may become necessary to redo the validation of the federates. In addition, the overall validation needs to be regarded. This chapter will handle both cases.

The easiest way to guarantee the information exchange to contain the correct data is to anchor this in the data exchange protocol, as it is done with DIS [41]. This can easily be accomplished with simulators in the same domain where linking simulators is common knowledge (DIS was developed for the defence domain). For the world of CI where multiple domains are involved and linking of simulators is not common, it

is unlikely that simulator developers will adapt to a common data exchange protocol. Also, a disadvantage of DIS is the use of broadcast for data exchange. All information is sent to every federate, which can result in a big bandwidth consumption.

A more flexible approach is used in HLA with the use of Object Models (OM) [42]. HLA object models are composed of a group of interrelated components specifying information about classes of objects, their attributes, and their interactions. Every federate has its own federate object model (FOM), but it has to be compliant with the object model of the federation, the simulation object model (SOM). HLA uses a publish and subscribe mechanism for data exchange. Therefore, network bandwidth can be adapted to actual needs.

Although the HLA method gives a much more flexible approach for linking simulators compared to DIS, it is still based on a common data model, which must be implemented by all federates. Also, it is a purely syntactic model which works well for use within one domain, but it does not provide the semantic information about the modelled domains needed in a multi domain environment. Masucci et al. [43] describe several modelling and simulation approaches to analyse critical infrastructure interdependencies and conclude with an ontology based modelling and simulation solution called the DIESIS KBS architecture. The DIESIS KBS design incorporates a meta knowledge "world" infrastructure ontology (WONT), infrastructure ontologies (IONTs), a federation ontology (FONT) and gateway components (see Sect. 11.4.5).

As Masucci et al. [43] describe, the DIESIS KBS is designed for creating abstractions of critical infrastructure domains and to represent and formalize their parameters and dependencies. The KBS is intended to be used in a federated simulation environment to study the behaviour of infrastructures and their components under different conditions and constraints. The resulting federated environment will support complex simulation scenarios involving multiple infrastructures with different semantics and granularities (or fidelities).

Making all the federates in the federation understand each other's data solves only half the problem. Also, the way they use the data and the level to which extend the model describes the real world domain should be taken in account because these end up in defining the credibility or the final outcome of the federation. For this reason, verification and validation (V&V) of the federation should be performed from the early begin of the process up to the end. SISO [44] provides a generic methodology for verification and validation to support acceptance of models, simulations and data. The objective of the V&V effort is to develop an acceptance recommendation that convincingly shows why a federate or federation is acceptable or not acceptable for the intended use. This V&V objective is articulated as an acceptance goal. This high-level goal should be translated into a set of concrete and assessable acceptability criteria for the federate or federation. Relevant and convincing evidence should then be collected or generated to assess the satisfaction of these criteria. When it is convincingly demonstrated to what extent the federate or federation does or does not satisfy all these acceptability criteria, a claim can be made on whether or not the federate or federation is acceptable for its intended use (i.e., acceptance claim).

11.4 State of the Art in Federated MS&A for CIP

In the last decade, the awareness has grown that Critical Infrastructures are in a greater or lesser extent dependent on each other. Investigating, exercising and training of CI behaviour in case of an event can not be done by one system alone. Therefore the need for combining the interdependent systems in a simulation environment has grown. In these past few years, several initiatives have been taken to combine (parts of) the different simulations for critical infrastructures in or across different domains. This chapter describes many of these efforts. The characterised works within this section can be divided roughly into three—not entirely disjunct—categories:

1. Special purpose federated simulation systems, consisting of a number of simulators (CI and others), additional system components, and a dedicated middleware for communication and synchronisation (IRRIIS, EPOCHS, ...),
2. Frameworks for modelling, simulation and analysis of CI using dedicated—for instance, agent-based—simulations (I2Sim, AIMS, IME, ...),
3. More general frameworks for setting up distributed federations and more general middleware for communication and synchronisation within federations (IDSim, ASimJava, ...), including (quasi-)standards (OpenMI, HLA, ...), and sometimes accompanied by proofs-of-concept (DIESIS, XMSF, WSIM, ...).

More elaborate presentations will be given for one example of a framework for central coupling (HLA, the High Level Architecture standard) and for one example of lateral coupling (the DIESIS approach). Older publications that include overviews on the state of the art in MS&A for CIP are [6, 45].

11.4.1 Special Purpose Federated Simulation Systems

With its SimCIP [21] modelling and simulation environment, the **IRRIIS** [20] project used an agent-based environment where components, subsystems and systems are represented by autonomous agents and the simulation is synchronised through a centralistic RTI-like simulation engine, LAMPS. Within the IRRIIS demonstrator, SimCIP orchestrates a federation of the SINCAL power transmission simulator and the NS2 network communication simulator, modelling the dependencies between power distribution and communication networks in a large European capital. The target application of IRRIIS was twofold, namely investigating dependencies and interdependencies, and risk reduction by communicating early risk indicators between operators of the two infrastructures [22]. The limitation of SimCIP though is that no standardised definition or workflow is proposed for the extension of the environment through new simulators.

The **EPOCHS** approach [46] was driven by the need to better understand the effects of integrating network communication systems into electric power control systems on the stability of the electric power systems. The task required the
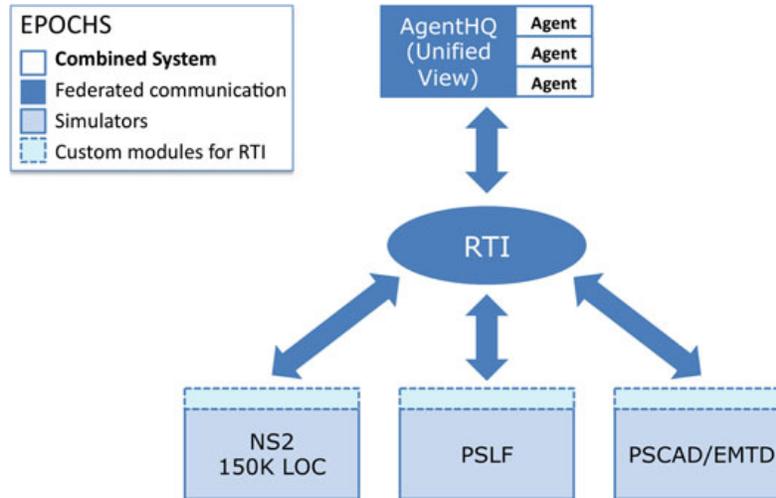
Fig. 11.4 Federation architecture of EPOCHS (after [46])

combination of three different high-fidelity simulation systems: A simulator for electromagnetic transient simulation (PSCAD/EMTDC), a simulator for electro-mechanical transient simulation (PSLF), and a network communication simulator (NS2 [47]). Each of these three simulators was designed as stand-alone simulator, that is, all three simulators lack built-in interoperability. Hopkinson et al. [46] implemented a specific solution for creating a federation of the three chosen simulators. It consists of three basic elements: a Run-Time Infrastructure for enabling time management and communication between the three federate simulators, an agent-based control interface for the user of the federated simulation, and individual extensions of the three simulators enabling the compatibility with the EPOCHS RTI in order to make them interoperable. The resulting architecture is shown in Fig. 11.4.

For enabling the compatibility with the EPOCHS RTI, the designers chose three different ways, based on the properties of the three different simulators [46, p. 5]. For the communication network simulator NS2, they used the fact that source code was available and added a new transport protocol for realising RTI access. The power simulator PSCAD/EMTDC allows external function calls. The EPOCHS designers used this feature for creating an external component that gets active at each time step of the simulator, reading and/or writing equipment values from/to the RTI before the simulation continues. For the power simulator PSLF, they used a similar solution. PSLF does not allow calling simulator functions, but allows writing extensions in its own programming language EPCL. The EPOCHS designers programmed a communication stub in EPCL that writes or reads simulation values to or from a file upon request from the RTI. The EPOCHS approach is an example of central coupling with non-standardised interfaces.

The motivation of Riley et al. [34] is the usage of simulation as a tool for analysis of communication network problems and validation of models of communication networks. For this task, the computer simulation of a communication network
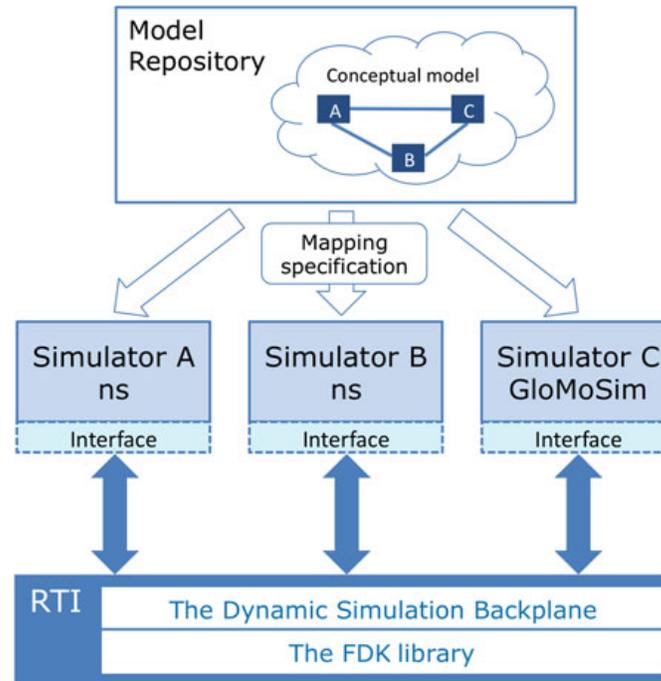
Fig. 11.5 Architecture for a high fidelity distributed federated communication network simulation, self-federating two instances of the Network Simulator ns (after [34])

requires a high degree of fidelity, which leads to high computational demand: "It would take several days to simulate just one minute of operation of this network" [34, p. 118]. Riley et al. propose a twofold solution that is suited to reduce the high ratio of simulation time versus real time. For the first part, they propose to distribute the simulation across multiple networked processors, which results in faster simulation. For the second part, they propose to setup the distributed simulation as a federation of several simulators, which allows simulating larger networks. In this case, the federates could also be several instances of the same simulator system (self-federated). Figure 11.5 shows the architecture concept that Riley et al. used for realising their distributed federated communication network simulation. It fits the generic scheme presented earlier in this chapter.

The Run-Time Infrastructure performs synchronisation and data distribution and uses a library called the Federated Simulations Development Kit (FDK library). A Dynamic Simulation Backplane ensures syntactic compatibility for data exchange between federates where possible, enables the exchange of meaningful event messages, checks for incompatibilities and provides more functions for communication between federates.

As an interesting experimental result, Riley et al. [34, p. 146] report that for two network simulators, PDNS and GTNets, they were able to show that self-federating each of these simulators enabled simulation of large network topologies (almost 2 million nodes) with "linear efficiency" up to 128 federates.
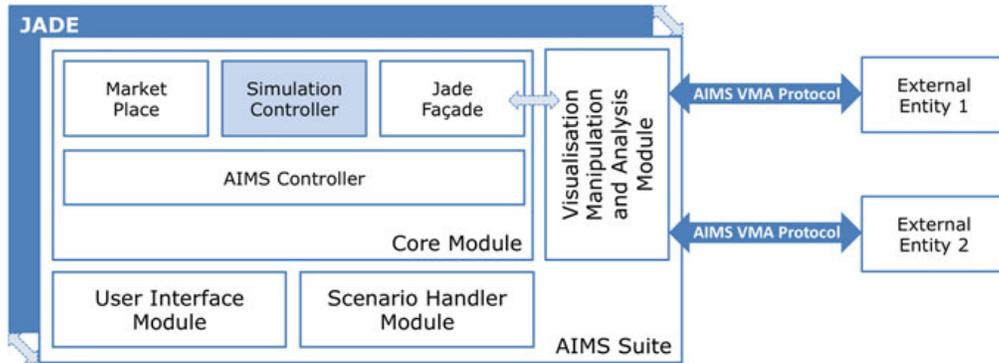
Fig. 11.6 AIMS modelling and simulation framework (after [49]). The agent-based CI simulation is executed in JADE (Java Agent Development Framework)

11.4.2 Frameworks for Modelling and Simulation of CI Using Dedicated Simulations

Bagheri, Ghorbani et al. developed the Agent-based Interdependency Modelling and Simulation **(AIMS)** suite [48, 49] especially for investigating (inter)dependencies between CI. Instead of employing off-the-shelf simulators, the authors used their own agent-based simulation, the AIMS simulator. Figure 11.6 shows the modules of the AIMS suite. They form a federation in the sense of our Fig. 11.1, but with only one dedicated agent-based simulator. For Bagheri et al., the analysis and visualisation part of MS&A were of particular importance, both on-line and off-line. They created a special middleware, the VMA (Visualisation, Manipulation and Analysis) entity protocol that allows integrating external entities, that is, analysis software modules, advanced visualisation tools, and scenario editors. In the on-line case, the latter entities enable users of the AIMS suite changing the models or scenarios while the simulation is running. All external entities need to conform to the VMA entity protocol in order to be integrated into the federation.

In [49], the authors describe their approach in detail. Also, they report the results of a case study of an electronic service provider that is dependent on two electricity suppliers and an Internet service provider (ISP). They use four scenarios in which they dynamically change the quality or availability of services of the electricity suppliers and the ISP in order to find out about the economical impact upon the electronic service provider. Finally, it is worth mentioning that Bagheri et al. are one of the few author teams that describe the workflow for setting up a federation (an instance of the AIMS suite, in this case). Since setting up federations requires considerable time and special know-how, this is valuable information.

I2Sim [19] provides a framework for discrete abstract modelling of dependent infrastructures from scratch. The I2Sim modelling ontology consists of 'cells', 'channels', 'tokens' and 'controls'. A cell is a functional or production unit, like a hospital, a power station etc, that requires some input and produces some output. The behaviour of an infrastructure is described as human readable table (HRT), created

by experts (typically, the operators of the infrastructure). For an electrical power infrastructure, such a HRT specifies the available output power for a certain number of infrastructure operating states. Each operating state of the infrastructure—or the cell representing it—is described by a number of input-output relations.

Channels connect outputs of one (source) cell to inputs of other (consumption) cells and transport tokens from the source to the production cell. A channel may have a loss coefficient that characterises the percentage of tokens that get lost during transport (like loss of voltage due to resistance of a power line). After specifying all cells and their input-output relations, all channels and their loss coefficients, I2Sim sets up a mathematical model of the entire system. For the simulation part, I2Sim performs time-driven discrete event simulation [50]. Events occur at each time step and trigger the recalculation of the coefficients describing all cells' operating status. At each time step, the simulated system is described by a set discrete time equations [50], represented by a 'system transportation matrix'. This representation allows identifying strong dependencies or critical vulnerabilities. As a test case study, the I2Sim team modelled and simulated the campus of the University of British Columbia at Vancouver, which has the size of a small city. The model took into account buildings and water, gas, power and road networks of the campus. The test case system performed damage assessment in case of disaster, expressed in the number of casualties, economical losses and loss of campus functions. And finally, the test case system provided advanced decision support capabilities, including what-if analyses for first responders.

The I2Sim approach has at least three advantages: It allows modelling at different levels of abstraction, it preserves the privacy of the contributing infrastructures by not requiring revealing lots of technical detail, and it simultaneously reduces the required infrastructure domain expertise of the modelling experts. Small drawbacks are that the fidelity of the modelling is limited and that the infrastructure behaviour description cannot be validated by the modelling experts: They need to trust the domain experts.

To conclude this section, we want to mention that Tolone et al. [15, 51, 52] used their Integrated Modeling Environment (**IME**) framework for creating mixed federations of their own special purpose simulations and external simulators. IME allows both that developers of federated simulations do the entire modelling and simulation from scratch and that they use existing simulators for the federation.

11.4.3 More General Frameworks and Middleware for Modelling and Federated Simulation

OpenMI [30, 31], the Open Modelling Interface, is a context based request-reply architecture that defines an interface allowing time-dependent models to exchange data at runtime. A recent application example is described in [53]. Data exchange between models to be linked only takes place if the models are OpenMI-compliant.

The quantities that are to be exchanged must be identified and matched. The models can then be linked at runtime. The very generic character of OpenMI leads to similar restrictions (complexity overhead) as in the case of HLA (cf. below). The development of OpenMI originated from the field of water related research and is promoted by the OpenMI association [31].

IDSim [54], the Interoperable Distributed Simulation framework, is a middleware for federated simulation that has been designed for distributed federated simulation. Following one of its initial design requirements, IDSim uses standard open technologies. IDSim's communication middleware is built on the open standard OGSI, the Open Grid Service Infrastructure, and abstract simulation models are represented in XMSF-based documents (XMSF: Extensible Modelling and Simulation Framework, [55, 56]). IDSim has been employed by the US-American facility NISAC (National Infrastructures Simulation and Analysis Center [57]) for demonstrating that it is feasible to integrate a federation of distributed simulation and a federation of distributed collaboration in the homeland security domain [58]. Within this federation of two federations, IDSim employs the HLA approach to federate the BioDAC simulation environment with the agent-based N-ABLE environment into a single simulation platform [58]. The IDSim software architecture is depicted in Fig. 11.7. In this architecture, IDSim clients enable the federate simulations communicating via OGSI. All communication between federates is routed through a central IDSim server that functions as orchestrator of the federation. The IDSim server holds status information of the federation and provides also the services for distributed simulation [54]. Data that needs to be recorded while the distributed simulation is running is logged by a storage service, while simulation models and configuration parameters are kept in XML repositories using XMSF-based syntax. Once again, this architecture matches the generic federation scheme that we depicted earlier in Fig. 11.1.

The Extensible Modeling and Simulation Framework (**XMSF**) [55, 56] has been developed with the goal of running HLA compliant simulators in a distributed fashion over the Internet and make them interoperable with other components needed for a federation (see Fig. 11.1). For this purpose, XMSF allows adding web services to HLA compliant simulators.

Besides IDSim, the **WSIM** (Web Services Internet Management) architecture [59] is another example that makes use of XMSF. The authors address the need for a sophisticated **interest management** [60], a concept that has been developed for reducing the amount of data that has to be transmitted between federates in order to optimise performance. The basic idea is to transmit only the data needed by a certain federate and only at certain points in time. WSIM extends that concept by using aggregated information and role based access control to clients within the federation. Figure 11.8 shows the scheme of the top-level architecture of a federation using WSIM. The development of XMSF and WSIM has been driven by the military domain, while IDSim has been developed for homeland security.

Other frameworks include **ASimJava** [61, 62], a Java based framework for federated and distributed simulation of large-scale physical systems as well as the aforementioned **IME** of Tolone et al. [15, 51, 52].
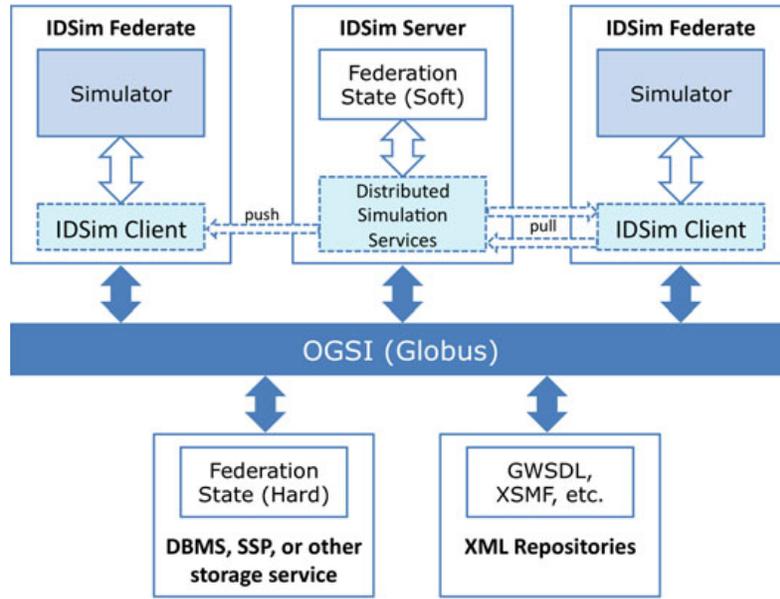
**Fig. 11.7** Federation architecture of IDSim, using an IDSim server as orchestrator of the federation (after [54])
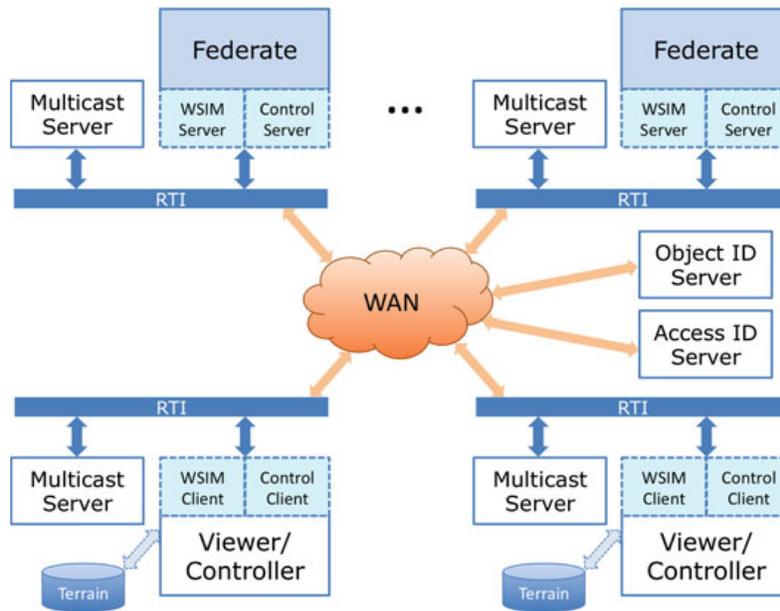
Fig. 11.8 Scheme of the distributed federation architecture WSIM using the XMSF framework (after [59])

11.4.4 The HLA Standard

In the Defence industry, coupling of simulators has been done over a long time and for various reasons. This paragraph describes the evolvement of this need and how it grew into its current standard called HLA [26]. It gives in short the benefits of the HLA approach, how it works and how the Critical Infrastructure Protection (CIP) community can benefit from these lessons.

In the early 1980s, training and education with the use of interactive simulators in virtual environments was very expensive. At the United States Defense Advance Research Projects Agency (DARPA) they realised that there was a need for multi-user simulation for real-time combat training. In order to achieve this in the most cost effective way, they came up with the idea to link single-user simulators in a multi-user environment. To prove this idea they developed the Simulator Network (SIMNET) as a wide area network of vehicle simulators like tanks, airplanes and helicopters together with computer generated forces (CGF). Based on the successes of SIMNET, a standard was derived for linking interactive simulators. This standard, called Distributed Interactive Simulation (DIS) is defined under IEEE standard 1278 in 1993 [41] and has had several revisions since. It is still used and a new version (version 7) is currently (2013) produced.

The DIS protocol is based on the principle that all participating simulators can act as a stand-alone simulator. Therefore all simulators keep all information necessary to create the (static part of the) virtual world. In order for the simulators to be able to interact with each other, every simulator sends the absolute truth about the (externally observable) state of the object it represents to all participants (broadcast). Every receiver has to decide for itself whether it is affected by these transmissions. For example, an airplane broadcasts its location. A radar receives this location and the internal algorithms determine whether this airplane is visible to the radar. The same holds for interactions like fire and detonations. The DIS standard contains information about update rates and dead reckoning algorithms. This allows simulators to join and resign the exercise without interrupting the others, and to lower the bandwidth consumption.

The the big advantage of the DIS standard is that the link is defined in a network protocol. This makes it easy to link simulators that comply to the standard. The downfall is that it is rigid: only the information contained in the standard is exchanged. Also it is limited to real time simulation and its broadcasting technique makes it network intensive. To overcome this, the United States Department of Defence (DoD) started the development of the High Level Architecture (HLA) in the late 1990s. HLA is defined under IEEE Standard 1516 in 2000 [26] and in 2010 revised as HLA evolved. HLA enables computer simulations to interact (that is, to communicate data, and to synchronise actions) with other computer simulations. The interaction between simulations is managed by a Run-Time Infrastructure (RTI) (Fig. 11.9).

Simulations used in HLA can be mathematical, rule-based, etc. and can be with or without human in the loop. If a simulator implementation is HLA-compliant, it is called a federate. HLA simulations, made up of federates, are called federations.
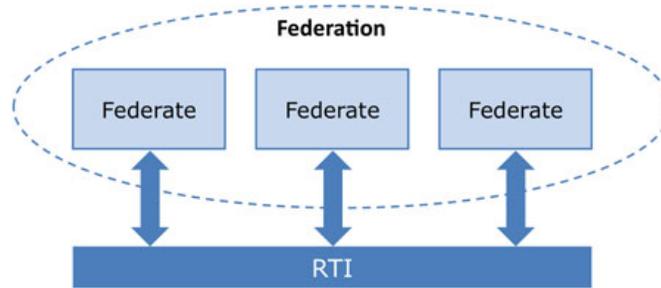
Fig. 11.9 Federation of simulators and Run-Time Infrastructure (RTI) middleware

Objects and interactions that are exchanged between federates in a federation need to be defined in a document. In HLA this is called the Object Model Template (OMT).

In the next paragraphs, we will present the HLA components: *HLA Rules*, *Interface Specification*, and *Object Model Template*.

11.4.4.1 The Rules of HLA

The core of HLA consists of a set of ten HLA rules which a federate or a federation must observe to be HLA-compliant [27]. The HLA rules are divided into two groups, five rules for HLA federations and five rules for HLA federates. The federation rules are aimed to create a federation, they include the following concepts:

- Documentation requirements: federations shall have a Federation Object Model FOM, documented in accordance with the HLA OMT.
- Object representation: all representation of objects in the FOM shall be in the federates, not in the RTI (Run Time Interface).
- Data interchange: during a federation execution, all exchange of FOM data among federates shall occur via the RTI.
- Interfacing requirements: during a federation execution, federates shall interact with the RTI in accordance with the HLA interface specification.
- Attribute ownership: during a federation execution, an instance attribute shall be owned by at most one federate at any given time.

The federate rules deal with the individual federates, they cover:

- Documentation: federates shall have a SOM (Simulation Object Model), documented in accordance with the HLA OMT.
- Control of object attributes: federates shall be able to update and/or reflect any in-stance attributes asw well as to send and/or receive interactions, as specified in their SOMs.
- Owner of object attributes: federates shall be able to transfer and/or accept ownership of attributes dynamically during a federation execution, as specified in their SOMs.

- Transfer of object attributes: federates shall be able to vary the conditions under which they provide updates of instance attributes, as specified in their SOMs.
- Time-management: federates shall be able to manage local time in a way that will allow them to coordinate data exchange with other members of a federation.

11.4.4.2 The Runtime Interface

The functional interfaces between federates and the runtime infrastructure (RTI) are defined by the interface specification. It has been adopted as IEEE standard (P1516.1). The RTI is not a part of the specification, but it is a software that matches the specification. In fact RTI provides the software services necessary for supporting an HLA-compliant simulation. There are different versions of RTI. The interface specification identifies not only the way federates will interoperate with the federation but also one with each other. The Run Time Infrastructure includes:

- Software providing common services to simulation systems
- Implementation of the federate initiated services in accordance with the HLA Interface Specification
- An architectural foundation encouraging portability and interoperability.

11.4.4.3 The Object Model Template

To achieve reusability and interoperability it is required that all objects and inter-actions, managed by a federate, are specified in detail with a common format. For this reason the Object Model Template (OMT) provides a standard to document the HLA Object Model information. In OMT three Object Models are defined:

- **The Federation Object Model (FOM)**: Every federation has only one FOM that introduces all shared information (e.g., objects, interactions). The FOM contem-plates inter-federate issues (e.g., data encoding schemes).
- **The Simulation (or Federate) Object Model (SOM)**: Every federate has one SOM, hence a federation can have several. A SOM describes salient characteristics of a federate and presents objects and interactions that can be used externally. The SOM focuses on the federate's internal operation.
- **The Management Object Model (MOM)**: The MOM identifies objects and interactions used to manage a federation.

Although the HLA standard originally was developed by the defence organisa-tions, it gives a general approach to connecting simulators of different fidelity and function. Therefore it's use can be much broader. The CIP community can have ben-efit from HLA as it is an open standard which provides all ingredients for linking simulators from different domains. Several commercial and non-commercial RTIs are available.

11.4.5 The DIESIS Approach to Semantic Interoperability

The EU funded project DIESIS (Design of an Interoperable European Federated Simulation network for Critical InfrastructureS) was a design study that assessed the technological, economical and organisational feasibility for European Infrastructures Simulation and Analysis Centre (EISAC) [16]. The distributed facility shall later be used by researchers, national security agencies and CI stakeholders in order to perform modelling, simulation and analysis for investigating a wide range of aspects of national and European CIs [24]. A prerequisite for establishing such a facility is the existence of flexible concepts for coupling heterogeneous simulation systems and their models. The DIESIS approach consists of an ICT architecture, a middleware layer for federating simulators and tools, a communication middleware for connecting distributed simulators and an ontology-based approach for achieving semantic interoperability (for detailed description of particular aspects see [24, 63, 64]).

Unlike HLA, the interoperability approach in DIESIS has been developed to fulfil specific requirements stated by the concept of EISAC: the ability to execute federated simulations of arbitrary (often large and complex) interconnected CIs in order to analyse their interplay, to identify dangerous situations and to assess risks under consideration of cross-domain dependencies. The proposed approach had to be able to get along with different time models and incompatible interface technologies provided by commercial closed-source simulators. Despite of its technological heterogeneity, the approach had to define a common superior modelling perspective that would allow to describe relations beyond the "world" of a single infrastructure. These requirements led to two basic concepts of DIESIS: lateral coupling of federates and separation of technical and semantic interoperability layers.

There are only few off-the-shelf CI simulators that support established interoperability standards like HLA. Experience shows that an attempt to find a common practicable interoperability solution for a set of CI simulators is often doomed to fail for several reasons. Firstly, the development of some specific features may require an enormous effort. Secondly, some particular couplings may be inefficient and significantly slow down the federated simulation. Finally, a desired global solution may not exist at all. Obviously, the increasing complexity of cross-domain dependencies and the growing number of different federates make the existence of a practicable solution even less probable. For this reason, DIESIS proposes a new interoperability approach that abandons the idea of a generic homogeneous architecture that uses a single RTI like HLA (see Sect. 11.4.4).

The proposed concept of **lateral simulator coupling** stands for the development of dedicated coupling links if and only if data exchange between the corresponding federates is required for the current analysis task (see Fig. 11.3). In such federation, pairwise couplings may coexist with clusters of centrally coupled simulators. The systematic development of coupling links implies a *scenario-oriented* federation design. This means that the specification of links is based on the knowledge about the involved domains and their interdependencies, about the runtime behaviour of particular federates as well as about the intended simulation output. In other words,

for realisation of coupling links, the connectivity of the federation has to be described both at technical and at semantic levels (see Sect. 11.3.1) by means of appropriate formalisms.

On the semantic level, the DIESIS approach employs an ontology-based representation to describe infrastructures, general dependencies, infrastructure elements and their relations [63]. DIESIS Knowledge Base System (KBS) uses Ontology Web Language (OWL) and Semantic Web Rule Language (SWRL) to define a *scenario* at three semantic layers [24]:

- **World Ontology** (WONT) is a template that provides basic logical concepts for describing infrastructures as well as their possible behaviours and interdependencies.
- **Infrastructure Ontology** (IONT) is based on the WONT template and describes one particular CI with its domain-specific properties and concrete elements (individuals). The KBS contains one IONT for each infrastructure represented in the federation. A IONT does not necessarily completely duplicate the underlying simulator model with all its facets. It is sufficient to model elements and relations that are involved in cross-domain activities.
- **Federation Ontology** (FONT) is dedicated to the modelling of dependencies among particular CIs. The FONT includes all IONTs and supports dependency modelling at general level (e.g., "a base station receives electric power from a power node") as well as in relation to concrete instances. The *FONT rules* (written in SWRL) express the dependency semantics (for example, "a base station is off if it gets no electric power") for particular relations.

The ontology-based model captures all facets of interplay among the infrastructures independently from the implementation of simulators and coupling links. However, its role is not limited to providing a guideline for link realisation. Data from KBS can be also used by the links at runtime for routing (i.e., sending internal state changes to the right federates according to dependency relations) as well as for automatic data transformation and filtering.

As already mentioned in Sect. 11.3.1, a problem of lateral coupling approach is the potentially large number of links that has to be developed for creating a federation as well as for adding a new member to an existing one. A possible solution is to implement similar links only once and to reuse them if possible. Creating lightweight links for particular tasks instead of complex "all-in-one" couplings significantly increases the probability that a resulting link can be reused for another pair of federates. The DIESIS architectural approach recommends the following four links types:

- **Time links** allow simulators to synchronise their internal clocks and to ensure the correct ordering of processed and sent events. A central synchronisation mechanism is possible but not required. Theoretically, a federation may contain clusters that internally use both conservative and optimistic synchronisation algorithms (see Sect. 11.3.2).
- **Data links** are used by simulators for exchanging their state changes and simulation results. Besides of individual implementations, it is possible to develop
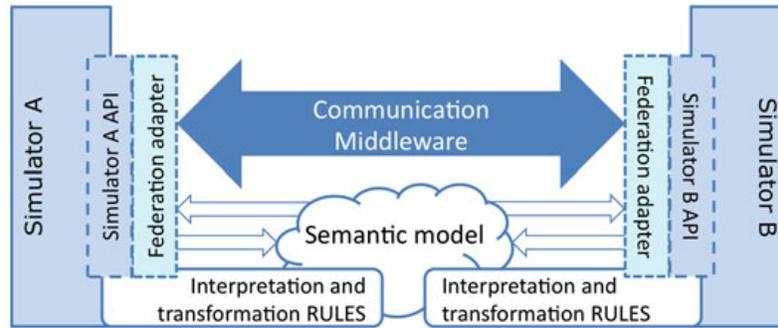
Fig. 11.10 DIESIS approach: simulator interoperability on semantic (*white blocks*) and on technical levels (federation adapters and communication middleware)

a common data routing, transformation and filtering algorithm that uses the dependency information from the KBS. Figure 11.10 shows the structure of such an "intelligent" data link (as it was created for the DIESIS demonstrator) that transforms data acquired from the simulator API according to KBS relations and rules before sending it to another federate.

- **Function links** serve the purpose of mutual invocation of function calls among federates. This may be used by a simulator, for example, to exploit the computational procedures of other simulators.
- **Control links** are employed to manage the runtime behaviour of federates, like starting, stopping or reconfiguring the simulators.

The realisation of links is considered as three-step process and follows the idea of clear separation of technical and semantic interoperability. In the first step, basic logical relationships among CIs have to be defined. In parallel to the formal ontology-based representation, this information can be visualised by means of *service networks*. A service network consists of *agents* that are connected by labelled directed *service links* (for more detailed definition and examples see [65]). In the second step, the required technological extensions (both agents and service links) have to be identified and added to the service network. Possible examples for such extensions are: central simulation control panel with corresponding control links, time management agents with their time links, visualisation and analysis modules, etc. In the final realisation step, the missing service network links and agents have to be implemented and deployed.

The realisation of a sufficiently complex demonstrator (four domains represented by simulators that do not support any common standards) showed the effectiveness and flexibility of the DIESIS approach. It turned out that the usage of a KBS in combination with a lateral coupling approach is a general advantage that significantly reduces the implementation effort.

11.5 Conclusion

Federated modelling, simulation and analysis is an invaluable method with various applications in the defence and civil security domains. It is a premier means of research and development of methods that aid in improving the resilience and protection of Critical Infrastructures, and, at the same time, a research subject of its own. Federated MS&A can be employed as exercise and training environment for crisis and emergency managers, and may serve as part of a decision support system for exploring different courses of action in case of a crisis or emergency. And finally, it can be used for testing and benchmarking new methods for CIP. The case studies and demonstrators reported in the state of the art literature expose impressive new capabilities that have a clear benefit for civil security and thus for society as a whole. The role model is the United States' NISAC that operates under congress mandate for ten years now. In Europe, capabilities like NISAC's are still missing.

A key enabling feature for federated MS&A is the interoperability of federates. They need to exchange data in a syntactically and semantically correct way, and at the correct points in simulation time. The latter requirement cannot be fulfilled for all possible combinations of possible different time models. Our review of the state of the art exposed that independent groups developed similar technical solutions for making federates interoperable, despite the existence of the interoperability standard HLA. As a matter of fact, HLA imposes strong requirements on HLA-compliant federate simulators, and the implementation of the Run-Time Infrastructure, a key interoperability technology, is not part of the standard. We conclude that these obstacles slow down the adoption of HLA in its original domain, defence. OpenMI is a well established M&S standard in domains related to water.

The situation is worse when it comes to interoperability of simulators for CIP. There are numerous simulators for several different infrastructures: railway simulators, electrical power network simulators, telecommunication network simulators and so on. However, almost all of these simulators have been designed for their domains only, not for becoming part of a federation. Some of them even do not have APIs. Although interoperability standards are desirable also for applications of federated MS&A for CIP, it is not likely that the makers of commercial simulators make investments into making their products compliant with some interoperability standard, as long as a convincing business model or a significant market for such an enhanced product is missing. As one solution to this problem, the DIESIS project suggests that the CIP research community joins resources and creates a repository of reusable interoperability solutions for CI simulators. Other groups avoided the problem by creating integrated simulators that cover several infrastructure domains. However, this approach is limited to a certain level of modelling abstraction. Whenever high-fidelity simulations are required, special purpose simulators are superior. Their integration into a federation then requires suitable interoperability middleware.

A second obstacle for a more wide-spread usage of federated MS&A is the fact that setting up federations is a time-consuming task that requires multi-disciplinary expertise. We appreciate that a few researchers have documented their expertise in

setting up federations and have created descriptions of the workflow for this task [49, 66]. However, standardising such workflows and training researchers in modelling and setting up federations adhering to such workflows should be considered as an aid to capacity building in European CIP research. This would be a first step to fill a security gap, as pointed out in [67]: Given the complexity of European CI systems, Europe would urgently need MS&A capabilities comparable to those in the USA.

References

1. Luiijf, H., Klaver, M.: International interdependency of C(I)IP in Europe (Internationale Verflechtung von C(I)IP in Europa). In: Proc. CIP Europe 2005—Critical Infrastructure Protection, GI CIS Forum, Bonn, Germany. (2005)
2. EC: Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection. OJEU, European Commission (2008)
3. Nieuwenhuijs, A., Luiijf, H., Klaver, M.: Modeling critical infrastructure dependencies. In Mauricio, P., Shenoi, S., eds.: Critical Infrastructure Protection II. Volume 290 of IFIP., Boston, MA, USA, Springer (2008) 205–214
4. Setola, R., Bologna, S., Casalicchio, E., Masucci, V.: An integrated approach for simulating interdependencies. In Papa, M., Shenoi, S., eds.: Critical Infrastructure Protection II. Volume 290 of The International Federation for Information Processing. Springer US (2009) 229–239
5. Luiijf, H., Nieuwenhuijs, A.H., Klaver, M.H., Eeten, M.J.V., Cruz, E.: Empirical findings on European critical infrastructure dependencies. Int. J. of System of, Systems Engineering **2**(1), (2010) 3–18
6. Pederson, P., Dudenhoeffer, D., Hartley, S., Permann, M.: Critical infrastructure interdependency modeling: A survey of U.S. and international research. Technical Report INL/EXT-06-11464, Idaho National Laboratory (August 2006)
7. Dudenhoeffer, D., Permann, M., Manic, M.: Cims: A framework for infrastructure interdependency modeling and analysis. In: Simulation Conference, 2006. WSC 06. Proc. Winter. (Dec. 2006) 478–485
8. Min, H., Beyeler, W., Brown, T., Son, Y., Jones, A.: Toward modeling and simulation of critical national infrastructure interdependencies. IIE Transactions **39**(1) (2007) 57–71
9. Laprie, J.C., Kanoun, K., Kaâniche, M.: Modelling interdependencies between the electricity and information infrastructures. In Saglietti, F., Oster, N., eds.: SAFECOMP 2007. Volume 4680 of LNCS. Springer, Berlin Heidelberg (2007) 54–67
10. Casalicchio, E., Galli, E., Tucci, S.: Federated agent-based modeling and simulation approach to study interdependencies in IT Critical Infrastructures. In: Distributed Simulation and Real-Time Applications, 2007. DS-RT 2007. 11th IEEE International, Symposium. (Oct. 2007) 182–189
11. Svendsen, N.K., Wolthusen, S.D.: Connectivity models of interdependency in mixed-type critical infrastructure networks. Inf. Secur. Tech. Rep. **12**(1) (March 2007) 44–55

12. Rosato, V., Issacharoff, L., Tiriticco, F., Meloni, S., Porcellinis, S.D., Setola, R.: Modelling interdependent infrastructures using interacting dynamical models. Int. J. of Critical Infrastructures **4**(1/2) (2008) 63–79
13. Rinaldi, S., Peerenboom, J., Kelly, T.: Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies. IEEE Control System Magazine **December** (2001) 11–25
14. Rinaldi, S.M.: Modeling and simulating critical infrastructures and their interdependencies. In: Proc. 37th Annual Hawaii Int. Conf. System Sciences (HICSS'04)—Volume 2, Washington, DC, USA, IEEE Computer Society (2004) 20054.1
15. Tolone, W.J., Wilson, D., Raja, A., Xiang, W.N., Hao, H., Phelps, S., Johnson, E.W.: Critical infrastructure integration modeling and simulation. In Chen, H., Moore, R., Zeng, E., Leavitt, J., eds.: Proc. 2nd Symposium on Intelligence and Security Informatics (ISI-2004). Volume 3073 of LNCS., Springer-Verlag (2004) 214–225
16. Rome, E., Bologna, S., Gelenbe, E., Luiijf, E., Masucci, V.: DIESIS—design of an interoperable European federated simulation network for critical infrastructures. In: Proceedings of the 2009 SISO European Simulation Interoperability Workshop (EURO SIW '09), San Diego, CA, USA, Simulation Councils, Inc. (2009) 139–146
17. Bloomfield, R., Chozos, N., Nobles, P.: Infrastructure interdependency analysis: Requirements, capabilities and strategy. Technical Report D/418/12101/3, Adelard LLP, London, UK (2009)
18. Luiijf, H., Klaver, M.: Critical infrastructure awareness required by civil emergency planning. In: Proc. 1st IEEE International Workshop on Critical Infrastructure Protection (IWCIP '05), Washington, DC, USA, IEEE Computer Society (2005) 110–118
19. Martí, J., Ventura, C., Hollman, J., Srivastava, K., Juárez, H.: I2Sim Modelling and Simulation Framework for Scenario Development, Training, and Real-Time Decision Support of Multiple Interdependent Critical Infrastructures during Large Emergencies. In: How is Modelling and Simulation Meeting the Defence Challenges out to 2015? Volume RTO-MP-MSG-060., NATO RTO Modelling and Simulation Group Conf., Vancouver, BC, Canada (October 2008) 16.1–16.14
20. Klein, R., Rome, E., Beyel, C., Linnemann, R., Reinhardt, W., Usov, A.: Information modelling and simulation in large interdependent critical infrastructures. In Setola, R.E.e.a., ed.: Proc. 3rd International Workshop on Critical Information Infrastructures Security (CRITIS '08). Volume 5508 of LNAI., Berlin, Springer-Verlag (2009) 36–47
21. Usov, A., Beyel, C.: Simulating interdependent critical infrastructures with SimCIP. European CIIP Newsletter **4**(3) (November/December 2008) 6–8
22. Balducelli, C., Pietro, A.D., Lavalle, L., Vicoli, G.: A middleware improved technology (MIT) to mitigate interdependencies between critical infrastructures. In de Lemos, R., Giandomenico, F., Gacek, C., Muccini, H., Vieira, M., eds.: Architecting Dependable Systems V. Volume 5135 of LNCS. Springer, Berlin / Heidelberg (2008)
23. Tofani, A., Castorini, E., Palazzari, P., Usov, A., Beyel, C., Rome, E., Servillo, P.: Using ontologies for the federated simulation of critical infrastructures. Procedia Computer Science **1**(1) (2010) 2301–2309
24. Usov, A., Beyel, C., Rome, E., Beyer, U., Castorini, E., Palazzari, P., Tofani, A.: The DIESIS approach to semantically interoperable federated critical infrastructure simulation. In: Advances in System Simulation (SIMUL), 2010 Second International Conference on. (August 2010) 121–128
25. Adinolfi, F., Monica, M.D., Masucci, V., Olivadoti, S., Servillo, P., Spizuocol, C., et al.: DIESIS Deliverable D2.2: Final technology analysis and assessment. Technical report, CRIAI (2009)
26. IEEE: IEEE 1516–2000: High level architecture (2000)
27. IEEE: IEEE standard for modeling and simulation (M&S) high level architecture (HLA)—framework and rules. Technical report, IEEE (2000)
28. IEEE: IEEE 1516–2000: High level architecture—framework and rules (2000)
29. IEEE: IEEE 1516–2000: High level architecture—federate interface specification (2000)
30. Gregersen, J.B., Gijsbers, P.J.A., Westen, S.J.P.: OpenMI: Open modelling interface. J. of Hydroinformatics **9**(3) (2007) 175–191
31. OpenMI Association: http://www.openmi.org last accessed 2013-01-16

32. Fujimoto, R.M.: Parallel simulation: parallel and distributed simulation systems. In: Proceedings of the 33nd conference on Winter simulation. WSC '01, Washington, DC, USA, IEEE Computer Society (2001) 147–157

33. Chandy, K., Misra, J.: Distributed simulation: A case study in design and verification of distributed programs. Software Engineering, IEEE Transactions on **SE-5**(5) (sept. 1979) 440–452

34. Riley, G.F., Ammar, M.H., Fujimoto, R.M., Park, A., Perumalla, K., Xu, D.: A federated approach to distributed network simulation. ACM Trans. Modeling and Computer Simulation **14**(2) (April 2004) 116–148

35. Mattern, F.: Efficient algorithms for distributed snapshots and global virtual time approximation. Journal of Parallel and Distributed Computing **18**(4) (1993)

36. Fujimoto, R., McLean, T., Perumalla, K., Tacic, I.: Design of high performance rti software. In: Distributed Simulation and Real-Time Applications, 2000. (DS-RT 2000). Proceedings. Fourth IEEE International Workshop on, IEEE (2000) 89–96

37. Jefferson, D.R.: Virtual time. ACM Trans. Program. Lang. Syst. **7**(3) (July 1985) 404–425

38. Sokol, L.M., Stucky, B.K.: MTW: Experimental results for a constrained optimistic scheduling paradigm. In Nicol, D., ed.: Distributed Simulation. Volume 22 of Simulation. Society for Computer Simulation (SCS), San Diego, CA (January 1990) 169–173

39. Steinman, J.S.: Breathing time warp. In: Proceedings of the seventh workshop on Parallel and distributed simulation. PADS '93, New York, NY, USA, ACM (1993) 109–118

40. Dickens, P., Reynolds, P.: SRADS with local rollback. Institute for Parallel Computation, School of Engineering and Applied Science, University of Virginia (1990)

41. IEEE: IEEE 1278–1993—standard for distributed interactive simulation (1993)

42. HLA-OMT: High-level architecture object model template specification version 1.3 (5 February 1998)

43. Masucci, V., Adinolfi, F., Servillo, P., Dipoppa, G., Tofani, A.: Ontology-based critical infrastructure modeling and simulation. In Palmer, C., Shenoi, S., eds.: Critical Infrastructure Protection III. Volume 311 of IFIP Advances in Information and Communication Technology. Springer, Berlin Heidelberg (2009) 229–242

44. SISO: SISO generic methodology for verification and validation (GM-VV) to support acceptance of models, simulations and data (2012)

45. Bagheri, E., Ghorbani, A.A.: The state of the art in critical infrastructure protection: a framework for convergence. Int. J. of Critical Infrastructures **4**(3) (2008) 215–244

46. Hopkinson, K., Wang, X., Giovanini, R., Thorp, J., Birman, K., Coury, D.: EPOCHS: A platform for agent-based electric power and communication simulation built from commercial off-the-shelf components. IEEE Trans. on Power Systems **21**(2) (May 2006) 548–559

47. NS2: The Network Simulator: http://www.isi.edu/nsnam/ns last accessed 2013-01-16

48. Bagheri, E., Ghorbani, A.A.: A service oriented approach to critical infrastructure modeling. In: Workshop on Service Oriented Techniques, NRC-Canada (2006)

49. Bagheri, E., Baghi, H., Ghorbani, A.A., Yari, A.: An agent-based service-oriented simulation suite for critical infrastructure behavior analysis. Int. J. of Business Process Integration and Management **2**(4) (2007) 312–326

50. Rahman, H., Armstrong, M., Mao, D., Marti, J.: I2sim: A matrix-partition based framework for critical infrastructure interdependencies simulation. In: Electric Power Conference, 2008. EPEC 2008. IEEE Canada, IEEE (Oct. 2008) 1–8

51. Tolone, W., Lee, S.W., Xiang, W.N., Blackwell, J., Yeager, C., Schumpert, A., Johnson, W.: An integrated methodology for critical infrastructure modeling and simulation. In Papa, M., Shenoi, S., eds.: Critical Infrastructure Protection II. Volume 290 of The International Federation for Information Processing. Springer US (2009) 257–268

52. Tolone, W.J., Johnson, E.W., Lee, S.W., Xiang, W.N., Marsh, L., Yeager, C., Blackwell, J.: Enabling system of systems analysis of critical infrastructure behaviors. In Setola, R., Geretshuber, S., eds.: Critical Information Infrastructure Security. Springer-Verlag, Berlin, Heidelberg (2009) 24–35

53. Betrie, G., van Griensven, A., Mohamed, Y., Popescu, I., Mynett, A., Hummel, S.: Linking SWAT and SOBEK using open modeling interface (OpenMI) for sediment transport simulation in the blue nile river basin. Trans. of the ASABE **54**(5) (2011) 1749–1757

54. Fitzgibbons, J.B., Fujimoto, R.M., Fellig, D., Kleban, S.D., Scholand, A.J.: IDSim: An extensible framework for interoperable distributed simulation. In: IEEE International Conference on Web Services 2004 (ICWS'04), IEEE (2004) 532–539

55. Brutzman, D., Zyda, M., Pullen, J.M., Morse, K.L.: Extensible modeling and simulation framework (XMSF) challenges for web-based modeling & simulation. Findings and recommendations report: Technical challenges workshop, strategic opportunities symposium, MOVES Institute, Monterey, CA, USA (Oct. 22 2002)

56. Pullen, J.M., Brunton, R., Brutzman, D., Drake, D., Hieb, M., Morse, K.L., Tolk, A.: Using web services to integrate heterogeneous simulations in a grid environment. Future Generation Computer Systems **21**(1) (2005) 97–106

57. NISAC, National Infrastructure Simulation and Analysis Center, USA: http://www.sandia.gov/nisac Last accessed 2013-01-16

58. Linebarger, J.M., Fellig, D., Moore, P.D., Goldsby, M., Hawley, M.F., Sa, T.J.: Integrating software architectures for distributed simulations and simulation analysis communities. Technical Report SAND 2005–6642, Sandia National Laboratory (2005)

59. Morse, K., Brunton, R., Pullen, J., McAndrews, P., Tolk, A., Muguira, J.: An architecture for web-services based interest management in real time distributed simulation. In: Distributed Simulation and Real-Time Applications, 2004. DS-RT 2004. Eighth IEEE International Symposium on, IEEE (Oct. 2004) 108–115

60. Morse, K.L., Bic, L., Dillencourt, M.: Interest management in large-scale virtual environments. Presence: Teleoper. Virtual Environ. **9**(1) (February 2000) 52–68

61. Sikora, A., Niewiadomska-Szynkiewicz, E.: A federated approach to parallel and distributed simulation of complex systems. Int. J. Appl. Math. Comput. Sci. **17**(1) (March 2007) 99–106

62. Sikora, A., Niewiadomska-Szynkiewicz, E.: FR/ASimJava: a federated approach to parallel and distributed network simulation in practice. J. Telecommunications & Information Technology **2006**(4) (2006) 53–59

63. Tofani, A., Castorini, E., Palazzari, P., Usov, A., Beyel, C., Rome, E., Servillo, P.: An ontological approach to simulate critical infrastructures. Journal of Computational Science **1**(4) (2010) 221–228 Class-AB

64. Görbil, G., Gelenbe, E.: Design of a mobile agent-based adaptive communication middleware for federations of critical infrastructure simulations. In Rome, E., Bloomfield, R., eds.: Critical Information Infrastructures Security. Volume 6027 of Lecture Notes in Computer Science. Springer, Berlin Heidelberg (2010) 34–49

65. Beyer, U., Usov, A., Rome, E., Beyel, C., et al.: DIESIS Deliverable D4.1b: Final architectural design. Technical report, Fraunhofer IAIS (2009)

66. Tofani, A., Usov, A., Castorini, E., Rome, E., Görbil, G., Palazzari, P., Servillo, P., Hanckmann, P., Beyer, U.: DIESIS Deliverable D4.2a: Proof of concept. Technical report, ENEA (2009)

67. Hämmerli, B., Renda, A.: Protecting Critical Infrastructure in the EU. Technical Report CEPS Task Force Report, Centre for European Policy Studies, Brussels (March 2011)

Chapter 12
Multisystem Simulation: Analysis of Critical Infrastructures for Disaster Response

José R. Martí

Abstract National critical infrastructures (e.g., electricity, water, transportation, etc.) form large complex systems that sustain essential living functions. During large emergencies (earthquakes, tsunamis, floods, etc.) multiple critical systems suffer damage and the normal recovery processes of individual infrastructures are not sufficient to bring back the combined system of systems. Coordinated action among infrastructures is needed to make the combined system operational and save as many human lives as possible. The complexity of the combined system of systems and the uncertainties of the available data require an approach that limits the number of possible operational states and leads to robust real-time solutions. An optimum coordinated response needs to consider the interactions among the multiple layers of an effective disaster response: the physical layer of buildings, lifelines, and critical resources, the information and control layer, and the decision layer where choices are made as to the best responses. The solution framework discussed in this chapter provides a structure to capture these interactions.

Keywords Complex large scale systems · Interdependent critical infrastructures · System of dissimilar systems · Hierarchies of decisions · Virtual distributed simulation · Disaster response.

12.1 Disaster Response

Due to climatic and other geological changes in the planet, the occurrence of large disasters has been on the rise in recent years [1]. The recent Japan triple disaster (earthquake, tsunami, and nuclear plant meltdown) [2] dramatically illustrated that

J. R. Martí (✉)
Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada
e-mail: jrms@ece.ubc.ca
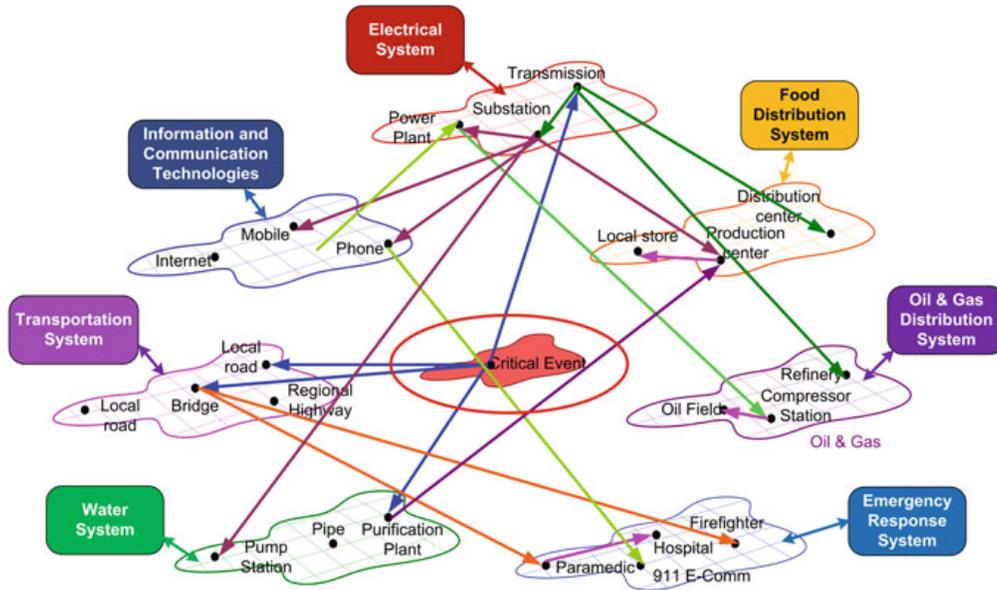
Fig. 12.1 System of systems

despite the advanced technologies available for early warning and monitoring of natural disasters, multiple critical infrastructures were severely affected and the response to the situation was far from optimal. Even though much can be done in deploying sensor systems and in preparing the population on what to do when a disaster strikes, the problem remains of understanding how multiple critical infrastructures can best coordinate the allocation of the available resources for an optimum overall response.

The problem of an effective coordination of the system response to minimize the consequences of a disaster is complex problem because of the size of the infrastructure networks that provide the vital survival resources (electricity, water, transportation, etc.) and the many interdependencies that exist among the objects and actors in the scenario. There are interdependencies at the level of the physical system (Fig. 12.1) (electricity is needed to operate water pumps, electricity and water, together with medicines, doctors, nurses are needed to operate a hospital, etc.), but also the system resources have multiple ownerships and multiple hierarchical levels (e.g., private versus government, municipal versus provincial versus federal, responders and operators versus managers, etc.). Simulation of this complex system of systems needs to include these multiple layers (Fig. 12.2).

In Canada we define ten national critical infrastructures: energy, water, food, manufacturing, finance, information and communication systems, transportation, health, safety and order, government and defence. Many of these systems have an essential role in saving human lives after a disaster occurs, and all of them play a role in the recovery of the system after the disaster.

The first minutes and hours after a disaster are critical for the overall management of the situation: injured victims have to treated within minutes or hours, the
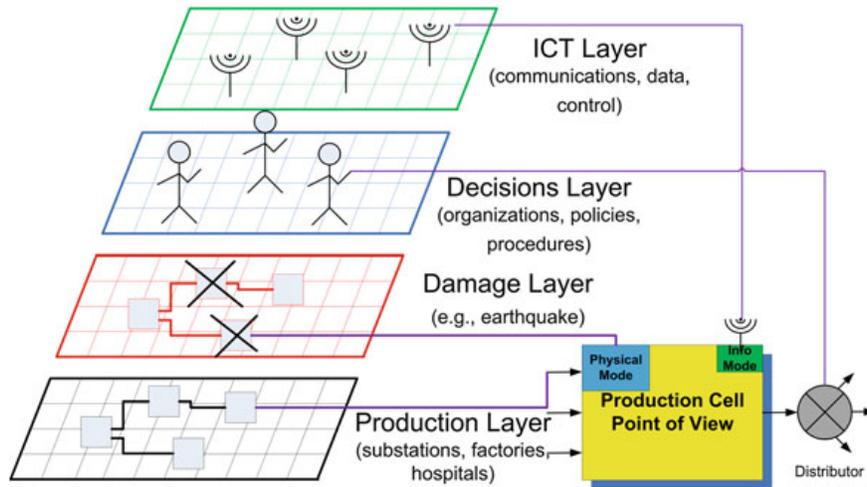
Fig. 12.2 Simulation layers

populations needs to be evacuated and a minimum level of critical services need to be restored within hours to avoid confusion and panic among the population.

There is a hierarchy and timing of needs. Critical survival needs correspond to basic human needs (e.g., Maslow's Hierarchy of Needs [3]) and the resources to provide these needs have to be available immediately after the disaster. Some of these needs include: physiological body needs (e.g., breathing, food, water, shelter) but also philia needs (e.g., whereabouts of your children, spouse, etc.). Other needs (e.g., property, financing, jobs, schools, etc.) can be restored in larger time frames.

Figure 12.3 illustrates the disaster timeline and its correlation with human needs. Even though the coordination of critical infrastructures is vital during the Response period, the Preparation period is equally or more critical. Time frames during the real time response are very short and the analysis of a large number of scenarios in Preparation time will discover system criticalities and best response strategies. A library of scenarios developed during peace time can be used as base-case scenarios for response optimization during the actual situation.

12.2 The Resources Allocation Problem

Consider the simple example illustrated in Fig. 12.4. Here we have an electric power substation supplying power to a hospital, water pumping station, and a residential area. During normal times, the 200 MW of the substation are distributed according to the needs of the served units. Assume now, that due to damage caused by a disaster on the transmission lines feeding the substation or a one o the transformers in the substation, only 100 MW are available. A decision has to be made on how to best allocate the available power. A "trivial" solution is to say that since the hospital is the most critical unit to save lives; all available power should be served to the hospital.
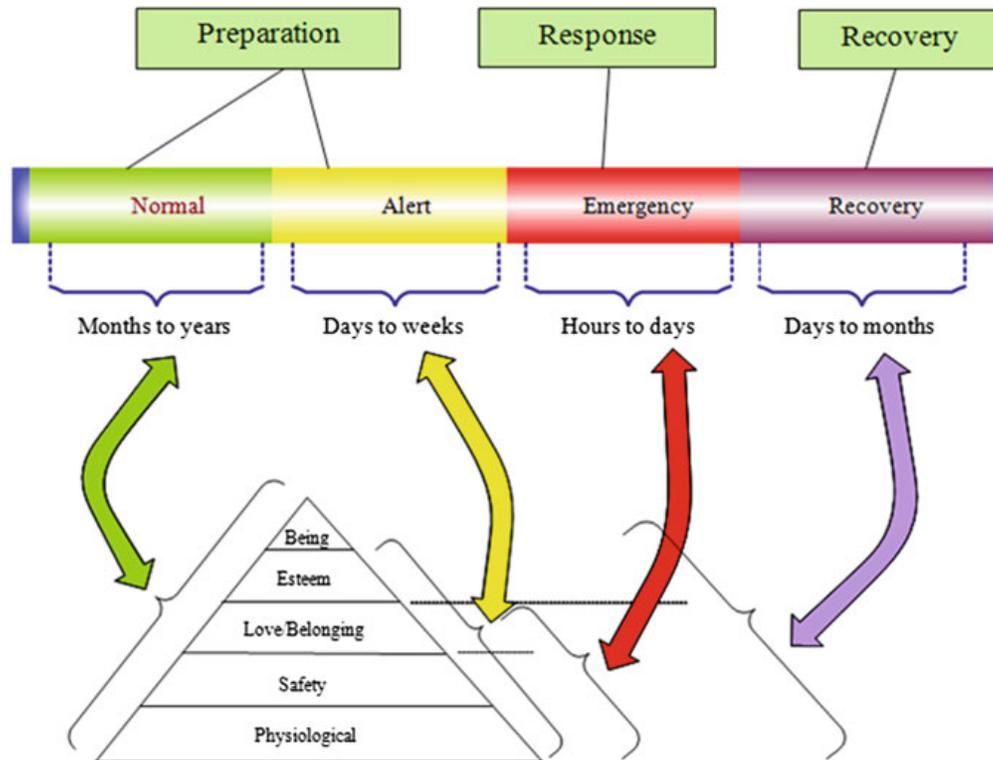
Fig. 12.3 Disaster timeline

Now the hospital will have all the power it needs. However, it ability to treat patients will be zero due to the lack of water! (This is a simplified example in which we assume that backup electricity and water reserves are not available, or have failed). The optimum solution will provide some electricity to the water station, which will depend on the level of operability (need) of the hospital.

A system of systems problem, which includes the capability of each system to produce resources under the damage caused by the disaster, and allocates these resources in an optimum manner, needs to be formulated. This is the motivation of the i2Sim solution framework discussed next.

12.3 The i2Sim Simulation Framework

As indicated in the example of Fig. 12.4, resources are produced by units operated by different infrastructures (e.g., electricity, water) and distributed by links that go from these infrastructures (electrical wires, pipes) to the units (hospital, water station) that require these resources for their operation. Figure 12.4 is a local view of distribution of resources that require extensive networks for their production and delivery. For example, the power system network can comprise tens of thousands of stations and
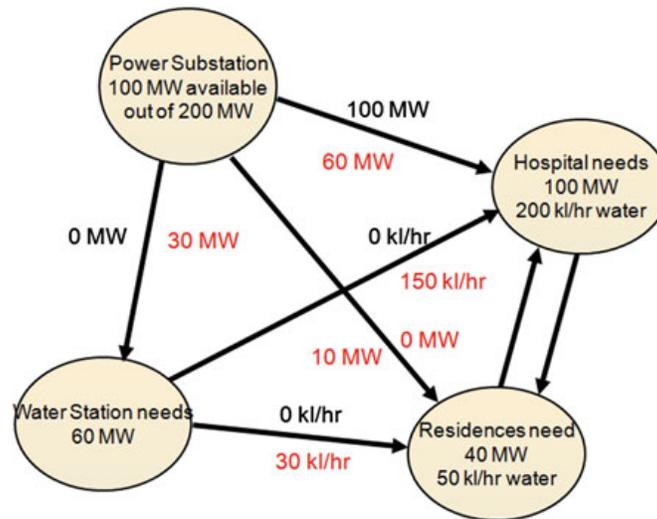
Fig. 12.4 Resources allocation problem

transmission lines to carry electricity produced in a hydroelectric dam located hundreds of kilometers away to the local substation that feeds the hospital. To simulate these large complex systems at the level of their local interactions requires a number of abstractions.

12.4 i2Sim Solution Strategy

To cope with the complexity of representing large infrastructure networks at the level of their interdependencies, the i2Sim framework [4] is designed around the following solution strategy:

1. Solve the individual infrastructure (for example the power grid) with specialized domain simulators that can represent the details of that infrastructure.
2. Find an equivalent of the infrastructure at the point where the local resource interactions are to be determined (Thévenin equivalent concept in electrical networks).
3. Find the best local allocation of resources in the area of interest (for example downtown Vancouver).
4. Notify the infrastructure of the optimum allocation of its available resources from the point of view of the local subsystem.
5. i2Sim integrates the solution of dissimilar resource networks by providing a common ontological framework and a level of granularity that matches the characteristics of the situation.
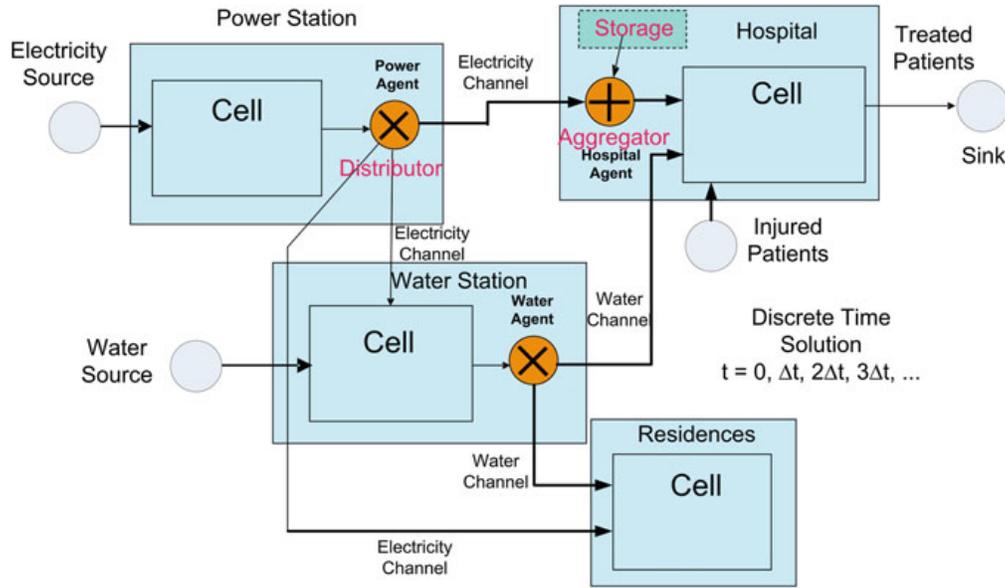
Fig. 12.5 Four-cell example

12.4.1 i2Sim Ontology

To combine multiple dissimilar subsystems, i2Sim defines a common ontology that captures the interactions among subsystems at a description level that can be understood by the different players (infrastructure owners, decision makers, and operators) and that has sufficient information to perform a mathematical optimization of best resources allocation.

The i2Sim ontology defines Base Types, Exchange Types, and Modifier Types, as described next. Figure 12.5 shows a simple example illustrating these types.

12.4.1.1 Base i2Sim Types

The base types are the citizens of the **i2Sim World.** The following base types are defined:

Tokens (Resources): Tokens are the resources needed in the system, for example, electricity, water, medicines, doctors, nurses, etc.

Cells (Production Units): A hospital cell takes input tokens: electricity, water, doctors, medicines, etc. and produces an output token: patients treated.

Channels (Transportation Units): A channel receives an output token from a cell and carries it to the input of a different cell. The electricity from the substation is transported by wires to the hospital; the water is transported by pipes, etc.

Distributors (Allocation Units): Given the available electricity output from the substation, how much of it should go to the hospital? How much should go to the water

pumping station? The Distributor interfaces the decision maker with the physical allocation of the resource (Fig. 12.2).

Aggregators (Adders): Electricity input to the hospital can come from the utility or from the backup generator (both tokens are added) (Fig. 12.5).

Reservoirs (Waiting Rooms): Arriving tokens are put into a waiting room until a request is received to let them out into the production cell.

Distributors and Aggregators constitute Control Units where decisions are made by external agents as to the best allocation of the output resources to the outgoing channels.

12.4.1.2 Exchange i2Sim Types

The exchange types communicate the **External World** with the i2Sim world. They bring in token from the external world into the i2Sim world. They also provide a mechanism for tokens to be exported out of the i2Sim world. The following exchange types are defined:

External Tokens: A Token brought in from the External World into the i2Sim World is an External Token. Once n the i2Sim world, the external token becomes an internal token and can be exchanged among internal cells.

Sources (Generators): Sources bring in External Tokens from the External World into the i2Sim World. For example, the electrical supply from the high-voltage transmission system.

Sinks (Terminators): Sinks move out Internal Tokens from the i2Sim World into the External World. For example, patients out of the hospital exit the i2Sim world.

12.4.1.3 Modifier i2Sim Types

Modifiers are external **Information Tokens** that are received as input into Cells, Channels, Distributors, and Aggregators.

Physical Mode Modifier (Physical Component Affecter): The damage caused by an earthquake or a terrorist act on a building, a piece of machinery or a lifeline will decrease the output of a cell or channel. Damage to an ICT component will prevent a control signal from exerting actions.

Human Mode Modifier (Human Affecter): The tiredness of a physician or the lack of guidance during a traffic jam will decrease the productivity of the hospital or the speed of the traffic channel.

Information Modifier (Knowledge Affecter): Lack of information of the state of the system will prevent the generation of the correct control signal to operate a Cell, Channel, or Control component. For example, if the RTU in a substation does not receive the correct control signal, the circuit breakers will not be able to reconfigure the substation to supply the priority loads.
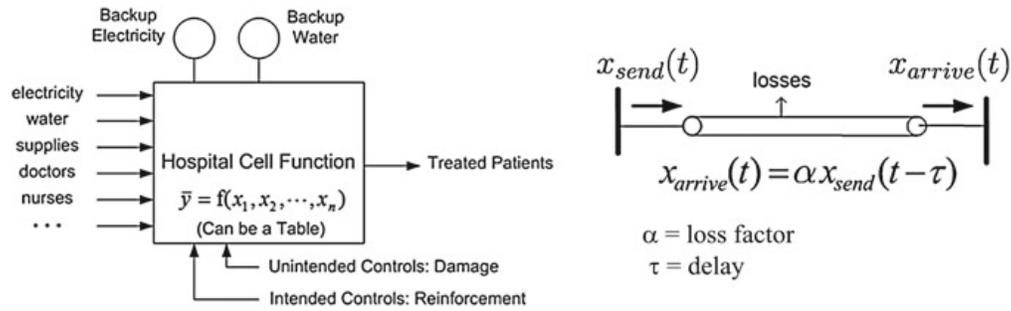
Fig. 12.6 Conceptual cell and channel models

The i2Sim world consists of Cells, Channels, Distributors, Aggregators, Tokens, Sources, Sinks, and Modifiers. The number of inputs and outputs depends on the component. Any component can have M external modifier tokens.

A Cell has N input tokens, all of different types, but only one output token. A Distributor has one input token and N output tokens, all of the same type as the input. An Aggregator has N input tokens, all of the same type, but only one output token, of the same type as the input. A Channel has one input token and one output token, both of the same type.

12.5 i2Sim Models

The conceptual models for cells and channels in i2Sim are shown in Fig. 12.6.

The cell model in Fig. 12.6 corresponds to a production unit model. This type of model was postulated by Leontief [5] to establish the relationship between inputs and outputs in a production factory. In i2Sim, this concept is extended to represent the interdependencies among infrastructure networks at the nodes where the resources produced by one network become inputs to the other networks.

The channel function is described by the main attributes of transportation mechanisms to carry resources from one node to another node which may be geographically away. These attributes are the time delay and the losses. Transportation of electricity through transmission lines happens at the speed of light and, therefore, has zero time delay in the context of the disaster time line. Similarly, water can be assumed to be available as soon as the water valve is turned on. However, leaks may occur in the water pipes because of earthquake shaking. In the case of vehicles transportation, the most important parameter is the time delay, even though losses may also exist due to accidents under damaged road conditions.

In the i2Sim realization, cells and channels are depicted as shown in Fig. 12.7, where the possible output values are discretized into five possible levels, which are colour coded. Green corresponds to the cell working at its maximum design output (rated value), yellow corresponds to the cell being able to produce only 50 % of its rated output, and red means that the cell is unable to produce any output. Blue
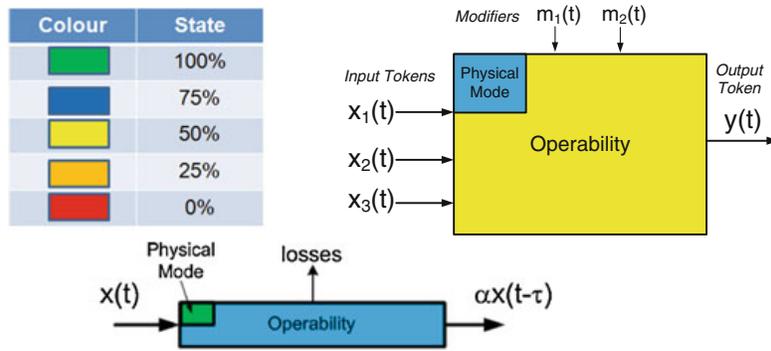
Fig. 12.7 i2Sim discretized cell and channel models

	y(t)	$x_1(t)$	$x_2(t)$	$x_3(t)$	$x_4(t)$	$m_1(t)$
Operability	Patients per hour	Electricity (kW)	Water (L/h)	Doctors	Nurses	Physical Integrity
100%	20	100	1,000	4	8	100%
75%	15	50	500	3	6	80%
50%	10	30	300	2	4	50%
25%	7	20	200	2	3	20%
0%	0	0	0	0	0	0%

Fig. 12.8 Human Readable Table (HRT) for a hospital ER unit

corresponds to 75 % and orange corresponds to 25 %. The reason for the cell not being able to produce its rated 100 % output can be the physical damage suffered by the cell during the disaster or the lack of availability of some needed input resource.

The concept of limiting the outputs to a finite number of states is essential to manage the complexity of the problem. Even though large scale utility networks such as the power grid or the telecommunications grid are usually scale-free networks and the number of possible states is statistically of polynomial order, when combined into a system of systems, the number of possible solution states can be too large for practical solutions, particularly in real time. Discretizing the number of possible states of Cells and Channels in i2Sim reduces the problem dimensionality but also it reduces the sensitivity of the solution to inaccuracies in the value of the data and in establishing the threshold levels in the HRT functions (Fig. 12.8). Even though five discrete states are defined in Fig. 12.7, in many situations where the knowledge of the information is very limited, three states (100, 50, 0 %) will still lead to useful results.

The colouring of the cell in Fig. 12.7 indicates that the cell is operating at 50 % of its normal capability while the physical state of the facilities is 75 % operational. The reason for not being able to operate at the physical capacity of 75 % is the lack

of some input resource or the effect of some modifier. The channel depicted in the figure is operating at 75 % of its normal performance even though there is no physical damage. This is due to the action of some modifier.

A key concept in i2Sim is the Human Readable Table HRT. The HRT provides the information needed to determine the operability (output level) of a cell or channel in terms of the physical damage, availability of input resources, and the effect of possible modifiers.

12.5.1 The Human Readable Table HRT

Figure 12.8 shows a Human Readable Table for the emergency unit of a hospital facility. This type of table is built from interviews with the hospital manager and operators.

When there is no physical damage and all input resources are available, the ER unit can treat 20 patients per hour. Assume now a disaster situation for which all electricity is available (either from the utility or from the backup generator) and all doctors are available. There is, however, some physical damage in the equipment that slows down the operation to 15 patients per hour. Also only six nurses are available. Under these conditions if all needed water (500 L/h) were available, the unit would be able to treat 15 patients per hour. However, only 300 L/h of water are available. This factor limits the output to only 10 patients per hour. As can be seen from this analysis, the scarcest resource limits the final operability, in this case to 50 % of the normal capacity. In this scenario, unless the water resource can be improved, only two doctors and four nurses will be needed to process 10 patients per hour and the additional doctors and nurses can be assigned to another hospital unit. In terms of resources, only 30 kW of electricity are needed and, assuming two feeders, the substation can be notified that only one of the feeders to this unit is needed.

The table of Fig. 12.8 is the "model" for the hospital unit. The inputs are x_1 to x_4 and m_1, and the output is y. Inputs x_1 to x_4 are tokens that are supplied by other cells in the i2Sim world or by external sources. Modifier m_1 is determined by the damage caused by the disaster. This information is an input provided by the Damage Layer (Fig. 12.2).

If the scenario is been run for disaster preparation purposes, the damage is calculated by some damage assessment software (for example Hazus [6]). If the scenario corresponds to a real time situation, the damage is being reported by field agents.

The HRT table allows the inclusion of human modifiers and information modifiers. The table in Fig. 12.9 shows an example of a human modifier (the doctors' shift hours).

In the HRT of Fig. 12.9, the available resources would allow operation at the 75 % level. However, the long shift of the doctors reduces their effectiveness to a lower level. Notice that an HRT may have more or fewer columns (inputs) depending on the particular scenario situation. For example, in the simulation performed for the Vancouver 2010 Winter Olympics [7], it was indicated by the hospital manager that

	y(t)	x₁(t)	x₂(t)	m₁(t)	m₂(t)	
Operability	Patients per hour	Doctors	Nurses	Physical Integrity	Doctors Shift Factor	Doctors Shift Hours
100%	20	4	8	100%	100%	10
75%	15	3	6	80%	75%	15
50%	10	2	4	50%	50%	20
25%	7	2	3	20%	25%	35
0%	0	0	0	0%	0%	> 48

Fig. 12.9 Human Modifiers in the HRT
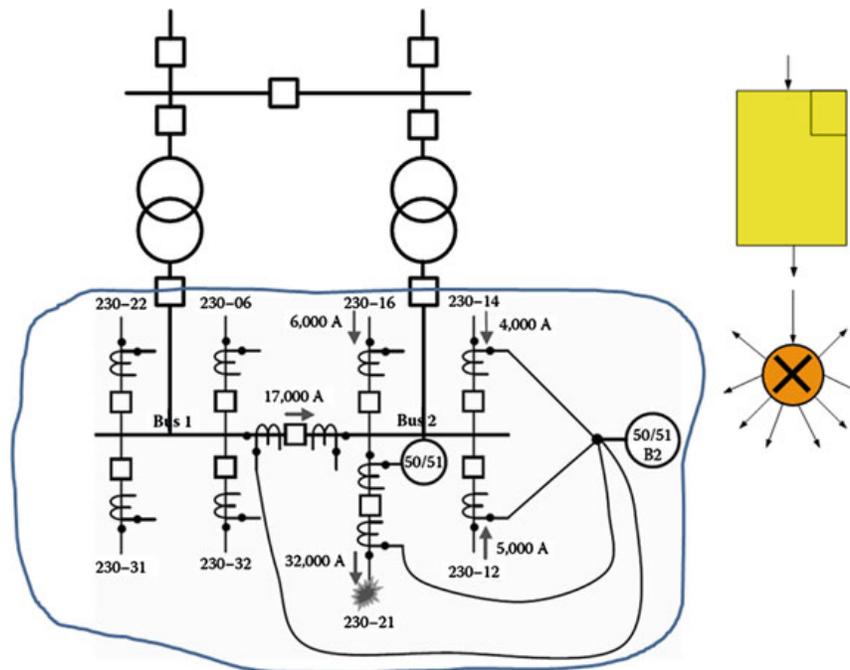
Fig. 12.10 Electrical substation layout

the availability of doctors during an emergency would not be an issue because the doctors' residences were located in the neighborhood of the hospital and the doctors would become available on very short notice.

The same HRT concept illustrated for a hospital is used for the other cells in the i2Sim model. Figure 12.11 shows the HRT for the power system substation layout of Fig. 12.10. Figure 12.12 shows the HRT for a water pumping station.

The electrical substation of Fig. 12.10 is being fed by two high voltage transmission lines of 40 MW each, through two transformers of 40 MW each. The substation

Operability	y(t)	x(t)	Condition
	Low Voltage Power (MW)	High Voltage Power (MW)	Transformers Working
Green	80	80	2
Yellow	40	40	1
Red	0	0	0

Fig. 12.11 Electrical substation HRT

	High Pressure Water (kL/h)	Low Pressure Water (kL/h)	Electricity (kW)	Pumps Working
	y(t)	$x_1(t)$	$x_2(t)$	
Green	500	500	50	10
Blue	350	350	35	7-8
Yellow	250	250	25	5
Orange	200	200	20	2-3
Red	0	0	0	0

Fig. 12.12 Water pumping station HRT

has eight feeders going out and two transfer circuit breakers. Assume, for the sake of illustration, that we are taking about the main substation in the campus of the University of British Columbia (UBC). Assume for simplicity of the explanation that each outgoing feeder is 10 MW and that two feeders go to the hospital, two to the residences, two to the classrooms, and two to the water pumping station. Suppose now that one of the two transformers has been damaged by an earthquake.

Since there are eight feeders out, the i2Sim Distributor (switchgear in the substation) will choose which four circuits will be supplied with 10 MW each and which four circuits will be shed. The high-voltage utility will be notified that only 40 MW will be required by UBC's substation.

Similarly to the case of the electrical substation, the HRT table of Fig. 12.12 for the water pumping station is established in terms of the number of water pumps that may be damaged by the disaster and the electricity available from the electrical substation to operate the pumps.

12.5.2 Channel Abstraction

The concept of Channel (Fig. 12.6) is an abstraction to obtain an equivalent of a complex system of lifelines. For example, to carry water from the water pumping station to the hospital, many pipeline segments may be used (Fig. 12.13). What is of interest for the i2Sim model is the amount of the water resource that arrives in the
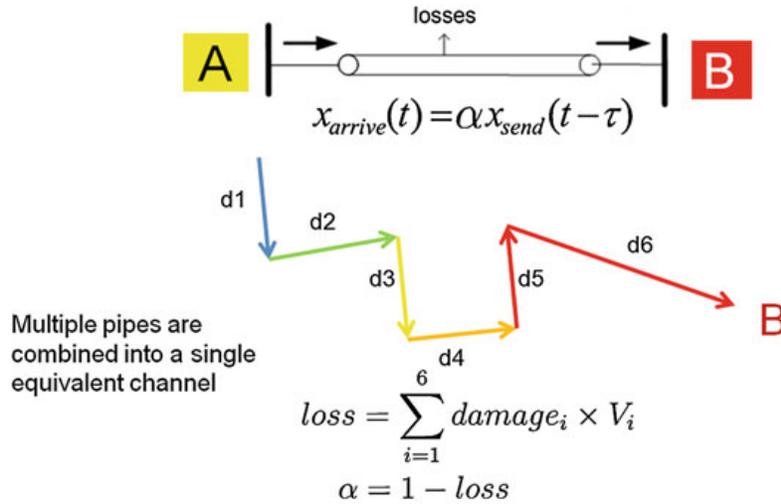
**Fig. 12.13** Channel abstraction
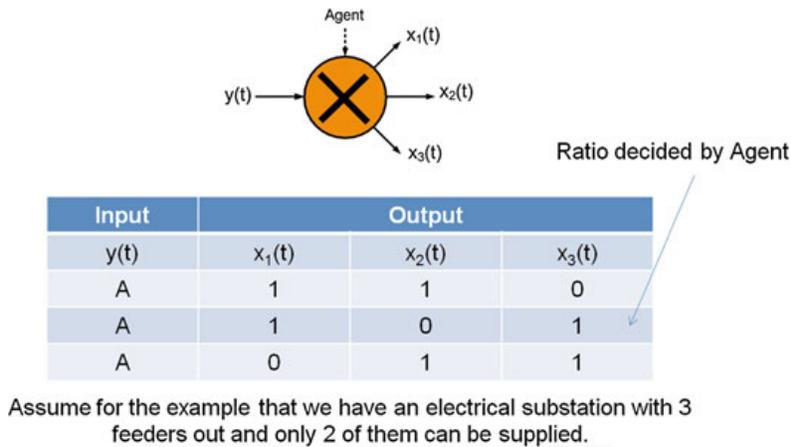
Fig. 12.14 Distributor at power system substation

hospital. The total losses in the water channel can be calculated from the combination of the individual losses in the water pipes. Damage assessment software can be used to make these estimates.

12.5.3 Decision Agents at Distributors

Figure 12.14 shows an i2Sim Distributor model for a power system substation. The Distributor decides how the electricity resource at the output of the substation is split among the cells that will receive the resource. The Distributor table will show what splits are possible according to the control mechanisms available at the substation (substation switchgear).

In the example, it is assumed that the amount of electricity that can be made available to the feeders is sufficient for only two out of the three feeders. A decision has to be made as to which feeders will receive electricity and which feeder will be disconnected (load shedding). This is a high-level decision that will impact the operability of the other cells in the system and eventually the optimization of the global output.

Normally, in disaster situations, the global optimization function is to save as many human lives as possible, and this will depend on the operability of the hospitals, the availability of ambulances, and the transportation times from the victims' site to the hospitals [7]. The decision of which load to shed should not be made at the level of the substation, or even at the level of the power utility company. Since it is a decision that will have a global impact, it should be made at the level of the decisions layer (Fig. 12.2) considering the effect of the decision on all the infrastructures involved (electrical, water, transportation, etc.) in terms of the global system function (save human lives).

After the decision is made, it needs to be carried out at the local substation level, either by site personnel (in the case of a manned substation) or by an automatic RTU/SCADA system in the case of an unmanned substation. If the communications system has been damaged during the disaster, the remote command decision might not be able to be implemented.

In the i2Sim simulation platform (Fig. 12.19), an optimization engine is coupled with the i2Sim simulator to find the optimum allocation decision in terms of the global system objective. For real-time decision support, the speed at which decisions can be optimized becomes critical and methods like genetic algorithms have proven to be too slow for the application. This is an area of current development, but some strategies like reinforced learning algorithms seem promising.

Depending on the size and complexity of the simulated system, optimization algorithms may prove to be too slow for real time applications. In these cases, a number of strategies can be used to produce reasonable solutions in real time. These range from increasing the computational speed by parallelization of multiple exploratory threads in multi-CPU environments, to pre-calculation of a large number of scenarios that can be used for a "good match" of a current situation. The "good match" of the current situation to a previous scenario may be available in a database of pre-calculated scenarios, or as done currently, "in the mind" of expert disaster responders. We hope that simulation will facilitate the task of these responders.

12.6 Matlab/Simulink Implementation

The solution of the i2Sim world of interdependencies among cells and channels in a subsystem representing a disaster area can be realized within the MATLAB/Simulink environment. Figure 12.15 shows the components of the i2Sim toolbox, while Fig. 12.16 shows a simulation scenario for downtown Vancouver developed for the 2010 Winter Olympics [7]. The downtown Vancouver model includes four power system substations, two hospitals, the water pumping station, and the BC Place and
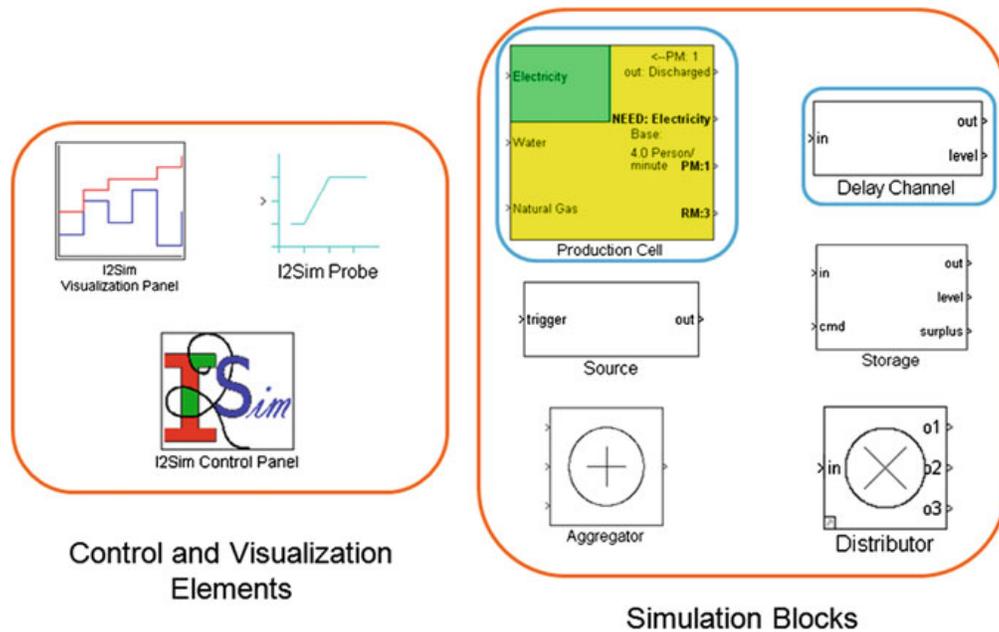
**Fig. 12.15**   i2Sim Toolbox for MATLAB/Simulink
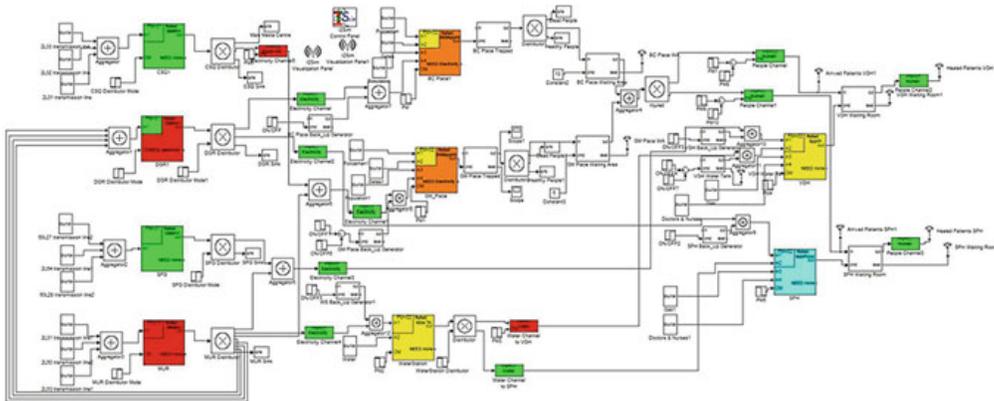
Fig. 12.16   MATLAB/Simulink model for downtown Vancouver

GM Place event venues. The simulation includes an evacuation model for an event that required the egress from BC Place, and a traffic model to simulate the traffic conditions to transport victims to the area hospitals. The egress model considered a number of Human Modifiers like density, demographics, guidance and rapid response. Similarly the traffic model included a number of modifiers, like traffic lights, guidance, and intersection closures.

Figure 12.17 shows the results of an earthquake scenario for the island of Guadeloupe.[1] The graph shows the number of patients arriving and being treated in the various hospitals in the island after optimization of the available system resources.

[1] MATRIX fp7 EU Project: Multi-Hazard and Multi-Risk Assessment Methods in Europe.
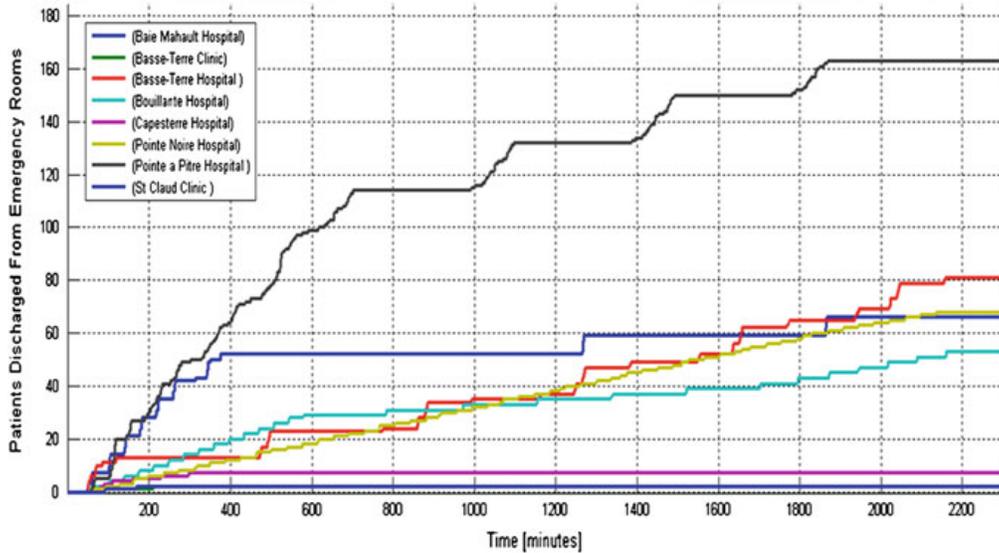
Fig. 12.17 Guadeloupe island disaster scenario. Patients discharged from hospitals

This is a large case that included the representation of 59 production cells, including power generation plants and distribution substations, water stations, hospitals, clinics, ambulance stations, and local communities; 349 channel cells, including roads and water pipes; 100 distributors, including power and water distribution and ambulance dispatchers; and other miscellaneous cells, including sources and reservoirs. Transportation system optimization, road conditions, and water availability were major issues in this study.

12.7 External Simulators

The i2Sim internal world resolves the interdependencies among cells and channels that use each other's resources. The purpose of i2Sim is to suggest the optimum allocation of the resources available inside the i2Sim world to maximize the system objective function (save as many human lives as possible). Damage caused by the disaster on the operability of cells and channels is a main reason for the scarcity in the production of resources. When a production cell cannot operate at full capacity due to physical damage, the resource produced by that cell is diminished and the other cells that depend on this resource (whether they have suffered physical damage themselves or not) cannot operate at their maximum capability.

In i2Sim, the external systems supplying resources to the i2Sim world are modelled as sources of that resource. How these sources are fed the resource requires the solution of the complete external infrastructure system associated to the source, for example, the power grid. i2Sim does not model the external power grid, but it assumes that sophisticated simulators are available to take into account all the op-
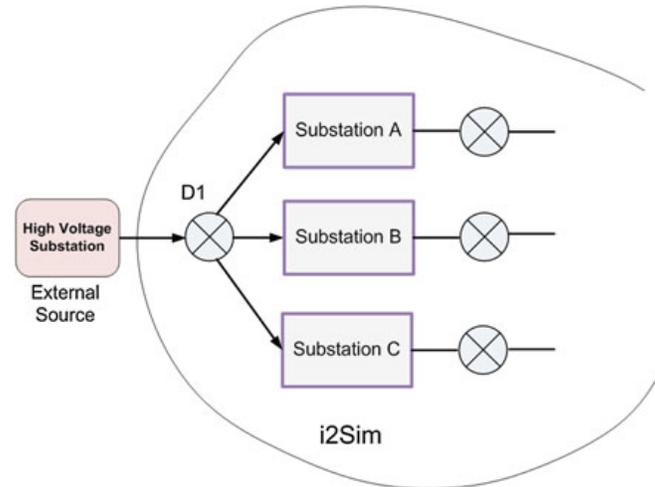
Fig. 12.18 External source distributor

erational constraints required for the feasible functioning of the grid. The possibly highly complex solution of the complete power system network is done by the "power system domain simulator".

In the example of Fig. 12.18, the domain simulator will tell i2Sim how much power will be available at Substation A, Substation B, etc. represented inside i2Sim. In case the external system cannot provide full power to all these substations, it will tell i2Sim what the possible allocation choices to supply these substations are. That is, it will tell i2Sim the settings of the D1 distributor. The optimum setting of D1 will be chosen by the optimization process inside i2Sim. This choice will then be communicated to the electrical utility.

12.8 Interfacing of Domain Simulators

The i2Sim architecture is designed to allow the interfacing of external simulators as plug-in modules. Figure 12.19 shows the software architecture design. The Enterprise Service Bus (ESB) is driven by a controller that keeps the solution time step and polls the attached modules. All information exchange is done through the common database. An external domain simulator is connected to the ESB through a software Adapter that translates the output of the domain simulator into the common i2Sim ontological language. For example, the power system simulator will indicate the amount of power that will be available at the various substations represented inside i2Sim. i2Sim itself is also connected to the ESB through an adapter and will pick up from the database the information provided by the domain simulator. i2Sim may also tell back the domain simulator (through the database) that not all the power that is available for the i2Sim substations can actually be used (example in Fig. 12.8). The external system might now want to allocate this extra power elsewhere.
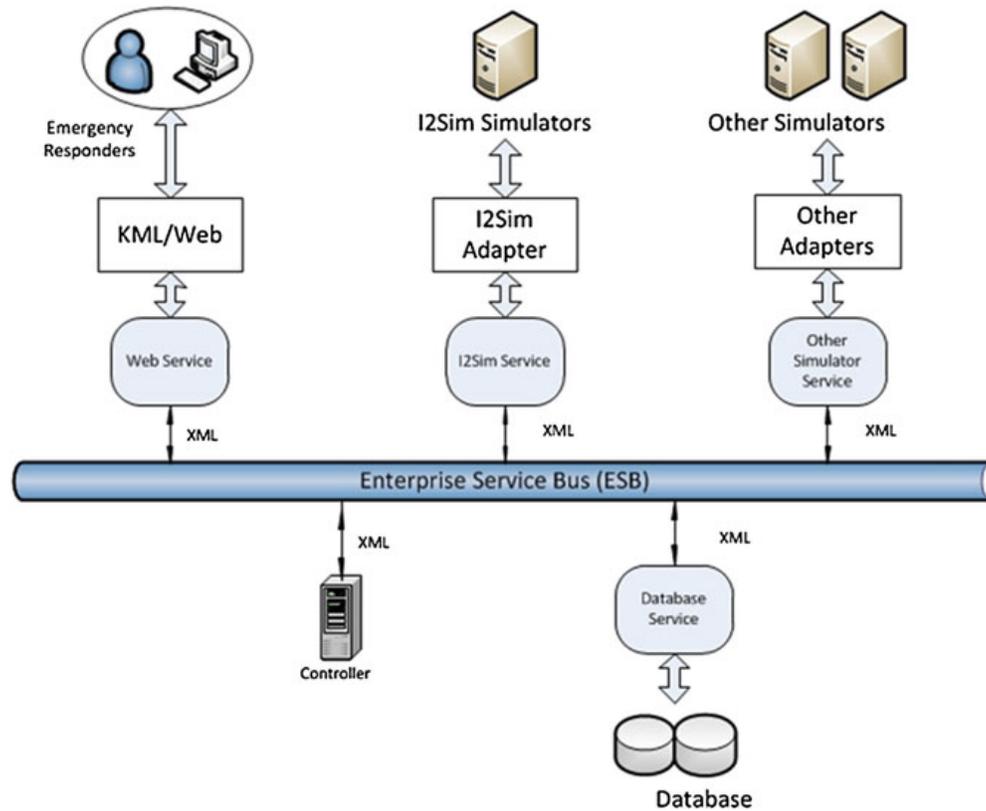
Fig. 12.19 i2Sim simulation platform. Enterprise Service Bus interface concept

Interfacing of external domain simulators can be very useful for planning of disaster scenarios where extensive areas served by provincial or national networks are affected. The architecture of i2Sim allows external domain simulators to be off-the-shelf software packages, that is, they do not need to be specifically written for the i2Sim environment. The only effort involved is in the writing of the Adapter to translate the normal input/output files of the simulator into a format and meaning that is understood by the i2Sim ontology. Figures 12.20 and 12.21 shows examples of i2Sim ontological translations for a commercial power system simulator PSCAD and a commercial water system simulator EPANet.

In the case of real-time disaster scenarios, the role of the domain simulator needs to be played by the utility company that provides the resource. That is, the availability of electricity has to be determined by the electric power utility, etc. This information has to be passed to the Disaster Manager of the Emergency Operating Centre (EOC) to be integrated into the i2Sim simulation. Privacy of information is a major issue for private utilities because of competitive concerns. However, the i2Sim framework goes a long way towards alleviating this issue by not requiring internal details from the domain providers. The only information needed is the availability of the resource at the points of interface with the i2Sim subsystem. The internal details of how the external system operates in order to make available a given amount of resources is of no interest to the i2Sim solution.
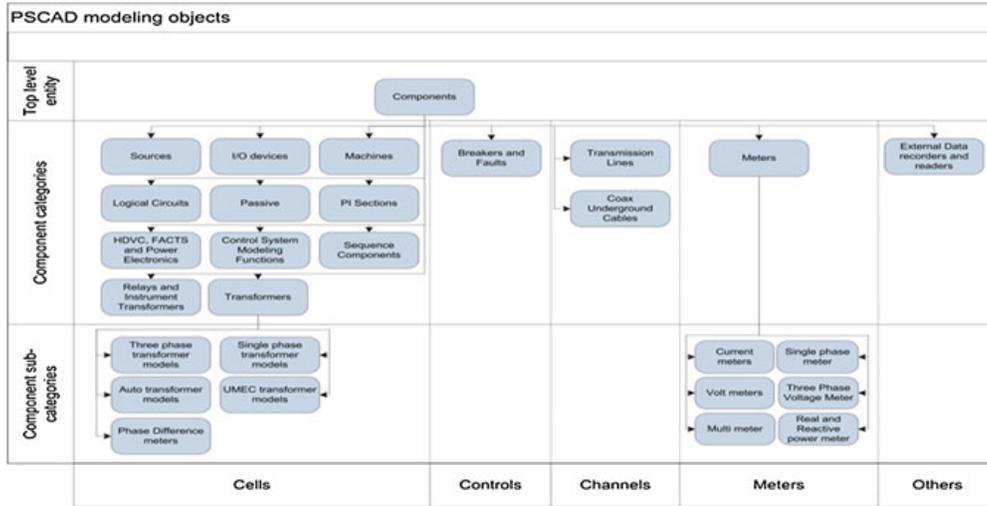
Fig. 12.20 PSCAD ontological mapping
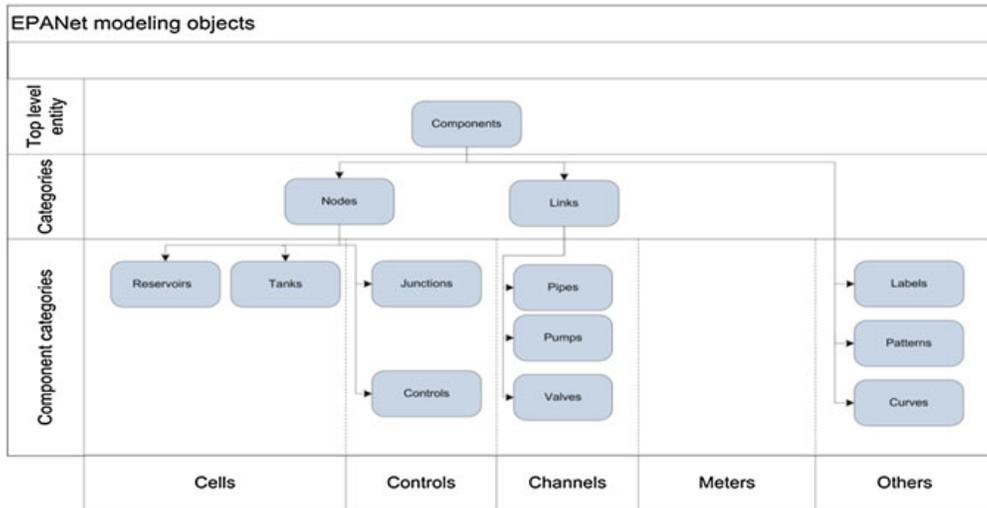
Fig. 12.21 EPANet ontological mapping

12.9 The i2Sim Solution Timeline

Figure 12.22 shows i2Sim's simulation timeline. A master clock in the ESB controller (Fig. 12.18) keeps the timing of the simulation process. The Master i2Sim finds a solution to the resources allocation problem at t = 0, Δt, 2Δt, 3Δt, … A typical time step for disaster simulations is 5 min for the first 10 h of the disaster.

The domain simulators will present updates when they change the amount of resources they can provide to the i2Sim sources. The physical damage assessment tools will present updates when the physical mode of cells and channels changes. Human and information modifiers will present their values when these values change.
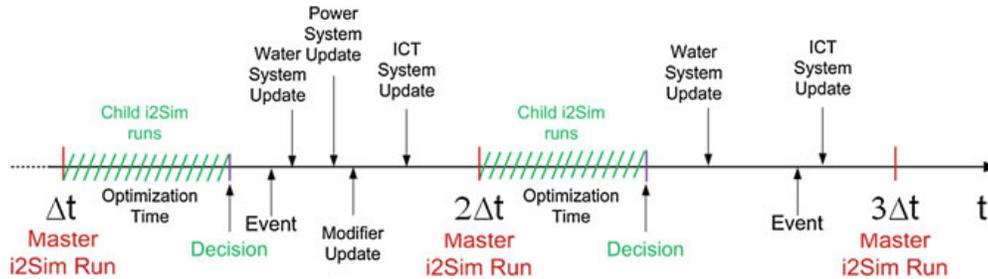
**Fig. 12.22** i2Sim simulation timeline
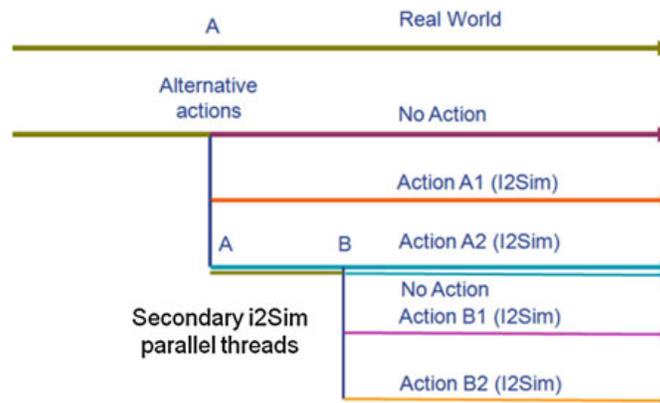
Fig. 12.23 What-if i2Sim scenarios

After all updates are received, the Master i2Sim simulator will perform the optimization of the allocation of available resources. This optimization process will normally involve the spawning of multiple i2Sim child runs to test different allocation scenarios (Fig. 12.23).

The i2Sim what-if scenarios allow decision makers to foresee the consequences of decisions before these decisions are applied to the real system. A software optimization engine can learn from these runs to suggest optimum actions.

For real-time applications, the external simulators need to complete their solution within the real-time clock interval. If an update is not received before the Δt solution cycle, the update cannot be included in the present cycle and has to wait until the next cycle. Similarly, if an optimal solution cannot be calculated within the Δt interval, the decision will have to be delayed or a suboptimal decision will have to be made. Planning scenarios are not constrained by execution speeds since the solution times do not need to match the simulated time.
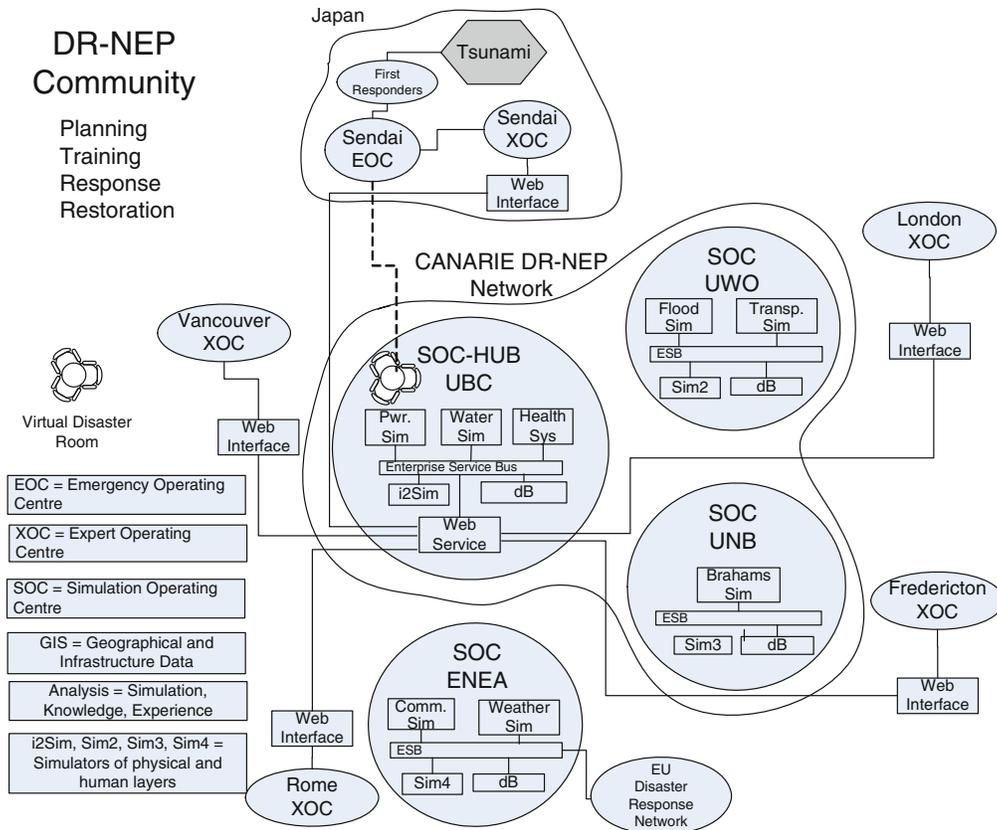
Fig. 12.24 DR-NEP community for disaster response

12.10 Integration Across Geographical Boundaries

The i2Sim solution architecture allows the integration of expertise and solution engines that can be located geographically apart from each other. It is often the case that an earthquake expertise centre is situated in a different location from a telecommunication expertise centre, etc.

CANARIE's[2] DR-NEP (Disaster Response Network Enabled Platform) network [8], based on the i2Sim simulation platform, interconnects research centres across Canada and internationally using very high speed (10 Gbps) optical fibre network. The speed and bandwidth of the connection allows the transfer of large amounts of data between sites and the interfacing of multiple simulation engines, attached to the architecture of Fig. 12.19, to work together in a seamless fashion, as if all simulators were located in the same laboratory (Fig. 12.24).

The DR-NEP network is currently being expanded to include sensor information from ground and under the sea monitoring stations. Through software adapters (Fig. 12.19), relevant data provided by these stations is stored in the ESB database.

---

[2] Canada's Advanced Research and Innovation Network (CANARIE).

Specialized software can access these data in real time to predict damage caused by the evolving disaster and other information relevant to disaster response (e.g., weather conditions). i2Sim can then pick up the damage information to update the physical mode of its cell and channel components, and other information that may trigger cell and channel modifiers.

12.11 Conclusions

Effective management of large disasters requires going beyond rescuing the victims and towards achieving a more optimum allocation of survival resources. In is in this aspect of effective management that large recent disasters (e.g., Japan, 2011, Haiti, 2010, Sichuan, 2008, Katrina, 2005, South Asian Tsunami, 2004) have fallen short and perhaps more lives could have been saved.

It is very difficult in large complex systems to predict and comprehend, even by highly skilled experts, the kinds of emergent behaviour that result from the interdependencies among multiple large dissimilar systems. The i2Sim multisystem simulation framework is an attempt to improve the management of resources during large disasters. The framework resolves the interdependencies among critical resources that can result in deadlock of the production and delivery of these resources if the consequences of these interdependencies are not understood. i2Sim does not replace domain simulators, which are needed to determine the availability of resources in a given area served by the domain utility. However, by explicitly modelling the interdependencies among domains, i2Sim can determine the best way to allocate the available resources within an area of concern and optimize the global system function of saving human lives in this area.

References

1. McClean, D.: World Disasters Report 2010. International Federation of Red Cross and Red Crescent Societies (2010)
2. UN Office for the Coordination of Humanitarian Affairs (OCHA): Japan Earthquake & Tsunami - Situation Report No. 6, Geneva (2011)
3. Maslow, A.: Toward a Psychology of Being. Van Nostrand, NY (1962)
4. Martí, J.R., Ventura, C.E., Hollman, J.A., Srivastava, K.D., Juárez, H.: I2Sim Modelling and Simulation Framework for Scenario Development, Training, and Real-Time Decision Support of Multiple Interdependent Critical Infrastructures during Large Emergencies. NATO (OTAN) MSG-060 Symposium on "How is Modelling and Simulation Meeting the Defence Challenges out to 2015?", Vancouver, 7–8 October (2008)
5. Leontief, W.: Input-Output Economics, 2nd Ed. Oxford University Press, NY (1986)
6. Federal Emergency Management Agency (FEMA): Hazus Methodology for Estimating Potential Losses from Disasters. http://www.fema.gov/hazus

7. Martí, J.R., UBC i2Sim Team: Modelling Critical Infrastructure Interdependencies in Support of the Security Operations for the Vancouver 2010 Olympics. Technical Report to Defence R&D Canada, DRDC CORA TR W7714–081111 (2010)
8. Martí, J.R., UBC/UWO/UNB Teams: Disaster Response Network Enabled Platform (DR-NEP. Technical Report to Canada's Advanced Research and Innovation Network (CANARIE) (2012)

Chapter 13
Addressing Interdependencies of Complex Technical Networks

Wolfgang Kröger and Cen Nan

13.1 Introduction

This chapter deals with large-scale technical systems, i.e., a wide-area network of physical-engineered infrastructures that function synergistically to provide a continuous flow of essential goods and services, groups within our societies or societies as a whole (increasingly) depend on. The most vital ones, such as the electric power and water supply system, information and communication technology (ICT), transport systems, are called **critical infrastructures**. They are subject to rapid technological and organizational changes (e.g., from monopoly to open competitive markets) and face multiple threats (e.g., technical-human, natural, physical, cyber, financial, contextual; either unintended or malicious); they may pose risk themselves (e.g., high-voltage lines or gas pipelines). In general, those systems have become more tightly integrated as well as more interdependent, also due to cyber-based host technologies for communication and control (SCADA[1] systems) moving from closed and dedicated to open and commercialized structures (see Sect. 13.2.2). As demonstrated by experience disruptions may start slowly, accelerate and cascade within and among infrastructure systems [1]. The "2003 Italian blackout" may serve as an illustrating example.

Those critical infrastructure systems have always been "complicated" but in recent years they have witnessed growing interconnectedness and interdependencies, have turned into **complex systems** (see Table 13.1 for contrasting juxtaposition).

---

[1] Supervisory Control and Data Acquisition.

W. Kröger (✉)
ETH Risk Center, ETH Zurich, Scheuchzerstrasse 7, 8092 Zurich, Switzerland
e-mail: kroeger@ethz.ch

C. Nan
Land Using Engineering Group, ETH Zurich, Universitätstrasse 16, 8092 Zurich, Switzerland
e-mail: cen.nan@usys.ethz.ch

Table 13.1 Contrasting complicated with complex systems (Acc. to [2])

Complicated systems (mechanical watches, commercial aircraft, nuclear power plants, etc.)	Complex systems (stock market, power grids, transport networks, www, social networks, etc.)
• Large number of highly connected components; frequency-consequence curves tend to follow a normal distribution	• Large number of highly connected components; frequency-consequence curves tend to show "fat tails" and follow power law distributions
• Components have well-defined rules and are governed by prescribed interactions	• Rules of interaction between the components may change over time and may not be well understood
	• Connectivity of the components may be quite plastic and roles may be fluid; interactions are not obvious
• Structure remains closed and stable over the time; limited range of responses to changes in their environment	• Systems are more open, respond to external conditions and evolve; interact with their environment
• Low dynamic, mostly linear behavior	• High dynamic and non-linear behavior; sudden regime shifts possible
• No adaptation; one key defect may bring the system to a halt	• Display organization without a central organizing principle (self-organization/emergence)
	• Inadequate information about the state of the influencing variables; probabilistic rather than deterministic behavior
• Decomposing the system and analyzing sub-parts can give an understanding of the behavior of the whole, i.e. the whole can be reassembled from its parts ("deductionism")	• The overall behavior cannot be described simply in terms of their building blocks; the whole is much more than the sum of its parts ("systems approach")

13.2 Understanding Complex Systems by Means of Exemplary Systems

13.2.1 Electricity Power Supply System

The electric power supply system (EPSS), consisting of power generators, high-voltage transmission and low-voltage local distribution grids, with transformers/substations in between, has become one of the most important critical infrastructures that modern societies and other infrastructures depend on. However, electricity is seen as common good; security of supply is a key issue but public lacks awareness of major blackouts. In Europe, while originally designed to serve a region and to allow for trans-boundary assistance in case of need, the EPSS has turned into an open system with given energy flux boundary conditions crossing neighboring countries without centralized control. Regional and vertically integrated monopolies are being
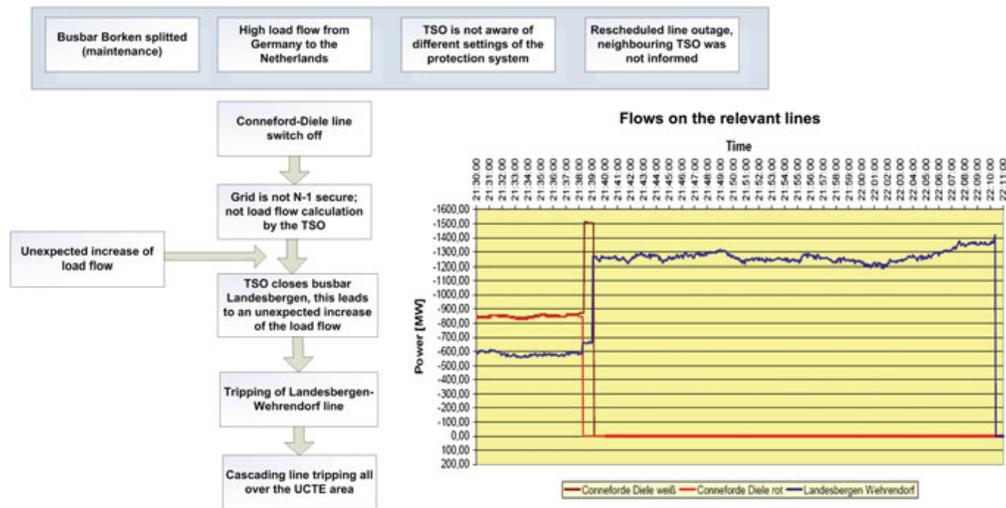
Fig. 13.1 Initial conditions and failure mechanisms leading to splitting of the ENTSO-E grid on 4 November 2006 (formation of areas at 22:20; re-synchronization of area 1 and 2 at 23:24, of area 3 at 23:57 h) [3]

replaced by an intricate market structure and stressing operation modes, closer to security margins. The risk of power outages spreading over wide geographic areas has increased. Furthermore, the integration of large shares of intermittent energy sources (wind and solar, increasingly at most suitable sites far away from consumer centers) has also made the power grid more vulnerable, often going along with lack of awareness and underestimation of complexity. As evidenced by the disruptive event of 4 November 2006, triggered by a planned, re-scheduled line cutoff (to let a new built vessel pass), the initial conditions can be manifold and of different types and the failure and spreading mechanisms are often hard to foresee and control (see Fig. 13.1). Finally the ENTSO-E[2] grid split into three areas of under (two)—and over (one)—frequency.

Table 13.2 depicts information about most recent major blackouts that happened in various regions of the world due to different reasons. Root cause analyses of them have revealed the following patterns:

- Operation of systems beyond original design parameters (high trans-border flows, integration of wind power, etc.).
- Malfunction of critical equipment and adverse behavior of protective devices; insufficient system automation in some cases (lack of investment).
- Lack of situational awareness and short-term emergency preparedness.
- Limited real time system monitoring beyond TSO (Transmission System Operator) control area and weak cross-border coordination in case of preparedness.
- Inadequacy of N-1 security criterion, of its implementation/evaluation.

---

[2] European Network Transmission System Operator-Electricity.

Table 13.2 Major blackouts of highly reliable bulk power systems

	Blackout	Loss [GW]	Duration [h]	People affected	Main causes
Aug. 14, 2003	Great Lakes, NYC	~60	~16	50 Mio	Inadequate right-of-Way maintenance, EMS failure, poor coordinadintion among neighbouring TSOs
Aug. 28, 2003	London	0, 72	1	500'000	Incorrect line protection device setting
Sept. 23, 2003	Denmark/Sweden	6, 4	~7	4, 2 Mio	Two independent component failures (not covered by N-1 rule)
Sept. 28, 2003	Italy	~ 30	up to 18	56 Mio	High load flow CH-I, line flashovers, poor coordination among neighbouring TSOs
July 12, 2004	Athens	~ 9	~3	5 Mio	Voltage collapse
May 25, 2005	Moscow	2, 5	~4	4 Mio	Transformer fire, high demand leading to overload conditions
June 22, 2005	Switzerland (railway supply)	0, 2	~ 3	200'000 passengers	Non-fulfilment of the N-1 rule, wrong documentation of line protection settings, inadequate alarm processing
Aug. 14, 2006	Tokyo	?	~5	0.8 Mio households	Damage of a main line due to construction work
Nov. 4, 2006	Western Europe (planned line cut off)	~14	~2	15 Mio households	High load flow D-NL, violation of the N-1 rule, poor inter-TSO coordination
Nov. 10, 2009	Brazil, Paraguay	~14	~4	60 Mio	Short circuit on key power line due to bad weather, Itaipu hydro (18 GW) shutdown
March 11, 2011	Northern Honshu	41	days		Grid destruction by earthquake and tsunami
July 30–31, 2011	India	~32	~2 days	620 Mio	High power demand due to extreme weather (heat) situation, weak inter-regional power transmission corridors

As "soft" (organizational, human) factors often dominate they cannot be ignored when analyzing the EPSS.

Due to pervasive use of cyber-based technology, partially unsecured like the internet, the risk of cyber attacks on the EPSS, and on the SCADA system and EMS (Emergency Management System) in particular, has increased but does not manifest as a trigger for blackouts yet. Parts of the EPSS spread over wide geographic and socio-political areas and are easily accessible, making them highly vulnerable to terrorist attacks; investigations have shown that "brute force attacks" (on more than single elements) are necessary to imperil the stability of a large-scale grid [4].

The tendency to growing instabilities may also be amplified by future trends within the EPSS:

- Future power system requires significant changes in the transmission and distribution system ("smartgrid"/"super grid") including RES-generation at most suitable sites and long-distance transport to consumer hubs.
- Means to better balance demand and supply will be given to "households"; the current generation of "smart meters" is unsecured introducing the risk of manipulation and cyber attacks ("worst scenarios" show grid collapse).
- Development of future market-oriented power supply systems are driven by political targets and demonstration of feasibility; vulnerability and security issues are often not sufficiently included.

Given the complexity and complex behaviors following disruptive events of the systems such as the EPSS it has been argued that reliability and vulnerability analysis have to go beyond the conventional approach of decomposition (e.g., fault tree analysis) or cause-and-effect/causal chain development (e.g., event tree analysis) to be able to capture emergent behavior and failure cascades, especially when strong interdependencies exist (see [5]). The behavior of the whole system can hardly be understood/described as the sum of the behaviors of its elements. Furthermore, the operational contexts including organizational factors, safety culture, coexistence of different technologies, etc, need to be adequately accounted for.

13.2.2 Industrial Control System

The growth of the worldwide interconnectivities of computing devices provides users new means to share and distribute information and data. In industry, this results in the adoption of modern ICTs and, subsequently, in an increasing integration of various facilities, i.e., industrial control system (ICS). In general, ICS is a term that encompasses several types of control systems, e.g., DCS (Distributed Control System), PLC (Programmable Logic Controller), SCADA, etc. ICS is typically used in modern critical infrastructure systems to enable the operators to continuously monitor and control them for the purpose of ensuring their proper operation [6]. Compared to other ICSs, SCADA system is normally used to monitor and control very large industrial process facilities such as electricity transmission facilities and oil
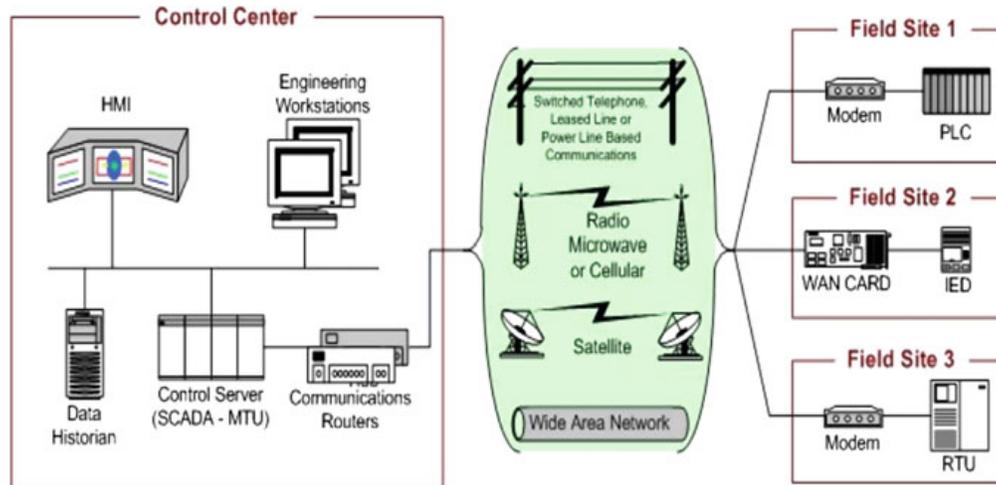
Fig. 13.2 General structure of a SCADA system [8]

and gas production facilities [7]. Its fundamental purpose is to allow a user (operator) to collect data from one or more remote facilities and send control instructions back to those facilities. For instance, voltage, frequency and phase angle are all important parameters in an EPSS and need to be continuously monitored for maintaining a normal operation environment.

Figure 13.2 shows the general structure of a SCADA system. There are four levels in a standard SCADA system hierarchy mainly based on the functionalities of devices. Level 1, the lowest level in the standard hierarchy, includes Field Level Instrumentation and Control Devices (FIDs and FCDs), e.g., sensors and actuators. Remote Terminal Unit (RTU), the level 2 in the standard hierarchy, is a rugged industrial common system providing intelligence in the field. It is a standard stand-alone data acquisition and control unit with the capabilities of acquiring data from monitored processes, transferring data back to the control center, and controlling locally installed equipments. Communication Unit (CU), the level 3 in the standard hierarchy, provides a pathway for communications between a control center and RTUs. Different protocols (e.g., Modbus and Profibus) and mediums are adopted by the CU. Most devices in the scope of the first three levels of the SCADA system hierarchy are installed (hardwired) in a substation. Master Terminal Unit (MTU), the level 4 in the standard hierarchy, can be regarded as a "host computer" issuing commands, collecting data, storing information, and interacting with SCADA operator who can communicate with substation level devices. Compared to the RTU, the MTU is a "master machine", which is able to initiate the communication either automatically by its installed programs or manually by an operator. Generally, three devices are included in a MTU: HMII (Human Machine Interface), control server, and engineering working station. The hardware configuration varies depending on the type and size of the system, while general functionalities are similar (see [9] for more information).

The trend from proprietary technologies to more standardized and open solutions together with the increased number of connections among ICSs and LAN/WAN
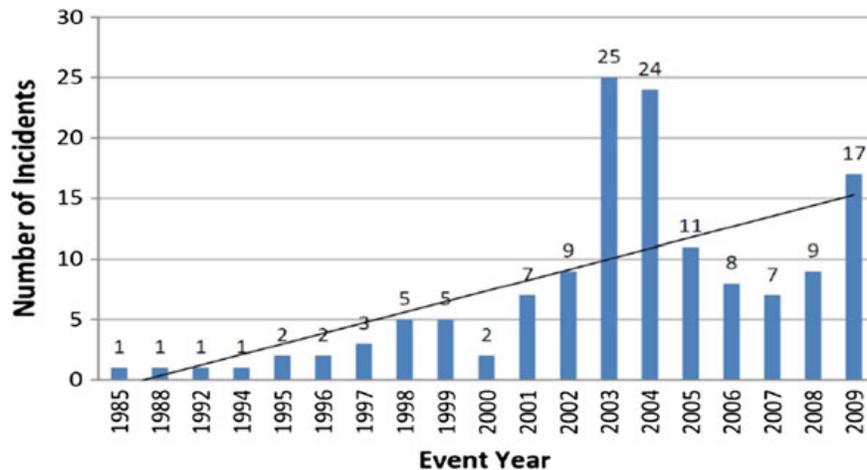
Fig. 13.3 Distribution of annual industrial security incident rates [17]

(Local/Wide Area Network) poses a significant threat. Originally, a SCADA system was designed as a point-to-point system connecting a monitoring or command device to remotely located sensors or actuators. By now, it has evolved into a complex network that supports the communication between a central control unit and multiple remote units using advanced ICT [10]. Having said this, extensive uses of them introduce new types of security threats to SCADA systems [11, 12]. For example, Stuxnet, a self-replicating computer worm, has recently challenged the securities of infrastructure systems for its capability of modifying the control logic of field level control systems through SCADA systems. This sophisticated "superworm" is a Windows-specific computer work, specifically written to attack SCADA systems, and was first discovered in June 2010. It should be noted that the only target of Stuxnet was Simatic WinCC, a Windows-based SCADA system developed by SIEMENS. Once inside the system, it uses certain exploits to infect other WinCC computers within the local network. According to [13], this computer worm infected Iran's nuclear enrichment facilities at Natanz, and other sites, and destroyed 30 % of its centrifuges by a self-destruct mechanism.

Recent surveys show that a number of attacks against ICSs, especially SCADA systems, have been reported over the years, e.g., the prominent Maroochy Shire accident in Australia (2000), the Florida power outage in USA (2008), etc [14, 15]. There are also numerous unreported incidents by asset owners and operators related to security issues in ICSs [16]. As seen from these incidents, threats to ICSs come from numerous sources, e.g., hostile governments, disgruntled employees, malicious intruders, human errors, technical failures, natural disasters, etc. Figure 13.3 shows annual industrial security incident rates from 1985 to 2009 based on records from RISI (Repository of Industrial Security Incidents).[3] As shown in Fig. 13.3, the annual incident rate gradually increased in the late 90's and peaked around 2003. It then

---

[3] RISI is a database including a number of technical incidents in which process control, industrial automation or SCADA systems were affected.
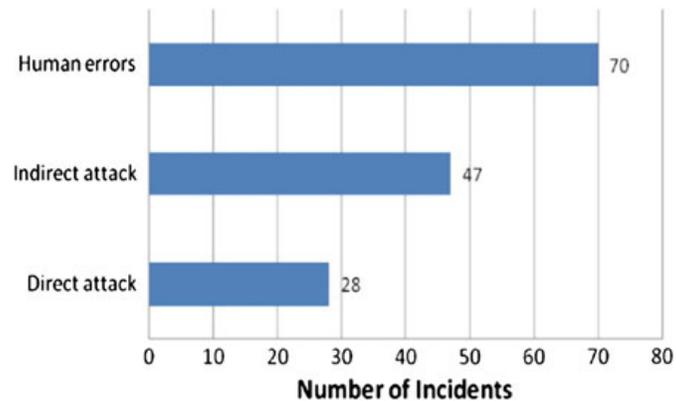
Fig. 13.4 Comparison of different types of industrial security incidents (1985–2009) [17]

declined sharply in the mid 2000's (2005–2007) and appeared to rise again in the late 2000's; a linear interpolation shows that its trend is increasing at probably 20–25 % per year over the last decade [17]. These incidents can be also grouped into direct attacks, indirect attacks (e.g., worms and virus), and human errors (Fig. 13.4). It should be noted that an incident can be classified into more one category. For example, an incident might may be caused by direct attacks and human errors. Unintentional incidents, e.g., equipment failures and malware attacks, also account for a significant number of incidents.

13.2.3 Railway System

Railway systems provide transportation services for passengers and goods in almost all countries and across borders. It is a large-scale infrastructure system that, if degraded, disrupted or destroyed, has serious impacts on the health, safety, security and well-being of citizens and on the effective function of the society. The 2009 Viareggio incident may serve as an example. On June 29, a freight train from Trecate, hauled by a locomotive with 14 bogie tank wagons derailed at Viareggio, Italy at 23:48 local time. The first wagon hit the platform of the station and overturned to the left, the next four wagons also overturned and the two following derailed but remained upright, the last seven did not derail, remaining intact on the track. The derailed wagons crashed into houses alongside the railway line causing a massive explosion that destroyed two blocks of flats, killing 22 people, injuring more than 40 and forcing around 1000 people to evacuate their homes (see [18] for this and other incidents).

In general, a railway system can be broken down to the following subsystems:

- **Infrastructure**: tracks, on-track equipment including switches, engineering structures (tunnels, bridges, etc.), associated station infrastructure (platforms, zones of access, etc.), safety and protective devices.
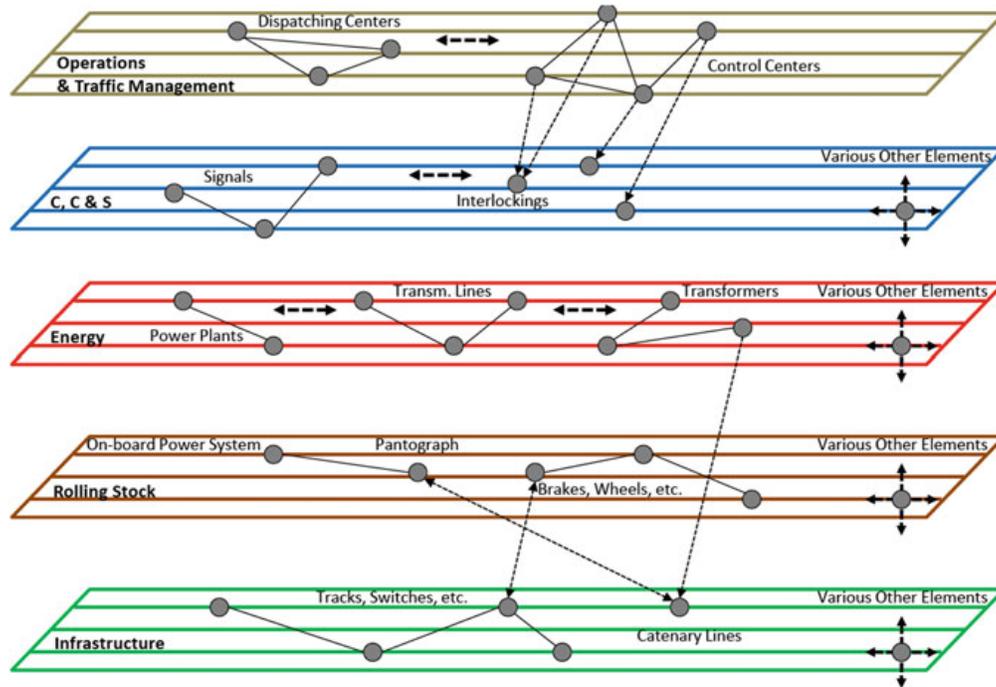
Fig. 13.5 Multilayer representation of the railway system [18]

- **Energy**: electrification system, including its own power plants, transmission systems, substations, transformers, overhead contact lines, etc.
- **Control, Command and Signaling**: all the equipments necessary to ensure safety and to command and control movements of authorized trains including track-side equipments such as radio block centers, interlockings, base transmission units.
- **Rolling Stock**: locomotives and wagons including all the various on-board equipment, named accordingly control equipment, structural components (brakes, wheels, car body, bogies, axles, etc.) and the power equipment (motors, main transformer, battery system, pantograph, etc.).
- **Operation and Traffic Management**: operation and control centers including the technical equipment and personnel at all levels of organization and operation.

Figure 13.5 shows a multi-layer representation of the railway system with various interacting hetero-geneous subsystems and associated components: parallel planes represent different subsystems while nodes represent various elements together with some of interconnections between them (arrows). The elements of the various layers depend on each other, depicted by various horizontal (inside a layer) and vertical (between layers) links. These links introduce direct and indirect (inter) dependencies and a failure of an element of the lower layers can cause cascading failures up to the top layers that could affect the function of infrastructure systems. For example, one plane represents the rolling stock subsystem whereas the parallel lines consist of the on-board power system (with nodes on this line being for the pantograph and various other elements), the on-board control system and so on.

Maintaining daily normal operation of railway systems is a highly challenging task and involves multi-dimensional, highly complex collections of technologies, processes and people and as such, the railway system is vulnerable to potentially catastrophic failures on many levels. In general, railway systems are subject to various hazards and threats:

- Sudden interruption of services due to loss of energy supply or communication and control.
- Operation of the system close to its limits (e.g., tight operational schedule).
- Malicious cyber-attacks on control systems.
- Accidents with injuries, fatalities or release of dangerous goods (e.g., derailment and/or collision) due to technical and/or human failures.
- Natural forces and environmental factors (e.g., landslides, extreme weather conditions) with consequences on operational availability and safety.

13.3 Interdependencies

13.3.1 Illustrating Evidence

Critical infrastructure (CI) systems have been continuously exposed to multiple threats and hazards. A single failure within any infrastructure system or even loss of its continuous service may be damaging enough to our society and economy while cascading failures crossing subsystems and/or even boundaries have the potential for multi-infrastructural collapses and unprecedented consequences. The importance of preventing or at least minimizing negative impact of cascading failures due to interdependencies among these systems has been recognized, not only by governments but also by the public, as a topic of CI Protection (CIP). The purpose of the protection is not just to identify the cause of failures and prevent them but also to halt ongoing cascading or escalating events before affecting other infrastructures. Therefore, it is vital to get a clear understanding of these often hidden interdependency issues and potential failure cascades, and to tackle them with advanced modeling and simulation techniques. In general, addressing the significance of interdependencies among infrastructure systems and uncertainties of their interactions is a challenge due to the complexity and perpetual nature of those systems, the lack of sufficient information clearly characterizing failure propagations, and the lack of modelling/simulation tools, by which system interactions can be comprehensively analyzed.

Nevertheless, it is still possible to find some evidences from many documented incidents through qualitative analysis of available information, which can help us to shed some lights on the understanding the characteristics of interdependencies [19]. The 2001 Baltimore tunnel fire may serve as an example: On July 18, a freight train with 31 loaded and 29 empty cars passed through the Howard Street Tunnel in Baltimore, USA. At 3:08 p.m., 11 cars derailed while the lead locomotive was about 1,850 feet from the east portal. Four of them were tank cars and one contained tripropylene. The derailment caused the puncturing (2-inch-diameter hole near the
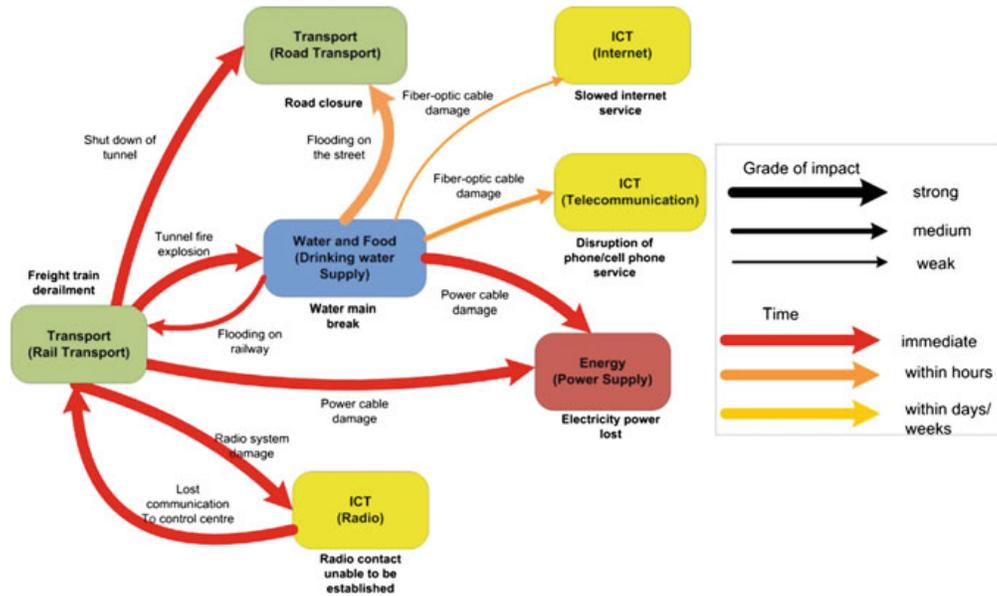
Fig. 13.6 Interdependency graph of 2001 Baltimore tunnel fire [19]

bottom of the tank) of the car carrying tripropylene and the subsequent ignition of this flammable liquid. The fire spread the contents of several adjacent cars, creating the heat, smoke, and fume that blocked the access of the tunnel for five days and eventually shut down the down-town area. As shown in Fig. 13.6, a technical failure (freight train derailment) occurring in the railway system continued to propagate into other infrastructure systems due to interdependencies. For instance, the break of the water mains (a failure within drinking water supply system) due to the tunnel fire/ explosion flooded the tunnel and damaged power cables and fiber-optic cables. The radio system was also damaged due to the derailment and therefore, the radio contact between train crewmembers and corresponding control center could not be established. About 1,200 Baltimore buildings lost electricity and both internet and telephone services were interrupted [20].

The 2012 India power blackout, occurred on July 30 and followed by another power outage on July 31, can serve as another example. The incident is the largest blackout in history, affecting 620 million people. It started from a tripped transmission line, which caused the failure of a substation. The cascade then spread further beyond this substation and led to a massive power outage throughout 22 states of India. Vital infrastructure systems were affected: railways and airports were shut down; health services provided by several hospitals were interrupted; drinking water services were interrupted due to the failure of electric pumps. This incident also demonstrates that the breakdown of such a complex infrastructure system is often the result of a relatively slow system degradation escalating into a fast avalanche of component failures, which finally lead to failures of directly or indirectly coupled systems.

These two incidents, as well as others such as the 2003 North America power blackout, 2004 Rome telecommunication node failure, 2005 Hurricane Katrina, etc, are regarded to be rare. It can be argued that the probability of future occurrence of

similar events could be relatively low. However, negative consequences of events, triggered by one single event, developing into fast cascades crossing system boundaries, can be worsened significantly due to interdependencies among systems. Analysis of those "low frequency, high consequences" disruptive events can help us to understand what can be expected due to interdependencies, even if in different contexts and scales. For example, cascades are directional in both cases, the 2003 North America and the 2012 India blackout, meaning that most of affected infrastructure systems have unidirectional relationships (dependencies) with power infrastructure systems.

13.3.2 Definition and Dimensions

From a technical perspective, the term *dependency* depicts a linkage between two systems through which the state of one system influences the state of the other, whereas *interdependency* is a bidirectional relationship through which the state of each system is correlated to the state of the other [21]. Interdependency can be of six different types: the first of three types can be referred as *direct* while the last three can be referred as *indirect interdependencies,* see below for a brief definition based on work done by Rinaldi et al. [21] and modified by the authors:

  (i) **Physical**—the state of one system depends on the material output(s)/flows(s) of the other, e.g., a pipeline network provides gas to fuel a power station while the electricity generated is used to power compressors and controls of the gas supply network;
 (ii) **Geospatial**—components of multiple infrastructure systems are in close spatial proximity and a local event is able to affect all these components, e.g., earthquake, flooding or a fire;
(iii) **Informational**—infrastructure systems are interconnected via electronic, informational links, e.g., a SCADA system monitors and controls elements of the electric power grid—likewise, it may provide pieces of information or intelligence supporting another infrastructure or a decision making process elsewhere;
 (iv) **Socio**—an infrastructure system affects another one via socio factors such as public confidence, trusts, culture issues, etc;
  (v) **Policy/procedure**—an infrastructure system affects another one due to factors such as market structure, organizational change, etc;
 (vi) **Finance**—an infrastructure system affects another one due to factors such as market condition, finance crisis, etc.

Figure 13.7 shows six dimensions for describing interdependencies including the six types. The "coupling and response behavior" of interdependent systems deserves special attention, as it directly influences whether the infrastructures are adaptive or inflexible when perturbed or stressed. As shown in this figure, the degree of coupling can be tight or loose, which addresses the nature of correlation of a disturbance in one system to those in another. The coupling order is either directly connected (first-order-effect) or indirectly through one or more intervening infrastructures
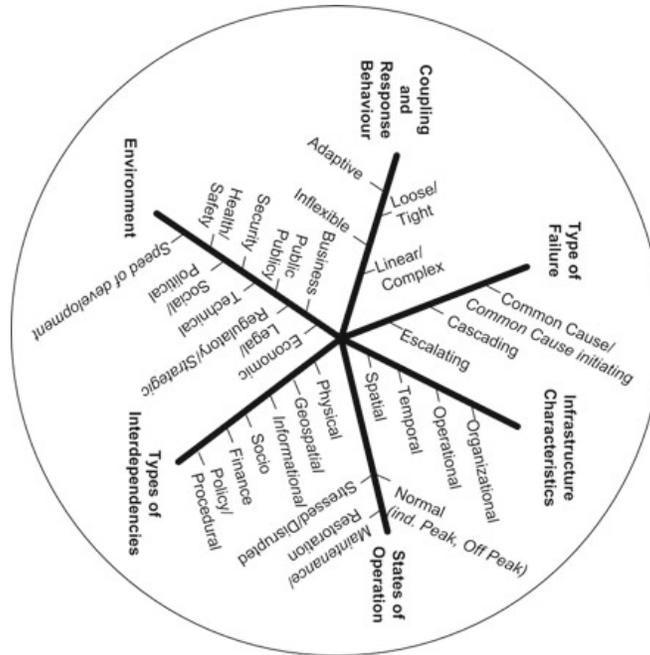
Fig. 13.7 Six dimensions for describing interdependencies (according to [21], modified by the authors)
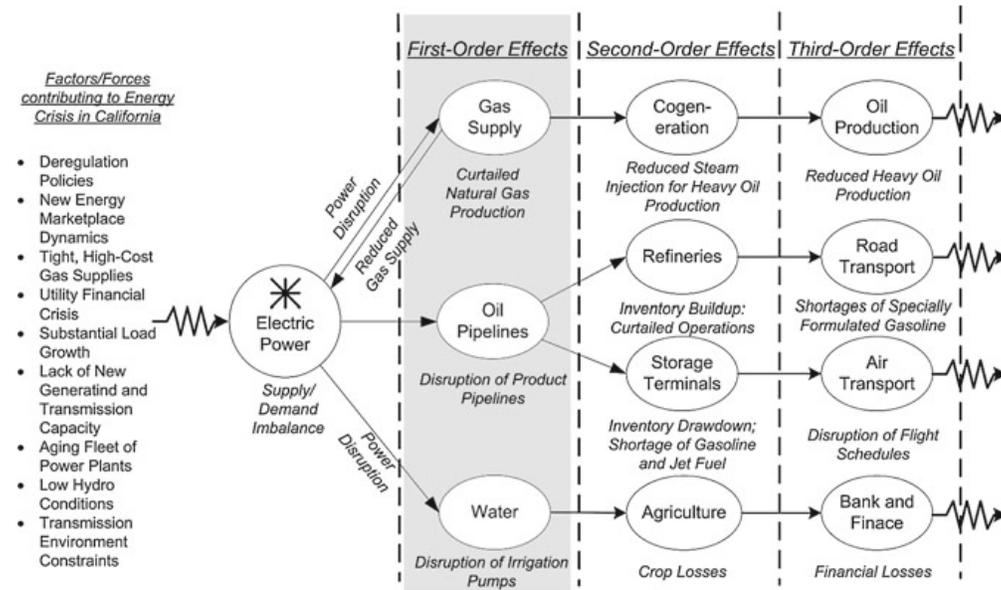

Fig. 13.8 Examples of nth-order interdependencies and effects taking Energy Crisis in California as a basis [21]

(second-order up to n-order effects, see Fig. 13.8 for illustration). The linearity or non-linearity/complexity of the interaction, i.e., whether or not systems can interact with others outside the normal scheme or operational sequence, not intended by design being subtle and difficult to detect, shows unfamiliar feedback loops.

Interconnectedness and interdependencies may have a positive or negative impact on the complex system behaviors indicating the need to find the right balance. Failures (negative impact) that arise from strong interdependencies (and coupling) can be classified as follows:

- **Common cause initiating events:** one event causing failure or loss of service of more than one infrastructure, e.g., areal external events such as earthquakes, floods, or extreme weather conditions, due to spatial proximity.
- **Cascade initiating events:** failure of one infrastructure causing failure or loss of service of at least another infrastructure, e.g., ruptures of mains of the water supply system.
- **Cascade resulting events:** failure or loss of service resulting from an event in another infrastructure, e.g., failure of gas lines due to loss of main electricity supply if compressors are electrically driven.
- **Escalating events:** failure or loss of service of one infrastructure escalating because of failure of another affected infrastructure, e.g., failure of the electric power system leading to failure of the SCADA system and by this affecting restoration of the electric power system.

Events being neither one of these four types maybe called independent. The types of non-independent events are not mutually exclusive.

13.4 Analyses of Interdependencies

The challenges regarding understanding, characterizing, and investigating interdependencies among infrastructure systems are immense and research in this area is still at an early stage [22, 23]. In recent years a great deal of effort has been devoted by researchers and two main directions can be distinguished, i.e., **knowledge-based** and **model-based** approaches.

13.4.1 Knowledge-Based Approaches

Knowledge-based approaches, e.g., empirical investigations or brainstorming, intend to use data collected by interviewing experts and/or analyzing past events to acquire information and improve the understanding of the dimensions and types of interdependencies. In order to address the question whether certain combination of infrastructure failures are more common than others, one of the early empirical investigation studies built a database using the collected information from a number of maintenance or operation accidents, reports of the US National Transportation

Table 13.3 Effect ratios [24]

Type of infrastructure systems	No. of times infrastructure systems caused failure of other infrastructure systems	No. of times infrastructure systems was affected by other infrastructure failures	Ratio of causing versus affected by failure
Water mains	34	10	3.4
Roads	25	18	1.4
Gas lines	19	36	0.5
Electric lines	12	14	0.9
Cyber/fiber/optic/telephone	8	15	0.5
Sewers/Sewage treatment	8	16	1.3
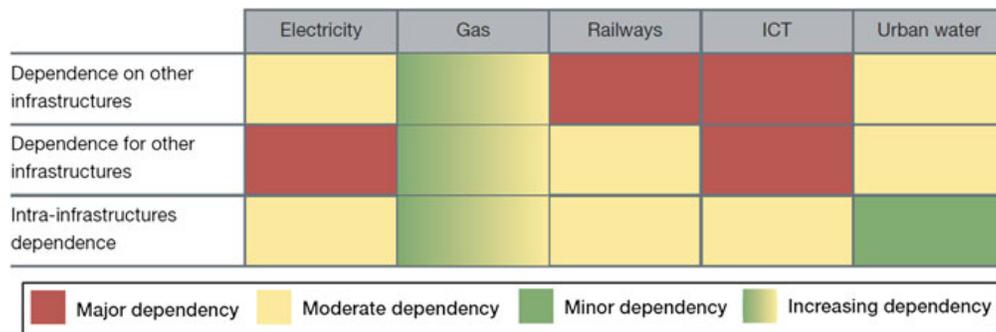
Fig. 13.9 Dependencies between critical infrastructures, according to Ref. [25]

Safety Board and news media searches [24]. The database mainly includes accidents that occurred from 1990 through 2004 in connection with failures during construction, maintenance or operation, or due to facility condition related to age of structures. Table 13.3 depicts the ratio of causing failure of another type of infrastructure versus being affected by failure of another type of infrastructure according to the database. As shown, water mains cause failures of other infrastructures more frequently while gas lines and telecommunication lines are more likely to be damaged by other infrastructures.

A policy brief of the International Risk Governance Council [25] also introduces an assessment based on brainstorming sessions among experts around the world and categorizes how dependent each infrastructure is on the others, how dependent the others are on it, and also how strong the intra-infrastructure dependencies are (Fig. 13.9). According to this report, among five reference infrastructures, electricity, railways and ICT are most important ones. Most infrastructures have a major dependency on the electricity infrastructure, while the railway infrastructure has a major dependency on other infrastructures. The ICT has a major dependency on others, as well as major dependence for other infrastructures.

The knowledge-based approach is straightforward and easy to understand. It is capable of providing a qualitative assessment on the severity of interdependencies

and can be considered as an efficient screening method. However, it is a purely data-driven approach, meaning that the accuracy of results depends on the quality and the interpretation of the collected information.

13.4.2 Model-Based Approaches

Model-based approaches aim to analyze interdependent infrastructure systems comprehensively by using advanced modeling/simulation techniques, capable of providing both quantitative and qualitative information. Even modeling single infrastructure systems is a challenging task because of their inherent characteristics such as dynamic/nonlinear behaviors and intricate rules of interaction with their environment due to their openness and high degree of interconnectedness. This task could become even more challenging when more than one infrastructure systems must be considered and interdependencies among them need to be tackled. Traditional approaches and methods based on decomposition and cause-consequence-relations such as fault and event trees reach the limit of their capacity [26, 27]. In recent years, a variety of advanced modeling approaches have been developed and applied, e.g., Input-output Inoperability Modeling (IIM), Complex Network (CN) Theory, PetriNet (PN)-based modeling, Agent-based Modeling (ABM), etc.

The **IIM approach** is an example of capturing interdependencies among infrastructure systems via the development of mathematical models. This approach is originally a framework for studying the equilibrium behaviour of an economy by describing the degree of interconnectedness among various economic sectors [28]. It assumes that each system can be modelled as an atomic entity whose level of operability depends on other systems and propagation between them can be described mathematically based on the basic Leontief high order mathematical model [29]. The IIM approach is capable of analyzing cascading failures and providing a mechanism for dependency measurement. In [30, 31], Haimes et al. applied this approach to study impacts of high-altitude electromagnetic pulse on electric power infrastructure. The great advantage of this type of mathematical model is its preciseness. However, deriving an appropriate representation of multiple infrastructure systems is not easy due to their inherent complexities. To overcome this difficulty, the task of analysing behaviours of interdependent infrastructure systems as a whole can be turned into the analysis of the aggregate behaviours of many smaller interacting entities.

The **PN-based approach** is a mathematical modeling language for the description of distributed systems which has also been used to represent/assess interdependencies among infrastructure systems. In this approach, components (subsystems) of infrastructure systems and their states are modeled using basic PN elements such as places, transitions, etc. In [32], the Swiss railway system is modeled using the PN-based approach for the purpose of vulnerability assessment, illustrated in Fig. 13.10. Elements of various subsystems such as track lines and transformers are selected and categorized as root causes potentially leading to single and/or common cause failures of the track lines. The core of the vulnerability analysis consists of integrating vari-
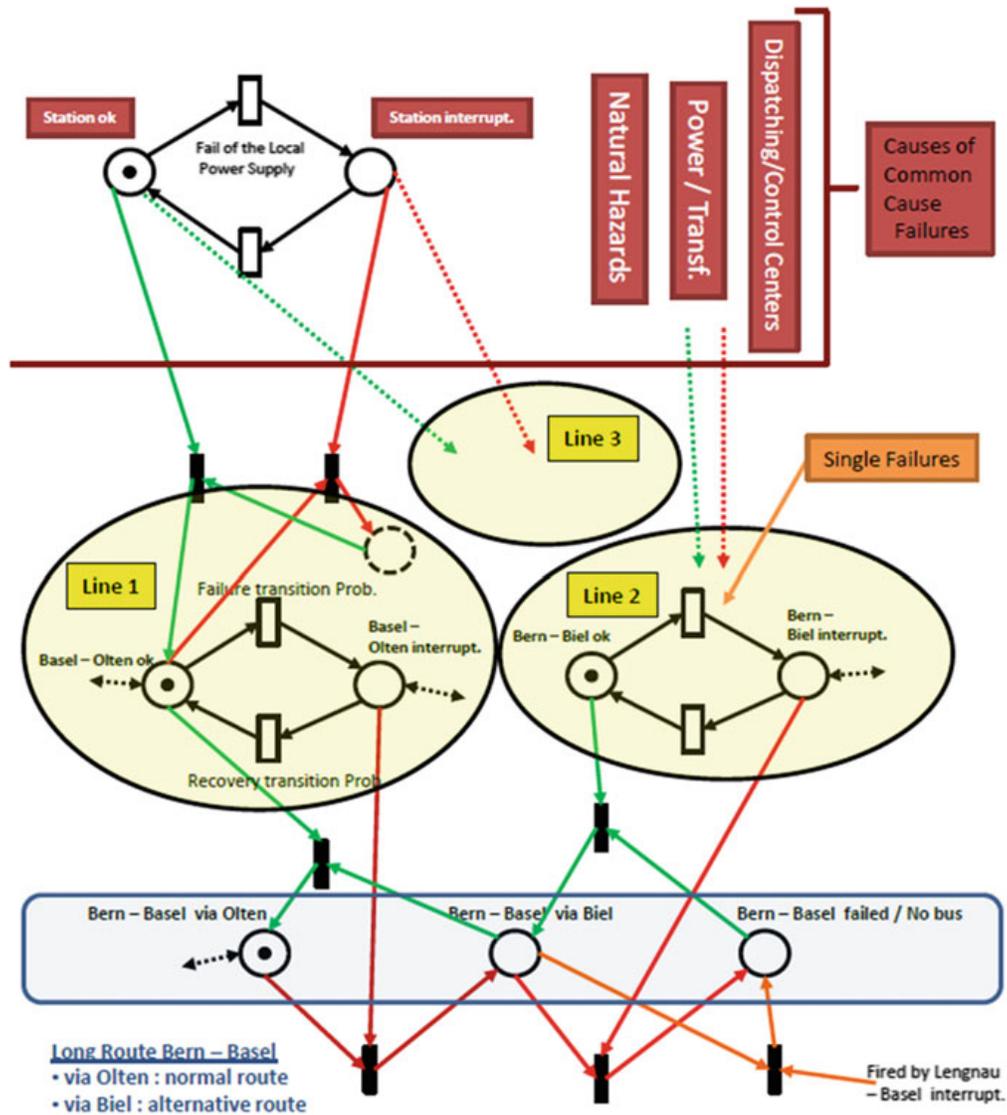
Fig. 13.10 The concept of PN-based modeling approach representing Swiss railway system [32]

ous risk factors affecting the system's operational performance in one "multi-layer" PN-based model.

This approach alone has difficulties representing infrastructure systems quantitatively and often needs to be combined with other methods. For example, in the Europe-wide project IRRIS (Integrated Risk Reduction of Information-based Infrastructure Systems), the PN-based approach is combined with the ABM approach to analyze and manage infrastructure interdependencies [33].

Fundamental elements of the **CN theory approach** are originally formed by graph theory [34]. A graph G (V, E) is composed by a set of nodes (vertices) V and the set of connections E between them. Each node (or vertex) represents an element of the system, while a link (or edge) represents the relation between corresponding
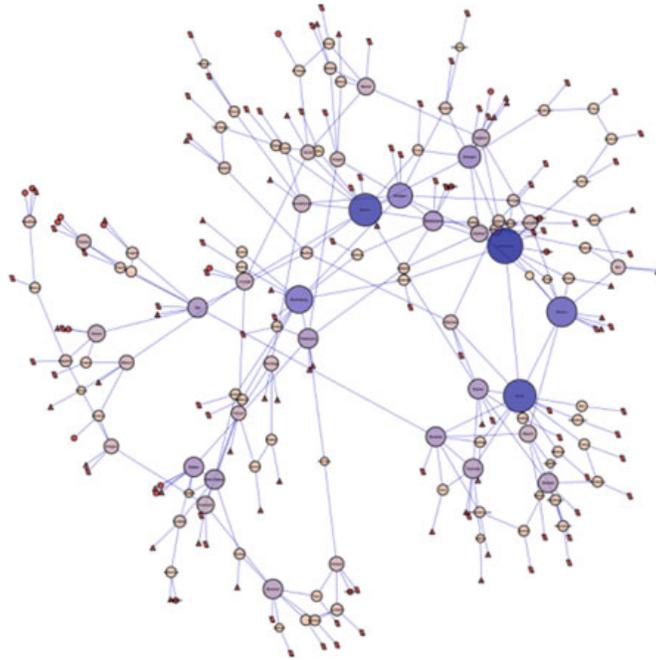
Fig. 13.11 Representation of the Swiss transmission grid using CN theory [4]

elements. A graph can then be drawn by plotting nodes as points and edges as lines between them. In general, a graph can be analyzed by well-developed parameters, e.g., the order/size of a graph, the weight/strength of a link, the degree/degree distribution/betweenness of nodes, etc. A complex network can be regarded as a graph with non-trivial topological features that do not occur in simple networks such as lattices or random graphs but often occur in real graphs. The CN theory is an approach capturing the coupling phenomenon as a set of nodes connected by a set of links and by this characterizing their topology. A number of modelling efforts have been made to adopt this approach for the development of infrastructure system models and interdependency-related assessments, demonstrating its capability of representing relationships established through connections among system components [35, 36]. In [4], the Swiss transmission grid is modelled and analyzed using the centrality analysis of this approach in order to perform heuristic investigations of potential malicious attacks (Fig. 13.11). In total, 242 nodes are developed to represent substations, loads, and power generating stations and 310 links to represent transmission lines.

The CN theory approach is based on the network model mapping physical configuration of the components (elements) of studied infrastructure systems and their (physical or logical) interconnections. The analysis of the topological properties of the network is able to reveal useful information about the structural properties, topological vulnerability, and the level of functionality demanded for its components. However, this approach lacks the ability to capture uncertain and dynamic characteristics of
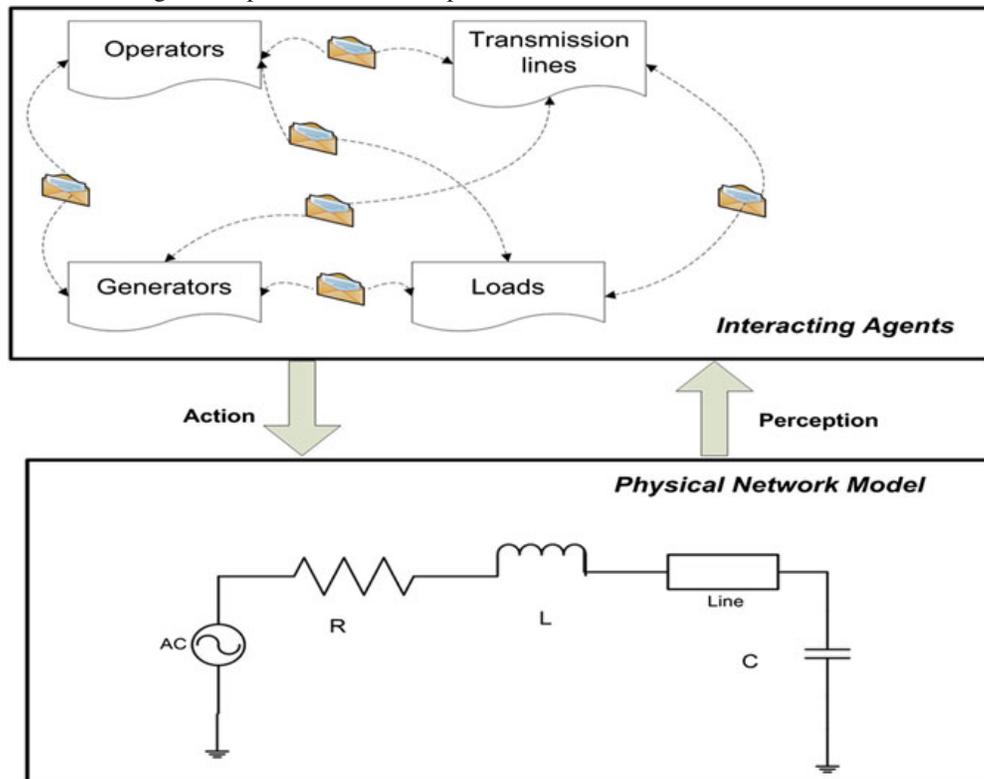
Fig. 13.12 Two-layer modeling concept (using application to the electric power supply system as an example)

infrastructure systems and system properties when dynamical processes, acting on the network, occur.

Using the **ABM approach**, each agent is capable of modifying its own internal data, its behaviours, its environments and even adapting itself to environmental changes. An agent can be used to model both a technical component (e.g., a transmission line), and a non-technical component (e.g., a human operator), while different agents interact with each other directly or indirectly. This approach is able to provide an integrated environment where a more comprehensive analysis of dynamic system behaviours can be performed by "looking-into" the component level of studied system(s) [37]. In [38], the Swiss transmission grid is modeled/simulated using the ABM approach for the purpose of system reliability analysis. Instead of only using nodes and links to represent substations and transmission lines respectively (recall CN theory modeling approach), agents are created to model various components of the system such as generators, busbars/substations, transmission lines, operators, and loads. The rules of behaviors of each agent are represented by using Finite State Machines (FSMs) and include both deterministic and stochastic time-dependent, discrete events. The model is developed using a two-layer modeling concept, illustrated by Fig. 13.12. Within this concept, the lower layer represents the separate modeling of the physical components by means of conventional, deterministic techniques such as power flow calculations, whereas the upper layer represents the abstraction of the

whole system (in this case, electric power system) with all its technical and non-technical components as individual agents. Overall, the ABM approach achieves a closer representation of system behaviors by integrating the spectrum of different phenomena that may occur, e.g., generating a multitude of representative stochastic, time-dependent event chains. However, this approach demands a large number of parameters defined for each agent, requiring thorough knowledge of the studied system(s).

It should be noted that other model-based approaches, which have also been applied by researchers but not discussed in this chapter, include **System Dynamic** [39], **Bayesian Network** [40, 41], **Dynamic Control System Theory** [42–44].

13.4.3 Comparison of Approaches

It is difficult to compare these (knowledge-based and model-based) approaches since all of these approaches have their own advantages and disadvantages. The knowledge-based approaches are straightforward and easy to understand, while the model-based approaches are more comprehensive and promise to gain a deeper understanding of behaviors of studied system(s). The level of this "deeper understanding" also varies: Some approaches are only capable of analyzing studied system(s) at the structure/topology level, which can be considered as appropriate approaches for the screening analysis, e.g., CN theory and PN-based modeling approaches, while some approaches are capable of capturing and analyzing dynamic behaviors of studied systems, e.g., ABM and IIM approach. Among all these, the ABM approach seems more promising than others, not just due to its capability for representing the complexity of any infrastructure systems, but also its modeling flexibility and adaptability. For example, the ABM approach can be integrated with many other modeling/simulation techniques and even be used to implement other models mentioned above.

13.4.4 Hybrid Modeling/Simulation Approach

13.4.4.1 Challenges and Basic Concept

Some of the model-based approaches which have been introduced and discussed in the previous section can be used to model interdependencies among infrastructure systems as well as single systems and interdependencies within, e.g., CN theory, PN-based and ABM approach. Some of them can only be used to model interdependencies, e.g. IIM. Due to inherent complexity of interdependencies among infrastructure systems, in practice, there is still no "silver bullet approach". Instead, it has proven necessary to integrate different types of modeling approaches into one simulation tool in order to fully utilize benefits/advantages of each approach and to optimize the efficiency of the overall simulation. One of the key challenges for developing
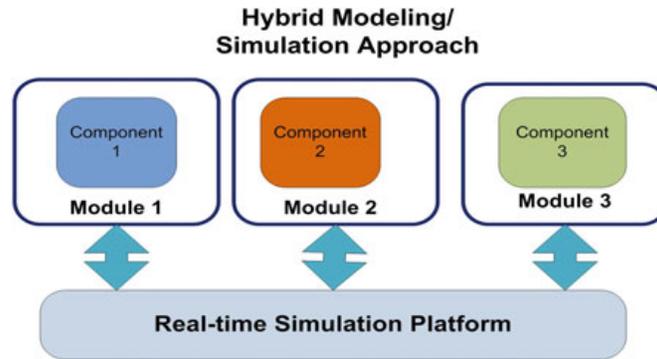
Fig. 13.13 Architecture of the hybrid modeling/simulation approach

such type of simulation tool is the required ability to create multiple-domain models, e.g., discrete and continuous time models, time-based and frequency-based models, and to effectively exchange data among them [45]. One solution for meeting these challenges and handling these technical difficulties is to distribute different simulation components by adopting the concept of modular design. The overall simulation platform can be divided into different simulation modules at first, which could be domain-specific or sector-specific simulation components, so as to make the best use of computational resources, and then distribute them across one simulation platform.

This so-called hybrid modeling/simulation approach, illustrated in Fig. 13.13, intends to integrate different modeling and simulation techniques, and can be considered as a successor of the traditional simulation approach in case multiple systems need to be simulated. It changes the way to design and develop simulation tools: Instead of building a "heavy weight" simulation component, a number of "light weight" components are developed interacting with each other over a real-time simulation platform, which not just potentially improves the efficiency and flexibility of the developed simulation tool but also decreases its overall complexity. Each distributed "light weight" simulation component is developed to represent its own system characteristics using appropriate modeling approaches. The information and control commands exchanged among simulation components are interpreted and processed over the network connection, allowing quick assembly of independently developed components without full knowledge of their peer simulation components.

13.4.4.2 High Level Architecture (HLA)

While several simulation standards do exist for supporting the distribution simulation components, the most widely implemented and applicable one is the HLA simulation standard [46], which is a general purpose high-level simulation architecture/framework to facilitate the interoperability of multiple-types models and simulations. In 1998, the first complete HLA interface specification was released to the public [47]. In 2000, HLA was approved as an open standard by the organization of

the Institute of Electrical and Electronic Engineers: IEEE Standard 1516–2000 [48]. Since then, the HLA standard has been revised and improved; the most current one is HLA-Evolved.

As an open IEEE standard, HLA has been widely adopted across various fields of simulation industries during the last decade. The EPOCHS (Electric Power and Communication Synchronizing Simulator) is an early attempt to distribute several individual simulators by adopting the HLA standard, which utilizes multiple research and commercial systems from various domains [49, 50]. Computer experiments show that "*the overall simulations have been sped up after distributing simulation components based on the standard of HLA*" [51]. Similar results are also observed while working on an agent-based framework for controlling activity flows between the ISS (Interactive Simulation Systems) components [52]. Furthermore, HLA has been applied to other industry fields such as the US border operation study [53], rail traffic safety system simulation [54], and many others [55–57]. Although, this standard has been questioned regarding its feasibility in the research field of interdependency study, it is still the most applicable and feasible one if compared to other similar simulation standards such as Distributed Interactive Simulation (DIS) and Aggregate Level Simulation Protocol (ALSP). One distinguished advantage of this standard is its support of live participants, meaning that the representation of the live world such as a human being, a real process instrumentation device and a field controller can be integrated into the simulation world. More details about the HLA standard can be found in [58]. While HLA is the architecture, a simulation standard, Run Time Infrastructure (RTI) is the software, the core element of the HLA standard, which provides common services to all participating federates.

13.4.4.3 Structure of the Experimental Simulation Platform

An experimental simulation platform has been developed to assess interdependency-related vulnerabilities between SUC (System Under Control) and its SCADA system by adopting the hybrid modeling/simulation approach (implemented using the HLA standard). The platform consists of four major components: SUC model, SCADA model, RTI server, and simulation monitor, all connected over a LAN (see Fig. 13.14).

The SCADA model is a discrete-event and agent-based model, developed by a failure-oriented modeling approach (Fig. 13.15). In this approach, the "agent state" is defined as a location of control with a particular set of reactions to conditions and/or events of its related agent. For example, open and close are two states defined for an agent representing a circuit break device. The "device mode" including both operational mode and failure mode is defined as the hardware status of corresponding simulated hardware devices. For example, failure-to-open and failure-to-close are two device modes defined for a field control device. The transition of various device modes can affect corresponding agent states. With the help of this modeling approach, technical failures of simulated devices of a SCADA system can be easily determined and corresponding failure propagations can be visualized/studied. The core of the device mode model is given by the state diagrams illustrated in Fig. 13.16, which
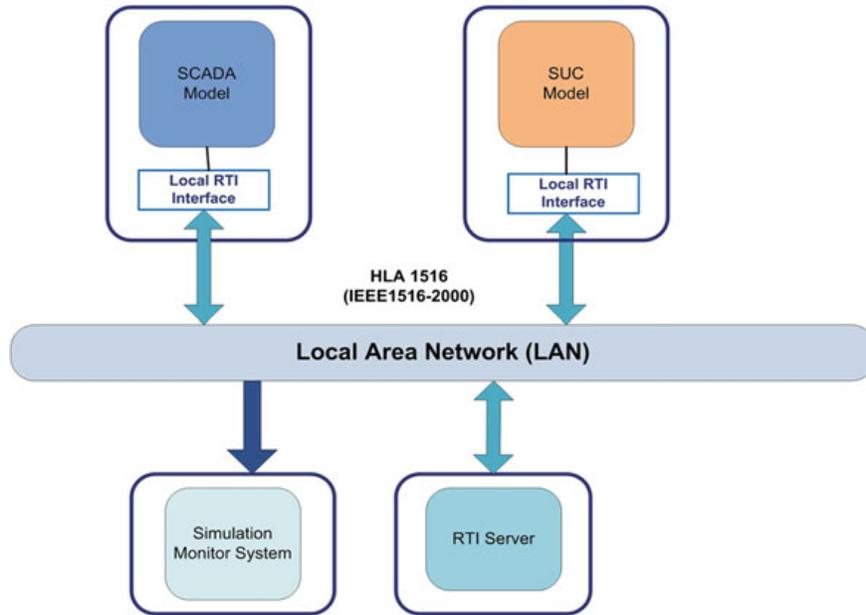
**Fig. 13.14** Architecture of the experimental simulation platform
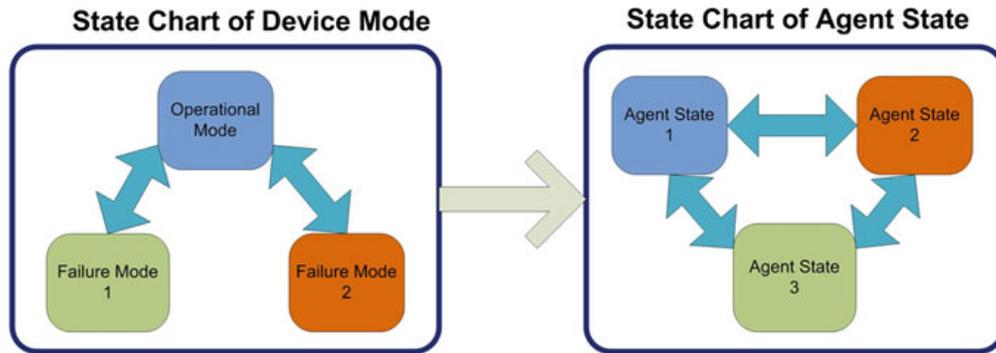
Fig. 13.15 Failure-oriented modeling approach

reflects a continuous-time, discrete-state Markov model describing failure behaviors of a studied device with one operation mode (left) and two failure modes (right) (see [59] for more details).

The SUC model is a continuous-time and agent-based model. The aim of this model is to investigate various system operating situations which could potentially result in a blackout of the Swiss electric power transmission network [5]. The SUC model simulates scenarios in a continuous time by means of conventional techniques such as power flow calculations. Since it was previously designed as a stand-alone model, no inputs from external models had been specified. To include this model in the experimental platform, a Java-based independent HLA-compliant interface is developed, which is responsible to process all inputs (outputs) to (from) the model (see [38] for further details).
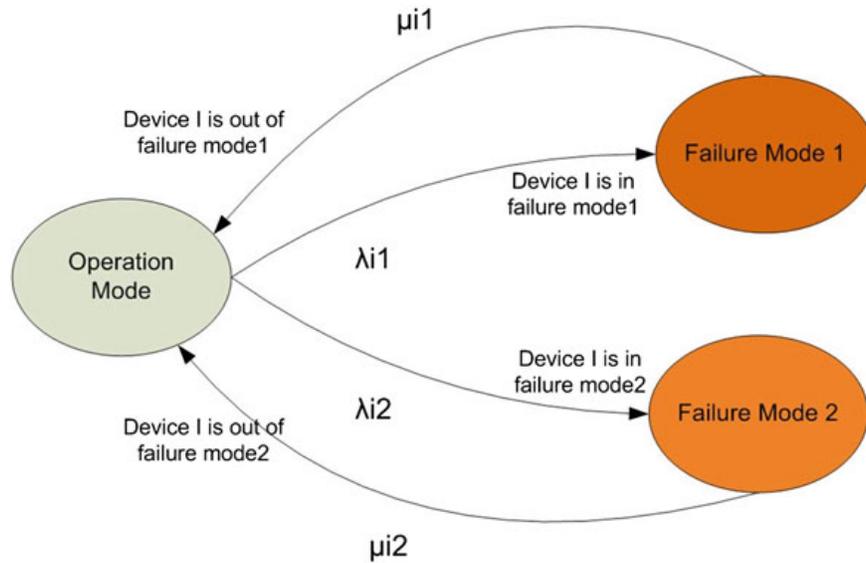
Fig. 13.16 State diagram of the device mode model (λ constant failure rate; μ repair rate)

The RTI server acts as the center of the experimental platform and is responsible for simulation synchronization and communication routing between all components, through the local RTI interface of each model. Each federate communicates with this server via its own local RTI interface and starts to follow central federation management. The simulation monitor system is a real-time tool, through which the simulation of two models can be observed.

13.4.4.4 Validation the Hybrid Modeling/Simulation Approach

To demonstrate the capabilities of the hybrid modeling/simulation approach, as well as of the simulation platform, for representing interdependencies among infrastructure systems, several experiments have been designed including feasibility and failure propagation experiments.

The purpose of the **feasibility experiment** is to study whether the HLA-compliant distributed simulation environment is capable to simulate interdependencies. In order to visualize the interdependency phenomena between SCADA and SUC, the scenarios that will trigger power line overload alarm are generated manually during the simulation. Generally, the maximum load each power transmission line can carry has been previously determined by its operator and is called *overload threshold*. If the real power flowing through a transmission line exceeds its overload threshold, this line is considered to become overloaded. An accidentally overloaded transmission line could cause a system collapse. Therefore, suitable corrective actions should be taken in order to alleviate the overloaded transmission lines. Normally, whenever a monitored transmission line is overloaded, an alarm will be generated and sent to the operator in the control center by the RTU of the SCADA system. If, after a

certain period, the operator fails to react to the overload alarm, then the protection devices such as disconnectors will automatically isolate the overloaded transmission line to minimize negative consequences. It should be noted that the procedure for handling a power line overload alarm is complicated and other factors should also be considered. In order to simplify this problem, it is assumed that the overload alarm failed to be handled correctly only if the operator fails to react to the alarm in time and the protection device fails to trigger. Three case study scenarios are developed by modifying parameters of corresponding agents in order to observe three different outcomes after the occurrence of the transmission line overload: (1) neither operator nor protection device react the alarm, (2) operator reacts alarm, (3) protection device is triggered after operator fails to react.

The observed simulation results from three case studies show that the propagation of cascading failures between infrastructure systems due to interdependencies can be simulated and visualized with the help of the experimental platform. Although the models are distributed, overall simulation performance is not affected and interconnections between models can still be efficiently handled (see [58] for more information).

To investigate the phenomenon of failure propagation and related issues, another experiment has been developed and conducted. In this so called **failure propagation experiment**, a number of tests are conducted by triggering single or even multiple technical failures in order to observe and study sequent events due to the failure propagation. For example, in a single technical failure test, which is mainly related to the investigation of the physical interdependency, the FID agent[4] is developed to represent a power flow transducer (PTi) measuring power flow (in unit of MW) transmitted in a selected transmission line that is included in the SUC model. It is assumed that the PTi is calibrated incorrectly due to the aging. A list of sequential events after the incorrect modification of the PTi's calibration value is recorded using a database during the simulation. As learned by studying these records, at certain time, the PTi's calibration value is modified incorrectly. As a consequence, the output of the PTi is more than its measured variable value should be. According to this wrong value, the RTU generates a wrong overloading alarm and sends it to the MTU causing the operator in the control room to make a wrong decision, i.e., to redistribute the power flow of a transmission line. As the result, the amount of power transmitted in this line decreases, although it should not. The measured variable from PTi, as part of the SUC, acts as physical input into the SCADA system. This relationship can be considered as the **physical interdependency**, which causes the failure of PTi to propagate from the SUC to the SCADA system and go back to the SUC (see [27] for more information).

---

[4] FID agent is an agent representing a field instrumentation device such as a sensor or transducer.

According to investigation results, analyzed based on both feasibility and failure propagation experiments, it can be concluded that three types of interdependencies can be simulated using the current experimental simulation platform: physical, cyber, and geographical interdependency.[5]

13.4.4.5 Brief Introduction of "In-Depth" Experiments

In order to investigate and identify interdependency-related (often hidden) vulnerabilities between the SCADA system and the SUC, three "in-depth" experiments are also developed and conducted.

The first experiment is the **substation level single failure mode experiment**, in which different failure modes of each substation level component (i.e., FID, FCD, and RTU) are evaluated by performing a number of tests related to each failure mode. In total 8 failure modes are defined for these substation level components such as FID FRH (Failure to Run (too high)), FCD FO (Failure to Open), RTU FRF (Failure to Run with Field Device), etc (see [59] for more information). One substation from the reference SCADA system including two transmission lines is randomly selected. During each test, the scenarios that will trigger power line overload alarm are loaded at the beginning of the simulation. Each test starts in the operation mode (a device mode) and one of the agent states. Within a given time period, the device mode of a respective component will go to one failure mode for which the transition time from is assumed to be exponentially distributed. After a given time period, the device mode will go back to operation mode for which the transition time is also assumed to be exponentially distributed. The transitions between different device modes have influences on corresponding agent states resulting in the change of behaviors of the SCADA system and SUC. According to the conclusion of this experiment, among all the simulated SCADA-related devices, negative effects caused by failures of the RTU device seem more significant on its interconnected SUC (see [59] for more information).

The second experiment, the **small network single failure mode experiment**, extends the scope of the first experiment to a small network including more components from the SCADA system and the SUC (40 substations and 50 transmission lines). In this experiment, one key substation[6] from the SUC model is selected for triggering the failure modes of substation level components during the simulation. For each single failure mode, two types of tests are implemented: normal and worse-case test. The modeling scenarios of normal case test are similar to of the tests in the first experiment. The worse-case test represents the worse-case situation when the operator is unable to handle any alarm received by the control center due to natural or technical failures (hazards), e.g., the failure of the control panel, flooding/fire in the control center, etc. The purpose of performing experimental tests under this situation

---

[5] Indirect interdependencies are not considered during these experiments.

[6] In this experiment, it is assumed that substations connecting more than 6 transmission lines are considered as key substations.

is to observe corresponding consequences if the SCADA system fails to monitor and control the SUC through the MTU. According to the conclusion of this experiment, on average, negative effects due to interdependencies are aggravated during worse-case tests, which have been demonstrated during FID FRH worse-case tests (see [60] for more information).

The third experiment, **whole network worse-case failure modes experiment**, extends the scope to the whole network including all simulated components of the SCADA system and the SUC, by which negative consequences caused by interdependencies can be observed and analyzed. In this experiment, instead of just considering single failures, double failures occurring simultaneously at different substations are also included. The same modeling scenarios defined in the worse-case tests of the previous experiment are applied, but in addition, two key substations and non-key substations are selected as exemplary substations. According to the conclusion of this experiment, failures of FIDs in both single and double failure tests show very strong degree of impacts. It is also observed in this experiment that the increase of the number of key substations could also lead to more significant negative consequences (see [60] for more information).

Based on the results from these experiments, vulnerabilities of the studied SCADA system due to its interdependencies with the SUC have been identified, which can hardly be obtained without an appropriate simulation tool due to the complexity of real systems. Furthermore, suggestions for potential technical improvements are proposed, which could be useful to minimize the negative effects and improve the coping capacity of both systems (see [60] for more information).

13.5 Conclusions

Large-scale/wide-area technical networks, such as critical infrastructures, have become increasingly interdependent going along with operational modes closer to their limits, thus stressing the systems. These tendencies and interdependencies, in particular, have dramatically increased the overall complexity of related infrastructure systems, turning them to "system-of-systems" and causing the emergence of unpredictable behaviors and negative impacts. Therefore, these systems become more vulnerable to cascading failures with widespread consequences. These interdependency-related issues should not only remain as a subject of theoretical research. The practical importance has been evidenced and highlighted by numerous major disruptive events (2001–2012) such as bulk electric blackouts and should not be underestimated.

These technical networks even continue to become more integrated and their behaviors may tend to become more complex. Understanding and characterizing them is a real challenge; research in this area is still at an early stage. It is essential to get a clearer understanding of their cascading behaviors by applying appropriate techniques. Consequently, modeling/simulating those systems will remain as a field of active research. Although progress has been made in advanced modeling and

simulation, more efforts are needed to further improve the methods/tools, to validate them and to scale them up to the level of "system-of-systems" and of the systemic nature of related risks.

In practice, there is still no "silver bullet" solution. Several approaches have been introduced and discussed in this chapter. Among these approaches, the CN theory is one of most frequently used techniques for topological analysis, while the ABM can be combined with other techniques such as the Monte Carlo simulation and offers the possibilities to include physical laws into the simulation and emulate the behavior of the infrastructure as it emerges from the behavior of the individual agents and their interactions. Combining different approaches and utilizing their strengths within one simulation tool by adopting the technique of distributed simulation using appropriate standards seem promising. This so called hybrid modelling and simulation approach has already proved its feasibility and applicability in recent research study and different types of experiments. Hopefully this approach will be adopted by researchers and practitioners in the field of risk analysis. With the help of this approach, traditional approaches such as the logic trees with limitations to capture the behaviour of those systems alone can also be combined with more advanced ones such as the ABM approach and used for more comprehensive system reliability/vulnerability analysis.

References

1. Newman DE, Carreras BA, Degala NS, Dobson I. Risk Metrics for Dynamic Complex Infrastructure Systems Such as the Power Transmission Grid. Proceedings of the 45th Hawaii International Conference on System Sciences, p. 2082–2090, 2012.
2. Kröger W, Zio E. Vulnerable Systems, Springer, 2011.
3. Habenberger J, Kröger W, Probst P, Raschke M, Schläpfer M, Birchmeier J. Stromversorgungssystem Schweiz. BABS Report: ETH Zurich, 2009.
4. Bilis EI, Kröger W, Nan C. Performance of Electric Power Systems under Physical Malicious Attacks. IEEE Systems Journal. Vol. 7(4), p.854–865, 2013.
5. Eusgeld I, Kröger W, Sansavini G, Schläpfer M, Zio E. The role of network theory and object-oriented modeling within a framework for the vulnerability analysis of critical infrastructures. Reliability Engineering and System Safety. Vol. 94, p.954–963, 2009.
6. Igure VM, Laughter SA, Williams RD. Security Issues in SCADA Networks. Journal of, Computers and Security, Vol. 25, p. 498–506, 2006.
7. Boyer SA. SCADA supervisory control and data acquisition. 3rd ed. Research Triangle Park: ISA; 2004.
8. Stouffer K., Falco J., Scarfone K. Guide to Industrial Control Systems (ICS) Security. National Institute of Standards and Technology; 2008.
9. Nan C, Kröger W, Eusgeld I. Focal Report: Study of Common Cause Failures of SCADA System at Substation Level, BABS Report: ETH Zurich, 2011.
10. Balducelli C, Bologna S, Lavalle L, Vicoli G. Safeguarding information intensive critical infrastructures against novel types of emerging failures. Reliability Engineering and System Safety. Vol. 92, p. 1218–1229, 2007.
11. Nai Fovino I, Carcano A, Masera M, Trombetta A. An experimental investigation of malware attacks on SCADA systems. International Journal of Critical Infrastructure Protection. Vol. 2, p. 139–145, 2009.

12. SWISSGRID: Die Nationale Netzgesellschaft. 2007.
13. Aitel D. Cybersecurity Essentials for Electric Operators. The Electricity Journal. 2013.
14. Slay J, Miller M. Lessons learned from the Maroochy water breach. IFIP International Federation for Information Processing. Vol. 253, p. 73–82, 2008.
15. (FPL) FPaLC. FPL announces preliminary findings of outage investigation. 2008.
16. Christansson H, Luiijf E. Creating a European SCADA Security Testbed. IFIP International Federation for Information Processing. Boston: Springer; p. 237–247, 2007.
17. Zhou L. Forcal Report: Vulnerability Analysis of Industrial Control Systems - Part B: Statistics and analysis of industrial security incidents, Challenges of ICS security research. BABS Report: ETH Zurich, 2011.
18. Trantopoulos K. Focal Report: Vulnerability of Critical Infrastructrures-Rail Transport Switzerland. BABS Report: ETH Zurich, 2010.
19. Kröger W, Nan C, Trantopoulos K, Zhou L, Eusgeld I. Report: Interdependencies. BABS Report: ETH Zurich, 2009.
20. Railroad Accident Brief: Accident DCA-01-MR-004. In: Board USNTS, editor.NTSB/RAB-04/08.
21. Rinaldi SM, Peerenboom JP, Kelly TK. Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies. IEEE Control Systems Magazine. 2001; Vol. 21: p. 11–25.
22. Griot C. Modelling and simulation for critical infrastructure interdependency assessment: a meta-review for model characterisation. International Journal of Critical Infrastructure. Vol. 6, p. 363–379, 2010.
23. Pederson P, Dudenhoeffer D, Hartly S, Permann M. Critical Infrastructure Interdependency Modeling: A Survey of U.S and International Research. Idaho National Laboratory; 2006.
24. Zimmerman R. Decision-making and the vulnerability of interdependent critical infrastructure. IEEE International Conference on Systems, Man and Cybernetics. p. 4059–4063, 2004.
25. International Risk Governance Council. Policy Brief: Managing and reducing social vulnerabilities from coupled critical infrastructures. Geneva, Switzerland: IRGC; 2007.
26. Kröger W. Critical infrastructure at risk: A Need For A New Conceptual Approach and Extended Analytical Tools. Reliability Engineering and System Safety. Vol. 93, p. 1781–1787, 2008.
27. Eusgeld I, Nan C, Dietz S. "System-of-systems" Approach for Interdependent Critical Infrastructures. Reliability Engineering and System Safety. Vol. 96, p. 679–686, 2011.
28. Leontief WW. Input-output economics. 2nd Ed ed: Oxford University Press, New York; 1986.
29. Setola R, De Porcellinis S, Sforna M. Critical infrastructure dependency assessment using the input-output inoperability model. International Journal of Critical Infrastructure Protection. Vol. 2, p. 170–178, 2009.
30. Haimes YY, Horowitz BM, Lambert JH, Santos JR, Lian C, Crowther KG. Inoperability Input-Output Model for Interdependent Infrastructure Sectors. I: Theory and Methodology. Journal of Infrastructure Systems. Vol. 11, p. 67–79, 2005.
31. Haimes YY, Horowitz BM, Lambert JH, Santos J, Crowther K, Lian C. Inoperability Input-Output Model for Interdependent Infrastructure Sectors. II: Case Studies. Journal of Infrastructure Systems. Vol. 11, p. 80–92, 2005.
32. Trantopoulos K. Focal Report: Methods for the Vulnerability Assessment of Multi-layer Infrastructure Networks-The Swiss Rail System. BABS Report: ETH Zurich, 2011.
33. Klein R, Rome E, Beyel C, Linnemann R, Reinhardt W. Information Modelling and Simulation in large interdependent Critical Infrastructures in IRRIIS. IRRIS Report 2007.
34. Steen Mv. Graph Theory and Complex Networks: An Introduction. 1 ed: Maarten van Steen; 2010.
35. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S. Catastrophic cascade of failures in interdependent networks. Nature. Vol. 464, p. 1025–1028, 2010.
36. Johansson J, Hassel H. An approach for modelling interdependent infrastructures in the context of vulnerability analysis. Reliability Engineering and System Safety. Vol. 95, p. 1335–1344, 2010.

37. Eusgeld I, Nan C. Creating a simulation environment for critical infrastructure interdependencies study. IEEE International Conference on Industrial Engineering and Engineering Management, p. 2104–2108, Hong Kong, 2009.

38. Schläpfer M, Kessler T, Kröger W. Reliability Analysis of Electric Power Systems Using an Object-oriented Hybrid Modeling Approach. 16th power systems computation conference. Glasgow. 2008.

39. Min H-SJ, Beyeler W, Brown T, Son YJ, Jones AT. Toward modeling and simulation of critical national infrastructure interdependencies. IIE Transactions. Vol. 39, p. 57–71, 2007.

40. Di Giorgio A, Liberati F. Interdependency modeling and analysis of critical infrastructures based on Dynamic Bayesian Networks. 19th Mediterranean Conference on Control & Automation (MED), p. 791–797, 2011.

41. HadjSaid N, Tranchita C, Rozel B, Viziteu M, Caire R. Modeling cyber and physical interdependencies - Application in ICT and power grids. Power Systems Conference and Exposition. p. 1–6, 2009.

42. D'Agostino G, Bologna S, Fioriti V, Casalicchio E, Brasca L, Ciapessoni E, et al. Methodologies for inter-dependency assessment. 5th International Conference on Critical Infrastructure (CRIS). p. 1–7, 2010.

43. Casalicchio E, Bologna S, Brasca L, Buschi S, Ciapessoni E, D'Agostino G, et al. Interdependency Assessment in the ICT-PS Network: The MIA Project Results. In: Xenakis C, Wolthusen S, editors. Critical Information Infrastructures Security: Springer, Berlin Heidelberg; p. 1–12, 2011.

44. Fioriti V, D'Agostino G, Bologna S. On Modeling and Measuring Inter-dependencies among Critical Infrastructures. Proceedings of the 2010 Complexity in Engineering: IEEE Computer Society, p. 85–87, 2010.

45. Bloomfield R, Chozos N, Nobles P. Infrastructure interdependency analysis: Introductory research review. 2009.

46. Gorbil G, Gelenbe E. Design of a Mobile Agent-Based Adaptive Communication Middleware for Federations of Critical Infrastructure Simulations. Proceedings of CRITIS 2009. 2009.

47. DoD. Department of Defense (DOD): High Level Architecture Interface Specification. 1998.

48. IEEE. IEEE Standard for Modeling and Simulation High Level Architecture (HLA)—Framework and Rules. IEEE Std 1516–2000, 2000.

49. Hopkinson KM, Giovanini R, Wang XR. EPOCHS: Integrated Commercial Off-the-Shelf Software For Agent-based Electric Power and Communication Simulation. Proceedings of the 2003 Winter Simulation Conference. p. 1158–1166, 2003.

50. Rehtanz C. Autonomous systems and intelligent agents in power system control and operation: Springer; 2003.

51. Lees M, Logan B, Theodoropoulos G. Distributed Simulation of Agent-based Systems with HLA. ACM Transactions on Modeling and Computer, Simulation. Vol. 17(3), 2007.

52. Zhao Z, Albada DV, Sloot P. Agent-Based Flow Control for HLA Components. Simulation. Vol. 81, p. 487–501, 2005.

53. Beeker ER, Page EH. A Case Study of the Development and Use of a MANA-Based Federation for Studying U.S. Border Operations. Proceedings of the 38th Conference on Winter Simulation, p. 841–847, 2006.

54. Lieshout Fv, Cornelissen F, Neuteboom J. Simulating Rail Traffic Safety Systems using HLA 1516. Atos Origin Technical Automation; 2008.

55. Ezell BC. Infrastructure Vulnerability Assessment Model (I-VAM). Risk Analysis. Vol. 27, p. 571–583, 2007.

56. Möller B, Löfstrand B, Lindqvist J, Backlund A, Waller B, Virding R. Gaming and HLA 1516 Interoperability within the Swedish Defense. 2005 Fall Simulation Interoperability Workshop. 2005.

57. Zacharewicz G, Alix T, Vallespir B. Services Modeling and Distributed Simulation DEVS / HLA Supported. Proceedings of the 2009 Winter Simulation Conference (WSC). p. 3023–3035, 2009.

58. Nan C, Eusgeld I. Adopting HLA standard for interdependency study. Reliability Engineering and System Safety. Vol. 96, p. 149–159, 2010.
59. Nan C, Kröger W, Probst P. Exploring critical infrastructure interdependnecy by hybrid simulation approach. ESREL 2011. p. 2483–2491, Troyes, France, 2011.
60. Nan C, Eusgeld I, Kröger W. Analyzing vulnerabilities between SCADA system and SUC due to interdependencies. Reliability Engineering and System Safety. Vol. 113, p. 76–93, 2013.

Chapter 14
Financial Networks

Stefano Battiston and Guido Caldarelli

14.1 Introduction

The financial system performs vital functions for the world economy. Very often one of more aspect of this system can be described by means of a complex graph. In this chapter under the generic name of *financial networks* we indicate several different systems all related to the world of finance. Such a coarse graining is justified by the fact that in all the various situations we always find similar behaviours. We shall present here a series of examples passing from the study of stock-price correlations to the study of the web of exposures between different companies, and finally to the lending of money between banks.

Indeed in every of the abovementioned systems we encounter similar mathematical structures. Furthermore we are interested in similar basic questions. More particularly we always find a scale-free architecture, a scale-free distribution of centrality and betweenness. At the same time, in all these cases we want to know which institutions is more important for the stability of the whole system, what is the global impact of a local bankruptcy, and finally how we can act on the system in order to change its properties or to recover the initial stability.

These questions have similar answers, with details changing from one case to another, but in any case related to the issue of centrality and controllability.

The possibility to provide to regulators a set of simple indicators that can be used as a thermometer of the financial situation in order to prevent crises is one of the most challenging perspectives. For this reason more and more often scientists and research groups involve regulators in the research activity. The recent crisis has spurred a profound debate about the role of policy and regulations in financial markets. The debate has drawn the attention of researchers from many areas of

S. Battiston (✉)
ETH-Zentrum Systemgestaltung, WEV G 201 Weinbergstrasse 56/58, Zürich 8092, Switzerland

G. Caldarelli
IMT Alti Studi Lucca, Piazza San Ponziano 6, 55100 Lucca, Italy

science as well as of the civil society at large to the needs for new approaches to policy modelling. Overall, it has emerged as a prominent societal issue the need to build a sustainable global financial system that serves the global policy goals. In particular, many observers share the view that the current financial crisis should be seen as an opportunity to strengthen climate finance and not as an excuse to postpone the environmental objectives that were previously put forward.

It is well known that in financial markets, while contracts are beneficial to the parties involved they can also entail unforeseen (negative/positive) externalities to other parties [21, 36]. In particular, incentives for parties to take excessive risk as individuals lead to systemic risk for the market as a whole. According to the direct contacts of our Consortium with regulators (e.g., Bank of England, Bank of Italy, Deutsche Bundesbank, DG-Markt) [6, 16], the problem that many regulators are facing today is that (1) it is not clear what externalities could arise from certain contracts and (2) what could be done to mitigate the negative externalities and strengthen the positive ones. The problem is even more acute when linkages with the environmental sphere are introduced via climate finance [28].

The lack of data is the immediate cause of this situation. A first step has been very recently moved in the direction of collecting systematic information for instance for OTC derivatives. In December 2012, the European Commission has adopted new technical standards, the so-called European Market Infrastructure Regulation (EMIR). However, there are more fundamental causes. The continuous injection of financial innovations as well as their inherent complexity makes it difficult to keep track of the possible externalities [28]. In fact, the interactions of market players across the globe through various instruments such as OTC derivatives, security lending and repurchase agreements make, today more than ever, the financial market a complex network [8, 20] with highly non-linear dynamics [6, 15, 22]. Our current understanding of what undesirable systemic effects may arise and how to cope with them is very limited. Progress in this direction is vital for the well-being of the economy and requires combined efforts and competences [35]. This is the topic of the first section of this chapter that focuses on reconstructing the missing links (Fig. 14.1).

Linkages among financial institutions can have ambiguous effects: on the one hand they increase individual profitability and reduce individual risk, but on the other hand they propagate contagion and distress, thus increasing systemic risk (see Fig. 14.2, for an illustration). On this topic a significant body of work has grown in the recent years. The issue of cascade of failures was initially investigated in simple four-node graphs [1] and in ring structures. The fact that too many linkages can be a factor of systemic risk is a result that has emerged in various works, from various underlying mechanisms. In some models, the balance between positive and negative effects of the network density depends on the level of market capitalization [31], in others on the persistence in the dynamics of financial fragility [5], and yet in others on the level of market liquidity [4]. More links in the network can result also in the so-called robust-yet-fragile property, namely the network resists well to most of the shocks but break down entirely in a non-negligible fraction of cases [15]. Moreover, the tendency towards complexity in the instruments and at the same time towards homogeneity in the risk models and investment strategies adopted by banks is per se
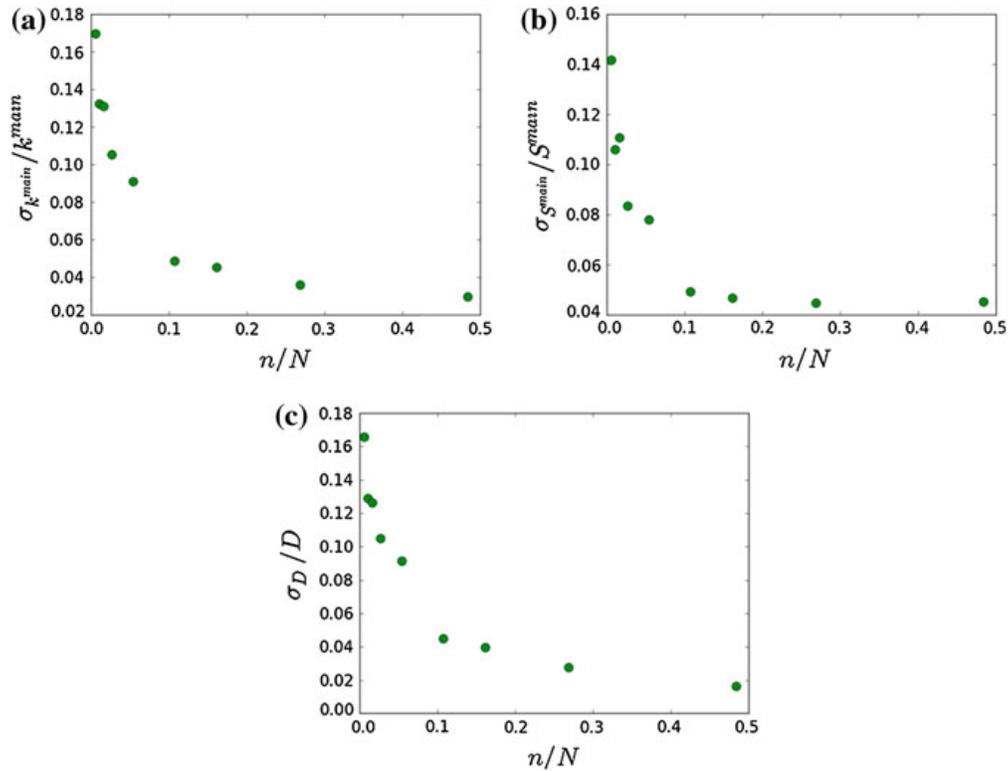
Fig. 14.1 The WTW network. The plots from *top left* represent respectively: **a** the relative error in the estimation of average degree of the main core $\sigma^{k}{}_{main}/k_{main}$ computed with real WTW network **b** the same as in (**a**) but for the relative error in the size of main core, **c** same as in (**a**) but for the density of the links D. In all the 3 plots it is shown that the quality of the reconstruction of the WTW network increases with the number of nodes used to generate the network ensemble

a source of systemic risk [22]. Finally, indirect linkages can emerge from the fact that banks invest in overlapping portfolios, resulting in a tension between the incentive to reduce individual risk and the social cost of having systemic crises.

Following this perspective we present here results that are mostly based on the activity of the FET Open project "Forecasting Financial Crises" a consortium of several European institutions and the European Central Bank whose continuous feedback proved crucial in order to refine our research.

14.2 Reconstructing the Missing Links

One first issue in the study of financial networks is to deal with partial and/or incomplete information. While the reconstruction of the original network may be proven to be very difficult, it has to be noticed though, that in most of the application we need only the statistical properties of it. Addressing this issue has many concrete applications. Typically we can consider a system made of financial institutions as
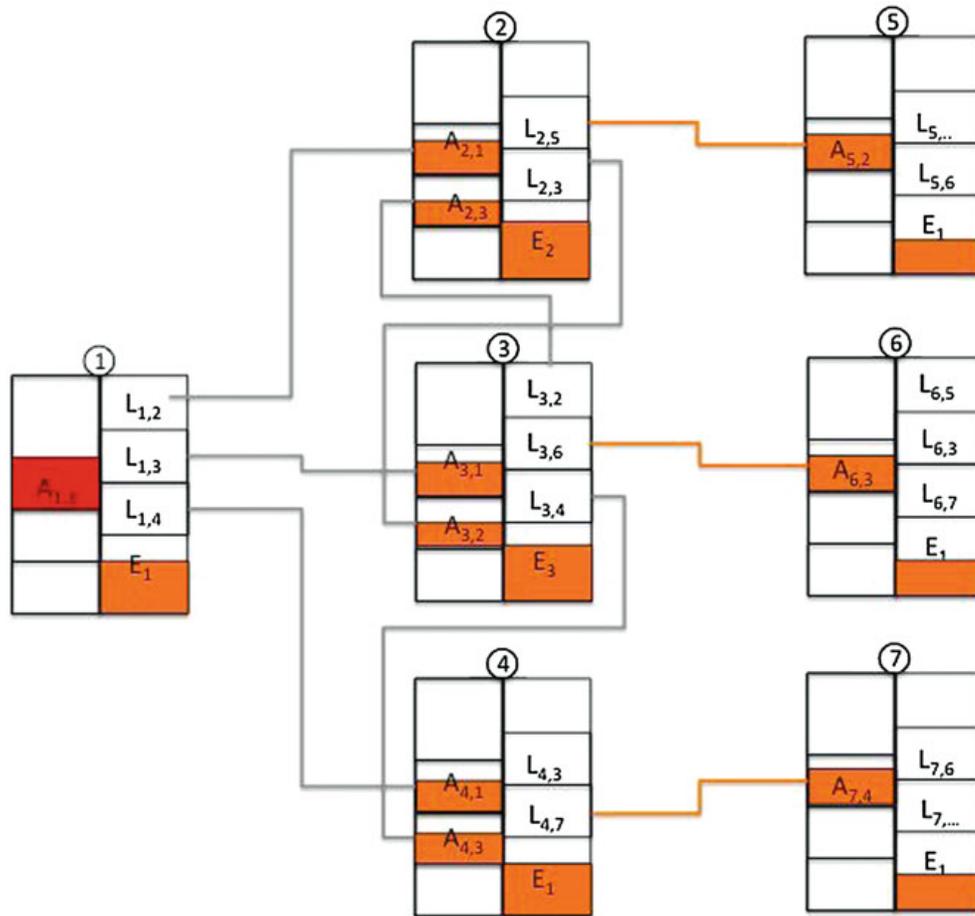
Fig. 14.2 Illustration of distress propagation across a network of banks connected via liabilities. Each block is a schematic balance-sheet of a bank

vertices and edges formed by various kinds of financial ties such as loans or derivative contracts. These ties result in dependencies among institutions and constitute the ground for the propagation of financial distress across the network. The resilience of the system to the default or the distress of one or more institutions depends on the topological structure of the whole network [4]. Unfortunately, the information that regulators are able to collect on the mutual exposures among institutions is very limited (since the confidentiality issues).

Various methods have been presented in order to reconstruct the network in the papers analysing systemic risk. One of the most successful is the so-called Maximum Entropy (ME) algorithm. This method assumes that the network is fully connected (for this reason this class of approaches is called "dense reconstruction methods"). Edges are weighted and these weights are obtained via a maximum homogeneity principle. This means that each node is assumed to bear a similar level of dependence from all other nodes. After that, the method proceeds by looking for the matrix that minimizes the distance from the uniform matrix (where every entry has the same

value), while satisfying certain constraints (imposed in this case by the budget of the individual banks). Such a matrix is found by minimizing an objective mathematical function known as the Kullback-Leibler divergence.

However, the hypothesis of "graph completeness" strongly limits the ME algorithm, since empirical networks show instead a largely heterogeneous degree distribution. Moreover, such "dense reconstruction" leads to an underestimation of the systemic risk [29, 32]. To overcome these limitations a sparse reconstruction" algorithm has been proposed. The procedure is similar to the one for dense graph. Again we minimise the Kullback-Leibler divergence and we obtain a matrix with an arbitrary level of heterogeneity given certain constraints. The latter approach is more reliable but leaves open the question of what value of heterogeneity would be appropriate to choose. Finally the density of connections must be specified ex-ante and it is not recovered by the algorithm.

Recently a third approach wanted to overcome these problems. The new procedure is called Bootstrapping Method (BM) and it is a new general method to deal with incomplete information [30]. This method does not aim at reconstructing the original network but rather to estimate its global properties.

In more detail, among all the possible topological properties, the authors focused on those that in the literature have been shown to play an important role in contagion processes and in the propagation of distress, i.e., the network density and the k-core structure [25]. For the resilience, they focussed on a recently introduced notion, DebtRank [5], that allows to measure the systemic impact of an initial shock on one or more nodes, whenever the links in the network represent the financial dependencies among nodes. It is also possible to determine the accuracy of the estimation upon the size of the subset of nodes for which the information is available.

In this method, the allocation of the links among nodes is carried out using the fitness model ([9, 18]). Differently from other network generation models, the fitness model distinguishes amongst different vertices. In particular it generates a network structure starting from a non-topological variables (fitness) associated to the nodes. This approach has been used in the past to reproduce the topological properties of several empirical economical networks, including the network of equity investments in the stock market [19], the interbank market [12], and the WTW [18].

The validity of this method has been proved on both synthetic networks as well as examples of real economic systems. In these few cases, there is full information on the system. The whole adjacency matrix is available and therefore it is possible to evaluate the accuracy of the method by using only part of the information. The two empirical cases of study presented in [30] are

- the World Trade Web (WTW), i.e. the network in which nodes are countries and links are trade volumes (in US dollars) among them,
- the interbank loan network of the so-called e-mid interbank money market.

The result of this analysis is that information on the degree of a relatively small fraction of nodes is sufficient to estimate with good approximation the above mentioned topological properties, as long as the fitness of all nodes is known.

For instance, with only about 7 % of the nodes (10 out of 185) we have a relative error of about: 7 % on the density, 10 % on the average degree of the main core, 7 % on the size of the main core. Similarly, t with about 7 % of the nodes the resilience can be estimated with a relative error within 10 %.

14.3 Evaluating the Impact of Linkages Through Centrality Measures: DebtRank

During the period March 2008–March 2010 many US and international financial institutions received aid from the US Federal Reserve Bank (FED) through emergency loans programs, including the so-called "FED Discount Window". Recently this dataset has been released thereby providing a unique and important opportunity to study the distribution of debt across institutions and across time.

One of the papers based on the analysis of this dataset wanted to estimate the impact of a node on the others. This is done by means of a novel measure inspired by feedback centrality. Such quantity termed DebtRank [5] takes recursively into account the impact of the distress of an initial node across the whole network. More particularly DebtRank of vertex i, is a number (i.e. dollars or euros) measuring the fraction of the total economic value in the network potentially affected by the distress or the default of node i. This quantity can be used to construct a ranking, but it is not itself a particular rank of the node considered.

Its computation differs from the methods based on the default cascade dynamics [4, 11, 29] in which, below the threshold no impact is propagated to the neighbors. In this respect DebtRank is more similar to other feedback centrality measure that have found successful applications in many domains ranging from rankings in the world-wide-web (e.g. PageRank) to corporate control in economic networks.

Feedback centrality can be considered as the in-flow in a non-homogeneous diffusion process. Exactly in this spirit the presence of a cycle in the network represent a potential problem. In this case there is an infinite number of reverberations of the impact of a node to the others and back to itself, which leads to no simple and measurable economic interpretation. DebtRank overcomes this problem by only allowing for walks that do not visit the same edge twice.

Consider a directed network where the nodes represent institutions and the links represent financial dependencies.

- Denote the amount invested by i in the funding of j as A_{ij}. Thus, A is the weighted adjacency matrix of the investment network. The total value of the asset invested by i in funding activities is $A_i = \Sigma_j A_{ij}$.
- Denote by E_i the tier capital of i giving the buffer of i against shocks. If $E_i < c$ (where c is a positive threshold) the firm defaults.

If the node i defaults, the node j faces a loss of A_{ji} (in the first instance, any recovery is excluded). Similarly also the node j defaults if $A_{ji} > E_j$. When the loss
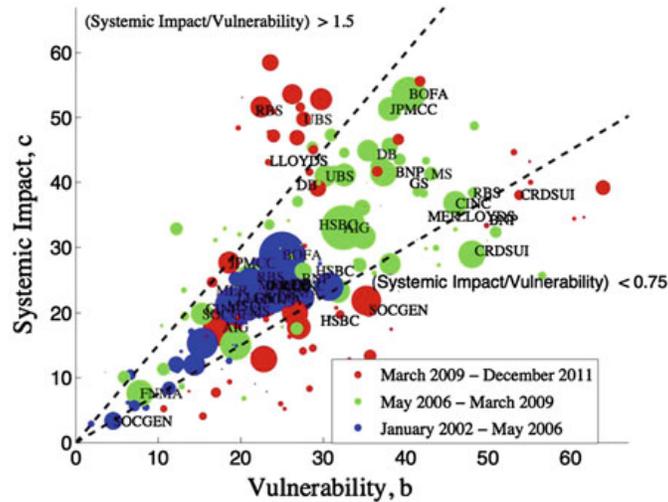
Fig. 14.3 DebtRank-like algorithms allow to monitor over time in a 2-dimensional plot those players that have at the same time high impact on the others and high vulnerability to other players' shocks (see [24] for more details on the calculations). As we can see, in the intermediate period (*green*) a number of players were at the same time highly vulnerable and systemically important

exceeds the capital the impact is fixed to 1, so that in general the impact of i on j can be defined as $W_{ij}=\min\{\,1,A_{ji}/E_j\,\}$. The total amount of the impact of institution i is $I_i=\Sigma_j\,v_j W_{ij}$.

The problem is to take into account the impact of i on its indirect successors, that is, the nodes that can be impacted from i at distance 2 or more.

Authors define an iterative equation of kind

$$I_i=\Sigma_j v_j W_{ij} + \beta \Sigma_j W_{ij} I_j$$

where the second term accounts for the indirect impact via the neighbours. The parameter β is a dampening factor.

In a cycle ($W_{ij} > 0$ and $W_{ji} > 0$), the impact of node i to j hits back on i and keeps cycling an infinite number of times (although with dampening). A single reverberation of the impact of i back to itself is realistic and mathematically acceptable. Further reverberations lead instead to an inconsistency because the impact could become larger than one. The reason is that if the impact is repeated several times through a cycle, then the impact of a node on another one is counted more than once. The same problem applies also to any cycle not involving i, but located downstream of i in the network. Removing the cycles altogether from the network and considering its corresponding acyclic graph would remove entirely the reverberation and cut many links, thus strongly underestimating the impact. In contrast, DebtRank is computed on the original network by excluding the walks in which one or more edges are repeated (Fig. 14.3).

14.4 Interbank Controllability

As shown above the network theory can provide some suggestions to improve the stability of financial systmes against crises. Especially in the credit sector, scientists tried to emphasize the systemic implications of distress in economic systems [1, 14]. Indeed, the fragility in specific countries, markets and financial institutions can propagate and damage the whole economy [2, 33]. For this reason an increasing research effort is being invested in the study of economic and financial networks [3, 4]. Interestingly, many (practically relevant) properties of these networks can be quantitatively investigated. For example, it has been widely recognized the role of Too-Big-To-Fail (TBTF) [35] hubs in determining the fragility of the system with respect to distress propagation or its resilience against link failures. The connectivity structure plays a fundamental role in it [1]. Against this background, Delpini et al. [13] investigated which (if any) policy could improve the stability of the network toward a less risky situation.

The basis of this work is to apply to the interbank money market the notion of *controllability structural* [26, 27] based on control theory concepts. The idea is that the whole network can be controlled by acting on specific drivers. An immediate application arises when a central bank must give credit to the banking system and therefore needs to know which actions (amongst many) are likely to affect the whole structure. The problem of finding the driver vertices of the system is completely solved by finding a maximum matching of the corresponding oriented graph [27].

Since the matching does not depend on the specific values of the weights (which are by the way largely unknown or affected by errors for the majority of real networks), the results hold in a variety of different situations explored with reference to the Italian case.

In particular it is possible to assess the controllability of interbank money markets empirically, focusing on the specific case of the Italian electronic trading system (e-MID), which is open to European banking players, and for which a time series of micro data is available [13]. Following the network evolution over time it is possible to detect the banks that are more relevant from a control perspective. For them, the authors considered the changes of the relevant topological and financial quantities, clarifying the role played by drivers in this system. Through this approach it is also possible to address the resilience of the network drivers, that is the correlation between driver sets at different times. At every aggregation scale and for every available network instance, it is possible to identify the set of driver nodes (Fig. 14.4a shows a daily network instance, with the driver nodes highlighted).

The number of maximum matchings for directed graphs of the size of the Italian interbank market is rather large. Searching for a maximum matching takes into account the graph edges only and enumerating all the possible control configurations is an intensive numerical task. Nevertheless, this can be simplified by assuming that some configurations may be more significant than others from a control perspective. Delpini et al. [13] selected the maximum matching with maximum weight (the weight being given by the sum of edge weights in the matching). The intuition behind
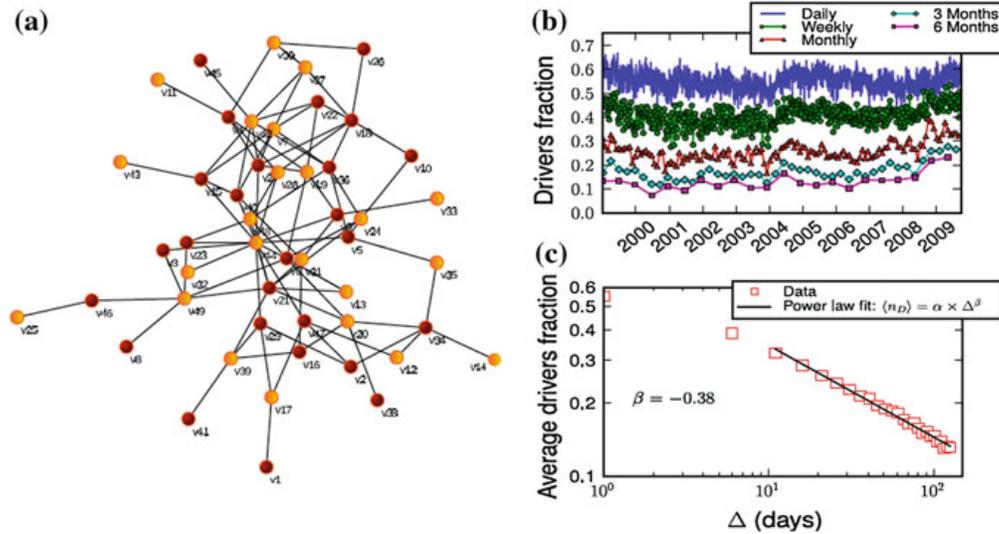
Fig. 14.4 **a** A sample snapshot of the daily interbank lending network. External inputs on the *yellow* nodes (drivers) allow to control the state of the whole system. **b** Time evolution of the fraction of drivers: at the monthly scale less than 40 % of the banks drive the system. **c** The average fraction of drivers decays with a neat power law scaling with the aggregation scale Δ

this choice is that cash flows are proxies for the strength of the influence relationship between two banks.

The main results are that there is no characteristic time scale for the fraction of drivers in the system. In other words it is not possible to select an optimal aggregation time. Rather, different levels of aggregation correspond to networks with different connectivity, which requires different control strategies. Different scales could serve different supervision purposes and policy makers could adopt the instruments that are more effective for the time horizon of interest for the control. The control set of drivers will change over time. However, inspection of the drivers resilience, shows also that the system is characterized by long-range memory. The survival function has a very slow, almost linear decay, and after 6 months nearly 60 % of drivers are still in the control set. Surprisingly, the level of the curve does not follow a monotonic trend with Δ: the control configuration is maximally stable at the monthly scale. This result supports the proposal of the Basel III Committee to introduce a 30-day liquidity coverage ratio, and suggests the monthly scale as a reasonable time window for observing the system.

This is important, since recent financial crisis has been forcing central banks to implement aggressive and creative policy actions. Radically new strategies have been proposed to cope with liquidity shocks within interbank markets. Traditionally, policies have been mainly based on liquidity injection through open market operations, but it has been proved that targeted intervention on individual banks could be more effective in guaranteeing and restoring the efficient allocation of credit. The above findings suggests the need for monitoring the system and keeping track of banks that are systemically relevant from a control perspective [13]. Since no characteristic

scale exists in the decay of the fraction of drivers with the time resolution, this implies that no optimal timing for bank supervision can be selected based only on that. Nevertheless, other network statistics, such as the persistence of control configurations, indicate the monthly scale as natural for observing the system. Another result of this analysis is that the more relevant banks to the overall state of the credit network are neither the most connected nor the top lenders. This strongly suggests the necessity to rethink the policies based exclusively on the TBTF specification of a systemically important institution. According to recent regulation proposals, the ECB has recognized that allowing recapitalization interventions directly on individual banks is a necessary procedure.

References

1. Allen, F. and Gale, D. "Financial Contagion" *Journal of Political Economy* **108**, 1–33 (2001).
2. Barabási, A.-L. Scale-Free Networks: A Decade and Beyond. *Science* **325**, 412–413 (2009).
3. Barabási, A.-L. "The network takeover" *Nat. Phys.* **8**, 14–16 (2012).
4. Battiston, S., Gatti, D., Gallegati, M., Greenwald, B., Stiglitz, J. "Liaisons dangereuses: increasing connectivity, risk sharing, and systemic risk", *J. Econ. Dyn. Control* **36**, 1121–1141 (2012a).
5. Battiston, S., Puliga, M., Kaushik, R., Tasca, P. & Caldarelli, G. "DebtRank: too central to fail? Financial networks, the FED and systemic risk". *Sci. Rep.* **2**, 541 (2012b).
6. Battiston S., Caldarelli G, Georg C-P, May R., Stiglitz J. "Complex Derivatives" *Nature Physics* **9** 123 (2013).
7. Betz, F. and Peltonen, T. "Tail dependence and Systemic Risk in European Banking", ECB mimeo (2012).
8. Caldarelli G., "Scale-Free Networks" *Oxford University Press* (2007).
9. Caldarelli, G., Capocci, A., Rios, De, L., Paolo., & Muñoz. Miguel a. scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters*, **89**(25), 258702 (2002).
10. Caldarelli G., Chessa A., Gabrielli A., Pammolli F., Puliga M. "Reconstructing a Credit Network" *Nature Physics* **9** 125 (2013).
11. Cont, R., Moussa, A. & Santos, E. B. "Network structure and systemic risk in banking systems" SSRN W.P. series (2010).
12. De Masi, G., Iori, G. & Caldarelli, G. Fitness model for the Italian interbank money market. *Phys. Rev. E* **74**, 066112 (2006).
13. Delpini, D., Battiston S., Riccaboni M., Gabbi. G., Pamolli F., Caldarelli G. "Evolution of Controllability in Interbank Networks" *Sci. Rep.* **3**, 1626; 2013, DOI:10.1038/srep01626.
14. Freixas, X. "Monetary policy in a systemic crisis". *Oxford Review of Economic Policy* **25**, 630–653 (2009).
15. Gai, P., Kapadia S. "Contagion in financial networks" *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **466** 2401–2423 (2011).
16. Galbiati M., Delpini D., Battiston S. "The power to control" *Nature Physics* **9**, 126 (2013).
17. Garlaschelli, D., Battiston, S., Castri, M., Servedio, V. D. P. & Caldarelli, G. "The scale-free topology of market investments" *Physica A* **350**, 2–4 (2003).

18. Garlaschelli, D., Loffredo, M. "Fitness-dependent topological properties of the world trade web" *Phys. Rev. Lett.* **93**, 188,701 (2004).
19. Garlaschelli, D., Battiston, S., Castri, M., Servedio, V., Caldarelli, G. "The scale-free topology of market investments" *Physica A* **350**, 491–499 (2005).
20. Garlaschelli, D., Capocci, A. Caldarelli, G. "Self-Organised Network Evolution coupled to Extremal Dynamics" **Nature Physics 3** 813–817 (2007).
21. Greenwald, B, C, and J E Stiglitz. "Externalities in economies with imperfect information and incomplete markets" *The Quarterly Journal of Economics* **101**, 229–264 (1986).
22. Haldane, A. G. & May, R. M. Systemic risk in banking ecosystems. *Nature* **469**, 351–355 (2011).
23. Kapoor, S. and Oksnes L., Hogarth, R., Green European Foundation (GEF) and Re-Define "Funding the green New Deal: Building a green financial system. A policy maker report from re-define" (2011).
24. Kaushik, R, and Battiston. S. "Credit Default Swaps and Financial Networks". arXiv:1205.0976 (2012).
25. Kitsak, M., Gallos, L., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H., Makse, H. "Identification of influential spreaders in complex networks" *Nat. Phys.* **6**, 888–893 (2010).
26. Lin, C. T. Structural Controllability. *IEEE Trans. Automat. Contr.* **19**, 201–208 (1974).
27. Liu, Y., Slotine, J., & Barabasi, A. Nature **473**, 167 (2011).
28. Mandel, A. "An index formula for production economies with externalities". *Journal of Mathematical Economics* **44** 1385–1397 (2008).
29. Mistrulli, P.: "Assessing financial contagion in the interbank market: maximum entropy versus observed interbank lending patterns" *J. Bank. Finance* **35**, 1114–1127 (2011).
30. Musmeci N., Battiston S., Caldarelli G., Puliga M., Gabrielli A. *Journal of Statistical Physics* **151**, 720–734 (2013).
31. Nier, E., Yang, J., Yorulmazerm T. and Alentorn, A. "Network models and financial stability" *Bank of England Working Paper No.* 346, April (2008).
32. Park, J., Newman, M. "Statistical mechanics of networks" *Phys. Rev. E* **70**, 066117 (2004).
33. Podobnik, B., Horvatic, D., Petersen, A. M., Urosevic, B. & Stanley, H. E. "Bankruptcy risk model and empirical tests". *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18325–30 (2010).
34. Schweitzer, F. et al. "Economic networks: the new challenges". *Science* **325**, 422–5 (2009).
35. Stern, G. H. & Feldman, R. J. "Too Big To Fail" (Brookings Institution Press, Washington, (2004).
36. Stiglitz, J E. "Capital market liberalization, economic growth, and instability". *World development* **28** 1075–1086 (2000).
37. Stiglitz, Joseph E. 2013. "Climate change and poverty have not gone away". *The Guardian*, Monday 7 January.

Chapter 15
Spatial-Temporal Quantification of Interdependencies Across Infrastructure Networks

Christopher Chan and Leonardo Dueñas-Osorio

Abstract As infrastructure networks become more complex and intertwined, the relevance of network interdependency research is increasingly evident. Interconnected networks bring about efficiencies during normal operations but also come with risks of cascading failures with disaster events. An adequate understanding of network interdependencies and realistic multi-system modeling capabilities enable the exploration of practical operation strategies and mitigation efforts applicable to existing or future coupled networked systems. This chapter examines recent efforts in quantifying infrastructure network interdependencies through spatial and time-series analyses to reveal the heterogeneity and complexity in their coupling. Furthermore, a combined spatial-temporal methodology is recommended for the future calibration and validation of theoretical and computational models of interdependent networks of networks. An example case study is demonstrated using data derived from the 2010 Chilean Earthquake in the Talcahuano-Concepción region, which highlights the richness in coupling strengths across infrastructure systems, both as a function of time and geographical extent. Insights for design and control of coupled networks are also derivable from joint spatial-temporal analyses of infrastructure interdependence and its evolution.

15.1 Introduction

From the World Wide Web to the national power grid, networks are an essential part of the world. Appearing in almost all aspects of modern society, a network connects individual components, or nodes, with links that join together multiple

C. Chan (✉)
Stanford University, 450 Serra Mall, Stanford, USA
e-mail: chris.chan@stanford.edu

L. Dueñas-Osorio
Rice University, 6100 Main Street, Houston, USA
e-mail: leonardo.duenas-osorio@rice.edu

nodes in a systemized network. Common networks include public transportation systems, electrical power grids, water systems, and intertwined social circles [14, 18]—examples that are frequently seen in everyday life. Hence, a more complex network can be defined as the coupled interaction of those individual networks with each other; for example, much of a city's service infrastructures, like the water supply network and subway systems, are dependent upon the greater regional power network, which may be contingent upon functioning telecommunications channels which govern the operations of a series of other networks and infrastructure systems [22, 34]. Similarly in fields differing from engineering, biochemically linked neurons in the body work both synchronously and in tandem with the other biological systems to sustain even greater functions [2, 4]. These interdependencies between networks create a complex web of networks of networks linked by both connectivity and dependency [13, 17, 29], which significantly increases efficiency, but also introduces greater risks in network security and reliability.

Network interdependency has especial relevance within infrastructure systems. As the world continues to urbanize, essential infrastructures have become increasingly interconnected and mutually dependent with new tech0nological advances [3, 13, 34]. Any disruption with critical infrastructure can result in what is known as cascading failures [6, 15, 36, 42] in which one failure causes a chain event result. On July 30 and 31 of 2012, more than 700 million people, roughly a tenth of the world's population, were plunged into darkness in northern India, as the three interdependent state power grids crippled one after the other [20]. The world's largest blackout paralyzed the interstate train system, affected local health services, and trapped 200 miners among other consequences [39]. The magnitude of the incident lucidly demonstrated the criticality of linked networks and manifested the extent that interdependent infrastructures can impact the world.

The inherent nature of increasingly connected and interdependent infrastructure systems implies that there will continue to be even greater risks and vulnerabilities during operations as well as even more pronounced repercussions in the event of external threats. Such dangers ranging from natural disasters to terrorist attacks can put the connected network of networks at risk of a falling domino effect. In fact, the reliability of critical infrastructure security became such an issue in the United States that in 1996 President Clinton issued an executive order to establish the President's Commission on Critical Infrastructure Protection (PCCIP) [8, 21, 30]. The importance of critical infrastructures means that it is necessary to not only model the complexities of infrastructure networks, but also quantify the inherent risk associated with interdependency.

As a result, much of recent research in the past decade has been devoted to examining and understanding infrastructure network interdependencies [34]. This can be done through a host of strategies, for example, identifying the mechanisms of interaction (physical, logistical, geographical, etc.), observing the interdependent characteristics (operational, spatial, temporal, etc.), and quantifying the coupling and response behaviors [24, 32]. This chapter will look at research and models to quantify the interdependencies of infrastructure networks and infer their potential effects on system performance. Models and simulations in the past have employed a variety

of techniques to measure and capture the complex interactions in interdependent networks with a mixture of approaches deriving from complexity and network theories, economic methods, probabilistic analyses, and data-driven approaches. For example, frequency analyses of infrastructure failure propagation incidents after hazards have been studied as a methodology for characterizing and empirically quantifying network interdependencies [7, 26]. From theoretical and simulation-based approaches, sandpile dynamics have been used with a multi-type branching process to analyze cascading loads in connected networks of different topologies [5]. With the increase in computing power, complex adaptive systems (CAS) have also been used to model interdependent networks as individual intelligent agents which cooperate and compete in the larger system. The agent-based CAS modeling could use sensors in the system to prevent cascading failures and is applied frequently among operational and socioeconomic networks [1]. Borrowing off of financial markets modeling, the Leontief economic paradigm, when applied to infrastructure networks, is an input–output model that uses an interdependence matrix to compute shared risk of inoperability of infrastructure systems [21]. Restoration of network services has also been modeled using multilevel interdependencies in a mathematical network flows model exploiting advances in operations research [25], while computationally intensive, flow dynamic-based methods that require large data sets have been applied to models of telecommunications, gas, power, and emergency systems [28, 33, 35]. In addition, the graph wavelets approach, which uses the wavelet transform to model changes in the network as a whole, has been used in spatial traffic flow analysis, which has the potential to impact multiple physical and social systems [10]. Finally, network reliability models that use a probabilistic quantification of interdependencies among networks provide flexibility to integrate with network theory and quantify performance correlations between infrastructure systems which provide unique insights for infrastructure engineering practice [23].

While these and many other approaches to modeling interdependent systems exist, for any method, the quantification of coupling, calibration of performance assessment models, and verification of predictions in a sundry of scenarios remains vital to research and practical applications in network interdependency. Recent approaches have utilized time-based analyses of multi-network performance to calculate coupling strengths by using temporal correlations of post-disruption restoration times [16]. At the same time, network interdependency can also be approached by looking at spatial correlation. Studies have taken stochastic external stresses to identify geographic vulnerabilities [31] or utilized kriging techniques in generating spatial correlation [40]. The methodology presented in this chapter allows for a novel approach to quantify joint spatial-temporal network correlations and reveals the heterogeneity in the interdependencies of infrastructure systems that simultaneously take into account the time-dependent and geographic relevance of the networks.

The following sections will briefly overview approaches that have been used to quantify interdependence, specifically with regard to time- and space-dependent methods and, furthermore, describe the unique spatial-temporal approach recommended by this study. After the methodological discussions, this chapter will focus on the application of the spatial-temporal approach to representations of data derived

from post-event analysis of the 2010 moment magnitude (M_W) 8.8 Chilean earthquake, and finally, the analysis of results and synthesis of insights will be followed by conclusions and suggestions for future work on interdependent infrastructure networks research.

15.2 Methodological Approaches to Quantifying Network Interdependencies

15.2.1 Temporal Methodologies

In order to quantify interdependencies across networks and enable the calibration of models of networks of networks, recent studies have taken a time-series approach in analyzing coupling strengths between infrastructure networks. Dueñas-Osorio and Kwasinski [16] explore such an approach by looking at utility service restoration responses. Utilizing post-disaster restoration information from the 2010 Chilean Earthquake, data from individual utility service systems, or lifeline system restoration curves, were collected, showing the gradual restoration of power, water, and telecommunication services available as a function of time. Auto-covariance and autocorrelations of the restoration data were calculated to assess temporal dependencies within the same system. To measure the coupling strength between different utility networks, the cross-correlation $\rho_{j,k}$ (Eq. 15.2) was calculated using the cross-covariances $\gamma_{j,k}$ (Eq. 15.1) for given time lags (or relative times between the restoration curves) using the following equations:

$$\gamma_{j,k}(h) = \frac{1}{1+n_j} \sum_{t=0}^{n_j-h} (x_{t+h,j}\bar{x}_j)(x_{t,k} - \bar{x}_k), \tag{15.1}$$

$$\rho_{j,k}(h) = \frac{\gamma_{j,k}(h)}{\sqrt{\gamma_j(0)\gamma_k(0)}} , \tag{15.2}$$

where $x_{t,j}$ or $x_{t,k}$ is the restoration value at time t of the jth or kth system, n_j is the maximum observation time, and h is the given time lag between the restoration curves of the systems. The cross-correlations provide a convenient dimensionless metric for quantifying interdependencies and analyzing the behavior of the systems across the restoration time series, and in this way, leading or lagging interdependence properties of the networks can be revealed. In order to achieve stationarity and make the time-series analysis tools suitable, the time-series data is transformed and second-differenced before the correlation analysis. The study notes high correlations between power and telecommunications systems (operational interdependency) as well as with water delivery (logistical interdependency). For example, correlation between fixed phone services and the regional power delivery system reaches 0.84 at a lag time of $h = 2$, highlighting the high level of operational interdependency

between power and telecommunications with the latter lagging in restoration. Similarly, strong correlation between power and water systems reaches 0.79, but at a negative lag time of $h = -13$, showing the leading tendency of the power system restoration on the ability of water system operators to coordinate the logistics of damage repairs. Outcomes are only observed several days later after intensive physical tasks of digging, welding, and replacing are completed. A mathematical relationship is then formulated as a measure of overall coupling strength $S_{j,k}$ (Eq. 15.3), reflecting both the time lag and the system correlations and demonstrates a high level of interdependence among power and telecommunication systems in regions with moderate level of damage, as well as strong intra-dependence within systems of the same type. In addition, the study reveals a high degree of infrastructure coupling between neighboring regions where the leading restoration of power and telecommunication systems directly affects closely linked restoration processes of networks in geographically close regions.

$$S_{j,k} = \begin{cases} -\rho_{j,k}(h)/(1 + \sqrt{|h|}) & \text{when } h \neq 0 \\ \rho_{j,k}(h) & \text{when } h = 0 \end{cases} \tag{15.3}$$

Using a time-series post-event interdependence quantification technique and analyzing the autocorrelations and cross-correlations in restoration data across systems, it is possible to not only capture the holistic operational and logistical coupling between two networks after a critical failure, but also identify leading and lagging relationships to improve performance and adopt effective mitigation actions for interdependent systems. The quantification of coupling strengths allows for potential applications in computational and theoretical predictive models as well as in disaster mitigation efforts for infrastructure operations and recovery. In the end, practical applications of quantified coupling strengths can include the exploration of system decentralization or the uncoupling of systems during emergency operations to enhance restoration as well as identification of specific physical or organizational factors affecting restoration rates.

15.2.2 Spatial Methodologies

Another significant dimension in quantifying interdependency, especially for infrastructure systems, is the geographic correlation between network elements and their performance. Spatial proximity among networks holds important relevance in modeling infrastructure interdependencies, most notably in the aftermath of natural disasters, such as earthquakes as demonstrated by Lee and Kiremidjian [27] and Rahnamay-Naeini et al. [31]. In modeling spatially-dependent systems, geostatistical techniques like ordinary point kriging utilize optimal least-squares predictions and can be employed in a probabilistic analysis of infrastructure networks to quantify their interdependencies as distributed in their service area spaces. Wu et al. [40] demonstrate the application of kriging surfaces on utility restoration records and

the calculation of spatial correlations to estimate the spatial distribution of network interdependencies among the lifeline systems of Sect. 15.2.1.

15.2.2.1 Ordinary Point Kriging

Ordinary point kriging is a geostatistical technique recently embraced by the spatial interdependence assessment methodology as a tool for spatial interpolation in the creation of restoration and correlation surfaces. Specific details are further discussed below as they are central to the formulation of the spatial-temporal approach proposed in Sect. 15.2.3 of this chapter.

In the analysis of spatial interdependency, a spatial surface of the infrastructure system parameter (e.g. service restoration) must be created by kriging using the mesh of original evaluation data points. With ordinary point kriging, interpolation of restoration records and their spatial variability necessitates an estimator variogram γ_E (sometimes refered to as a 'semi-variogram' governed by Eq. 15.4). The estimated variogram enables plotting spatial variation versus distance, and a parametric curve can then be fitted to model the spatial data [38]. Commonly used parametric models include the spherical, exponential, and linear models, etc.

$$\gamma_E(d) = \frac{1}{2N(d)} \sum_{i=1}^{N(d)} (z_{x_i} - z_{x_i+d})^2 \,, \tag{15.4}$$

In defining $\gamma_E(d)$, z_{x_i} and z_{x_i+d} are the restoration values at the evaluation points x_i and $x_i + d$, respectively. $N(d)$ is the cardinality of the set of pairs of points within a spatial lag or relative distance of d, and the lag interval d is defined as the Euclidian distance between points x_i and $x_i + d$, where the maximum lag is often set as the mean minimum distance between a given pair in the data set. Using a parametrically fitted variogram model (exponential, in the Chilean data case), the restoration values are then interpolated in a mesh grid by ordinary point kriging, which uses a weighted average of the other neighboring evaluation nodes with weight coefficients λ_i that are estimated by minimizing the mean-square error, or kriging variance, and satisfy the following constraints:

$$\hat{z}_p = \sum_{i=1}^{N} (\lambda_i z_{p_i}) \tag{15.5}$$

$$\sum_{i=1}^{N} (\lambda_i) = 1 \tag{15.6}$$

$$E(\hat{z}_p - z_p) = 0 \tag{15.7}$$

$$E((\hat{z}_p - z_p)^2) = 2 \sum_{i=1}^{N} \lambda_i \gamma(x_i, x_p) - \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \gamma(x_i, x_j) \tag{15.8}$$

where $E(\square)$ is the minimized estimation, N is the number of neighboring evaluation points, z_{p_i}, in the context of infrastructure systems after disruption, is the restoration level at a particular point of interest x_i, while $\hat{z}_p$ and z_p are the estimated interpolated value and true restoration value at the interpolation point of interest p respectively. The function $\gamma(x_i, x_p)$ provides the variogram value between the evaluation point at x_i and the particular point x_p, and $\gamma(x_i, x_j)$ is the variogram value associated with neighboring evaluation points x_i and x_j. Spatial interdependence between the two systems is calculated using cross-correlations among the stationary z_{pi} restoration values at the location of evaluation points x_i of a particular network (known as the reference network) and those of a second network (known as the adjuct network) while using a neighborhood set of $\hat{z}_p$ values around the evaluation points x_i of one system and the corresponding set of zp_{hat} values that are collocated in the other system to establish local interdependence strengths as a function of geographical location. Such interdependence strengths are finally used to create a kriging-based surface of local interdependence across systems, which is heterogeneous in contrast to typically assumed constant values of coupling strengths for all interdependence links.

15.2.2.2 Spatial Applications to Chilean Data

An application of spatial kriging techniques is performed on the post-disaster Chilean data in Wu et al. [40]. Cross-correlation of geographically distributed service restoration times for water and power networks are used first as a proxy for spatial interdependency quantification and mapping of local coupling heterogeneity into correlation surfaces. Then, correlation values across systems are synthesized as a function of relative angle and radial distance between restoration surfaces away from one of the system's set of evaluation nodes. Specifically, by shifting the adjunct system's kriging restoration surface from the surface of the reference system and calculating the correlation values, the spatial interdependency of the two systems can be found by the different cross-correlation and autocorrelation values, the latter being that in which the same network serves as both the reference and adjunct network. Such synthesized correlations are quantified using Pearson's coefficient (a measure of linear correlation) as well as Kendall's tau coefficient (rank correlation). The resulting data is plotted on polar coordinates to represent spatial interdependencies from a suite of relative angles and distances across restoration surfaces, forming global correlation plots. These plots demonstrate certain patterns in the analysis, specifically in the cross-correlation between power as the reference system and water as the adjunct system, where a southwest to northeast spatial directionality of the interdependence is evident, revealing spatial coincidence in the average restoration time trends along the noted geographical corridor, partially due to the topology of the systems and the patterns of damage. (Additional details are provided in Sect. 15.4 regarding directionality trends in spatial interdependencies.) Overall correlation plots are then created by averaging the global correlations across all angles, resulting in initial cross-correlation values as a function of relative distance between

restoration surfaces of 0.483 and 0.284 for Pearson's coefficient and Kendall's tau coefficient, respectively. The average correlation plots also reveal certain characteristics of interdependence, such as the average distance away from evaluation nodes until correlation values become negligible or reach zero, which can be interpreted as a measure for spatial coupling strength decay or interdependence length [40]. These interdependence lengths can readily inform utility operators about optimal system decoupling schemes as well as requirements for temporary back-up systems, given that the reach of interdependence is spatially confined to manageable distances.

15.2.3 Spatial-Temporal Methods

Both the time-series model and the spatial methods offer unique insights into the true nature of interdependent infrastructure networks, but emerging strategies in estimation and modeling of networks of networks attempt to encompass more realistic constraints, as when having both correlations in a single spatial-temporal model. By assessing the appropriateness of a series of assumptions regarding the covariance structure across systems in time and space, including full symmetry, separability and stationarity, a number of methodologies have been proposed that have wide ranging practical applications [19].

In order to combine spatial covariances (or variograms) with temporal ones, a variety of models are available, some of which are suitable for novel applications in infrastructure interdependency assessment problems. In order to be a valid covariance function, the combined variogram must satisfy the condition of positive definiteness, where a matrix remains nonnegative definite for all combinations of space-time coordinates [11, 19]. The Product Model separates temporal and spatial considerations and disregards dependence between space and time covariance, but offers a simple and efficient way to represent a spatial-temporal covariance. The Linear Model simply separates and sums up the two covariances, resulting in only a positive semi-definite function [11]. The simplicity of the Product Model and the Linear Model are often inapplicable to certain real world examples due to their inherent assumptions on separability and positive definiteness. As a result, Cressie and Huang derived a new approach to tackle nonseparable stationary covariance functions. By utilizing Bochner's Theorem and assuming integrability, positive definiteness can be proved using the Fourier transform on the spectral density, thereby creating a class of stationary spatial-temporal covariance functions [9]. However, the complexity of the Cressie and Huang model motivated the development of a more flexible and general Product-Sum Model which combines the simplicity of the product and linear models with the applicability that satisfies the conditions of a viable covariance function. The Product-Sum model enables, for the first time, the practical quantification of interdependencies in time and space and provides a reference for models of networks of networks to use in terms of their necessary input coupling information.

15.2.3.1 The Product-Sum Model

De Cesare et al. [11] promotes a Product-Sum model that, as its name suggests, serves as a simple hybrid between the Linear and Product models in the past. De Iaco [12] further refines a generalized version of the Product-Sum model and explores in depth how to fit data to the spatial-temporal variogram. The practical model introduced to represent spatial-temporal covariances is as follows:

$$Cov_{st}(h_s, h_t) = k_1 Cov_s(h_s) Cov_t(h_t) + k_2 Cov_s(h_s) + k_3 Cov_t(h_t), \qquad (15.9)$$

where Cov_s and Cov_t represent the separate spatial and temporal covariance models respectively, h is the generic lag, which is qualified by the subindices s and t for spatial and temporal, and k_1, k_2, and k_3 are constants determined by the individual variogram sills, defined as the plateauing γ limit of the variogram approximating the population variance [38]. The equivalent combined function using variograms is as follows:

$$\gamma_{s,t}(h_s, h_t) = (k_2 + k_1 C_t(0))\gamma_s(h_s) + (k_3 + k_1 C_s(0))\gamma_t(h_t) - k_1 \gamma_s(h_s)\gamma_t(h_t), \qquad (15.10)$$

where $\gamma_{s,t}$ represents the spatial-temporal variogram and γ_s and γ_t are the respective spatial and temporal variogram models. In addition, C_s, C_t, and C_{st} are the sills, estimated from the parametric curve corresponding to each bounded variogram. Each variogram is found by the following general equation, where $\gamma_s(h_s) = \gamma_{s,t}(h_s, h_t = 0)$ and $\gamma_s(h_t) = \gamma_{s,t}(h_s = 0, h_t)$:

$$\gamma_{s,t}(h_s, h_t) = \frac{Var(Z(s + h_s, t + h_t), Z(s, t))}{2}, \qquad (15.11)$$

where Z is a second order stationary spatial-temporal random field of the restoration levels and s, t are space and time values in the respective domains. By using the sills to calculate the three coefficients, positive definiteness can be guaranteed when $k_1 > 0$, $k_2 \geq 0$ and $k_3 \geq 0$, and the following expressions are used:

$$k_1 = [C_s(0) + C_t(0) - C_{st}(0, 0)]/C_s(0)C_t(0) \qquad (15.12)$$
$$k_2 = [C_{st}(0, 0) - C_t(0)]/C_s(0) \qquad (15.13)$$
$$k_3 = [C_{st}(0, 0) - C_s(0)]/C_t(0) . \qquad (15.14)$$

Spatial-temporal methods already have widespread applications across environmental monitoring and modeling, ranging from air pollution to wind speed monitoring, but the concepts can be adopted and customized to infrastructure network models as well. This chapter demonstrates next the relevance of the Product-Sum method to representations of both temporal and spatial data derived from utility restoration processes after the 2010 M_W 8.8 Chile Earthquake.

15.3 Application of the Spatial-Temporal Methodology to the 2010 Chilean Earthquake

The introduced spatial-temporal methodology is applied onto representations of lifeline systems restoration data throughout time and space derived from field-collected nodal values after the 2010 Chilean Earthquake [37]. A total of 93 evaluation nodes covering the service area of two infrastructure networks are used in the study, shown on the map in Fig. 15.1. The level of restoration at each of the evaluation nodes is represented as the percent of fully restored services provided by each of the power and water lifeline networks, which reflects both the level of damage inflicted on the node as well as the prioritization of the post-event restoration procedure. The temporal data at each node is assumed to follow the shape of restoration curves found in the local power and water delivery networks in the Talcahuano and Concepción region [16].

The joint representations of temporal-spatial data allows for the creation of a spatial-temporal variogram (Fig. 15.2) using the Product-Sum Model, applied to infrastructure lifeline system restoration in space and time. A given point on the surface of the spatial-temporal variogram can be interpreted as the restoration variance at a given spatial and temporal lag combination. Using a set of spatially-dependent, temporally-dependent, and spatial-temporal variograms on the data, the sills corresponding to each variogram are estimated using an exponential model variogram fit. For example, the spatial-temporal sill $C_{s,t}$ associated with the variogram shown in Fig. 15.2 is 0.1484. These values are used in the determination of the constants k_1, k_2, and k_3 according to Eqs. 15.12–15.14 to guarantee positive definiteness. The resulting spatial-temporal variogram offers a joint variability for every combination of lag in space and time.

Kriging is then applied using the spatial-temporal variogram by cutting the variogram surface such that spatial correlation or local interdependency can be observed in slices for each time lag. In order to calculate such local interdependencies, a local mesh of 101 points is constructed by kriging around each node for each system. Then, a correlation analysis can be performed between the water and power network local meshes (without relative displacements) for a given time, yielding a measure of local coupling strength in a particular service area. The Pearson's correlation values calculated for each node of the restoration surfaces are then kriged once again to depict the local distribution of correlation in the region per unit of time (Fig. 15.3). By translating the kriging restoration surfaces by a certain radial distance and angle (e.g., a radial mesh of 20 angle increments and 500 m radial increments up to 2,500 m away from evaluation nodes per restoration surface), the correlation of the restoration levels between the reference and adjunct networks can also be analyzed. Autocorrelation represents the dependency within the same network, while cross-correlation refers to the interdependency between different networks.

The shifting of one network surface from another reference network surface is performed not only for distinct lags in space but also time. The translations in time and space of the adjunct network surface will yield global correlation plots, which
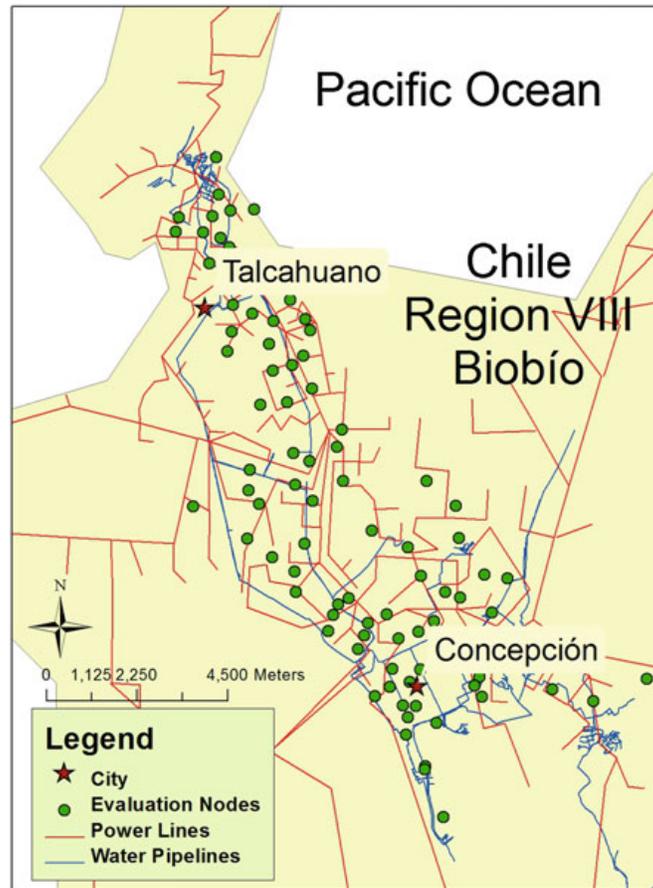
Fig. 15.1 Map of the evaluation nodes from the Talcahuano and Concepción region during the 2010 Chilean Earthquake along with streamlined transmission level power and water networks

can also be averaged to find overall interdependencies and their lengths of influence
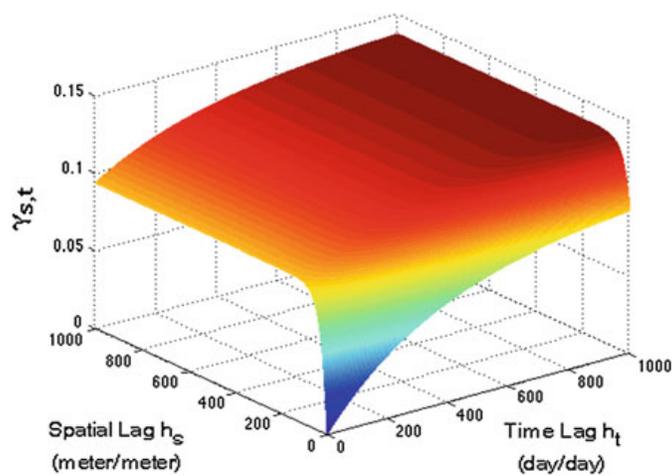

Fig. 15.2 The spatial-temporal variogram surface of the water system as a function of temporal and spatial lag normalized by the maximum value in each data set
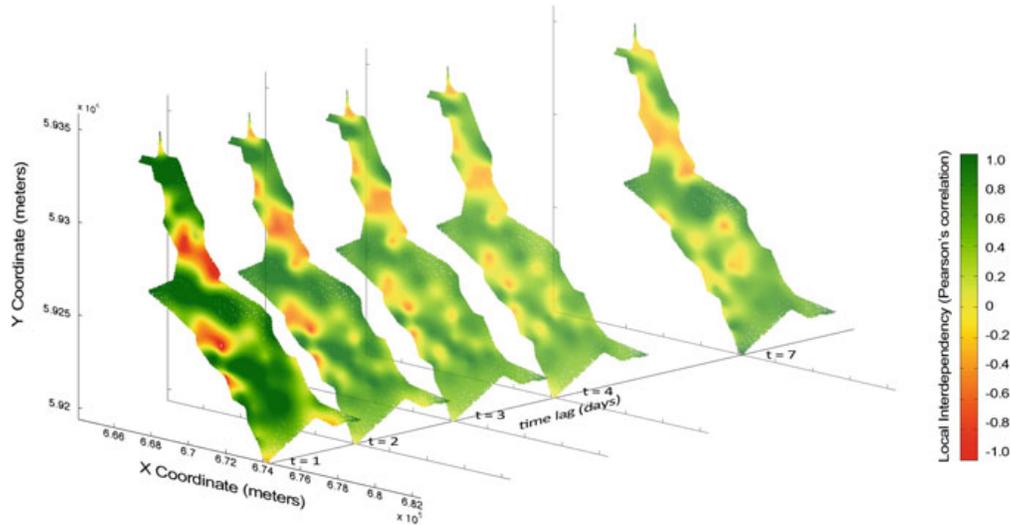
Fig. 15.3 Evolution of local interdependencies across space and time in the Talcahuano-Concepción region of Chile

as a function of time, as discussed in the next section. Note for now that Fig. 15.3 is the first depiction of interdependence evolution in time as a function of spatial location. The local correlation map highlights the richness and heterogeneity of the interdependencies, which differ from assumptions in early models of networks of networks that use homogeneous and static coupling strengths.

15.4 Analysis and Discussion of Synthesized Interdependencies

From the temporal evolution of the local correlations or interdependencies in the region (Fig. 15.3), it is possible to view a generalized summary of the spatial-temporal coupling between the power and water networks. What appears to be highly localized negatively or positively correlated regions in the map corresponds to local circumstances, demonstrating the spread of correlations across space, dependent upon the physical location of the networks, their state of damage, and adopted restoration strategies. The correlations of these areas are still evident through time, but the heterogeneity decreases as the time lag increases, verifying the intuitive result that correlation due to location diminishes as restoration levels of nodes reach higher and higher serviceability, thus reducing variability across them. After about a week after the earthquake event, there is a notable decrease in negative local correlations between the power and water systems, but as the majority nodes begin to reach full restoration, interdependencies are still visible although at reduced strengths.

Global correlation plots (Fig. 15.4) are created by displacing entire restoration surfaces relative to each other in space, so as to show spatial correlations in polar coordinates across time steps as a function radial distance and angle from
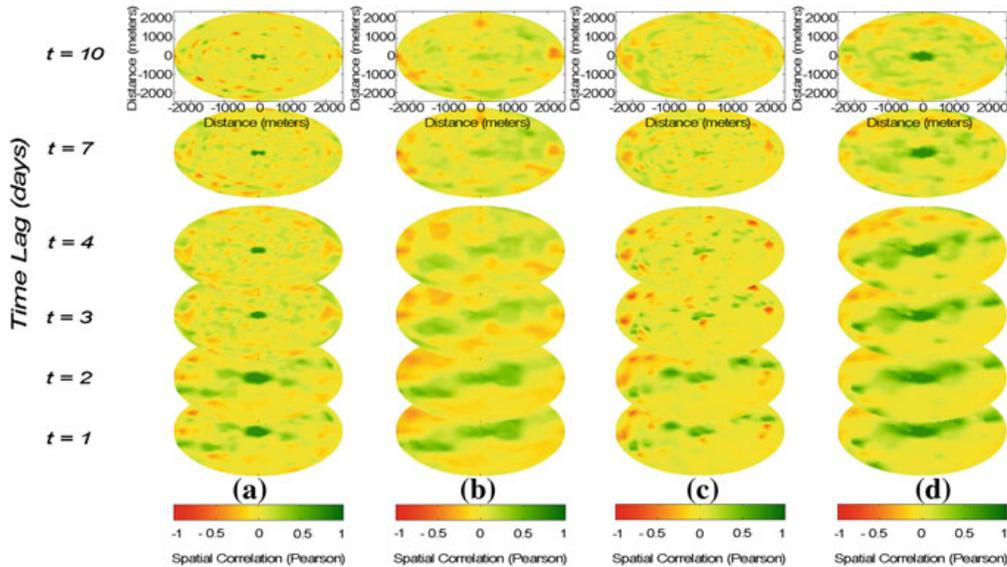
Fig. 15.4 Global correlation plots (with North taken in the upwards direction) synthesizing restoration surfaces between and across systems as a function of time as well as a function of radial distance and angle shown in four stacks depicting: (**a**) the power–power auto-correlation, (**b**) water-power cross-correlation (power as the reference network), (**c**) power-water cross-correlation (water as the reference network), and (**d**) the water–water auto-correlation

the shifted surfaces. The resulting polar-coordinate plots measure the global interdependence across translated maps and reveal the distance and direction in which interdependencies matter. The global correlation plots are presented in four stacks of adjunct-reference network pairs, depicting from left to right: the power–power auto-correlation, water-power cross-correlation (power as the reference network), power-water cross-correlation (water as the reference network), and the water–water auto-correlation. The auto-correlation plots start from a synthesized correlation of 1.0 at the center, since it intuitively represents the same network data. For all the plots, there are signs of directionality at earlier time steps, hinting at general correlations northeast or southwest of evaluation node sets. This may be due to the inherent shape of the analyzed region and associated networks, as well as the distribution of damage and the location of their main components [41]. However, traces of directionality are not evident as time lags increase. The two auto-correlations exhibit unique temporal trends, where the power auto-correlation moves from a central cluster of positive correlation to a relatively uncorrelated neighborhood, while the water network has a much larger neighborhood of high positive correlation that maintains directionality and a certain level of spatial correlation in all directions within a radial proximity of less than 500 m as time progresses. The auto-correlation of the water system (Fig. 15.4, Part d) shows the strongest directionality in recovery as the water system has a concentration of pumping stations and tanks, along with the main water treatment plant in the southwest to northeast direction. Cross-correlation plots are clearly anisotropic, particularly for large lag times, but reflect similar initial correlations
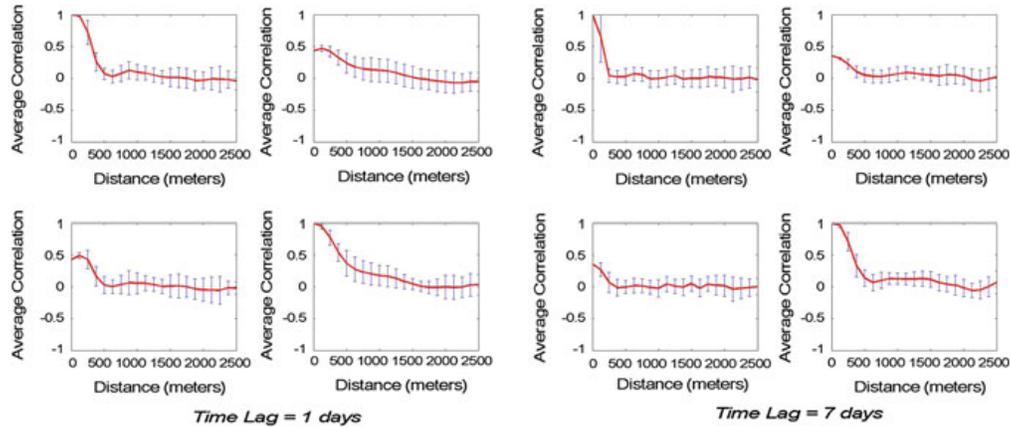
Fig. 15.5 Global interdependence plots of overall Pearson's correlation averaged over all angles as a function of radial distance at a time lag of 1 and 7 days. For each time lag, the plots are subdivided into four subplots depicting in a clockwise direction starting from the top left, the power–power auto-correlation, the water-power cross-correlation (power as the reference network), the water–water auto-correlation, and the power-water cross-correlation (water as the reference network)

with a diagonal trend of weak correlations northeast and southwest of the evaluation nodes, mainly contributed by the water system characteristics and the availability of electricity. At the same time, the trend is still stronger when the power system is taken as the reference system, highlighting a greater operational and logistical influence of the power system on the water system than vice versa. Overall, the global correlation plots provide an illustrative synthesis of the spatial interdependencies over time and reveal an inherent directionality in the infrastructure restoration trends which capitalize on alternative paths and follow directions perpendicular to the main axis of networks.

The overall correlation plots or average global interdependence plots (Fig. 15.5) further consolidate the data by averaging across all angles in Fig. 15.4 to obtain the average correlation for a given distance relative to the two displaced networks. The plots are compared at different points in time and include error bars indicating one standard deviation from the mean. For each time lag, the plots are subdivided into four interdependence plots depicting in a clockwise direction starting from the top left, the power–power auto-correlation, the water-power cross-correlation (power as the reference network), the water–water auto-correlation, and the power-water cross-correlation (water as the reference network). The off diagonal plots reveal the change in coupling strength between the power and water networks in the region. While initial cross correlations begin at under 0.5, it is evident that correlations exist up to approximately 2,000 m before reducing to almost zero at low time lags. Also, comparisons of the relationship between average correlation and distance from a time lag of 1 day to 7 days show a faster decay in correlation behavior with time. Plotting average global correlations versus distance over time can lead to the creation of surfaces that reflect the overall evolution of average coupling strengths between networks. While the average global interdependence plots allow for a succinct representation

of overall spatial trends of interdependencies, average plots may obscure trends that are evident in analyses of global or local interdependency, hence the need to jointly explore local, global and average global information for practical network of networks modeling and operation recommendations. Clearly, the geographically focused reach of interdependence effects offers insights into siting and timing of back-up resources to handle the propagation of interdependent cascades, as well as the sizing or capacity requirements of equipment and crews.

The spatial-temporal methodology presented in this chapter is shown to be applicable within the context of infrastructure network interdependency, and furthermore, takes into consideration factors traditionally unaccounted for such as time and space. The demonstrated approach highlights the multifaceted and evolving nature of infrastructure network interdependencies and allows for the depiction of heterogeneity of interdependence in space and its evolution in time. Insights from the analyses include the ability to reveal interdependence directionality as well as to identify and measure the length in which interdependence remains influential. The application of graphical and mathematical tools in the quantification of interdependence contributes to the better understanding of network coupling, and thus enables more comprehensive and accurate network of networks models to inform future design and mitigation actions. Knowledge of coupling strengths may lead to possibilities in the exploitation of interdependencies for efficient operations or the decoupling or reduction of network dependence during post-disaster or remediation periods.

At the same time, spatial-temporal analyses remain highly dependent on data availability as well as the reliable fitting of the space and time variograms. Limitations to the model include the assumptions inherent in the product-sum estimation model and the accuracy of the sill-dependent coefficients of the covariance function. While the global and average global correlation plots succinctly summarize general network dependency behaviors, they may over or underemphasize certain correlation aspects evident in local analysis, and thus motivate the need to study local interdependence plots as well. In the end, the spatial-temporal methodology can be used in conjunction with field observations and anecdotal data to support local to global trend interpretations. By comprehensively quantifying and understanding infrastructure coupling characteristics as well as informing theoretical and simulation-based multi-network models about interdependence, it is possible to ultimately capture inherent geographical and temporal interdependencies between critical infrastructure networks and exploit such properties to enhance operation, control, and restoration strategies.

15.5 Conclusions

Spatial-temporal analyses of network interdependencies are uncommon to complex infrastructure systems, thus the approach demonstrated in this chapter takes an important step towards understanding and quantifying the holistic relationships between networks of networks. Overall, this chapter serves to outline methodologies

used in analyzing interdependent networks and quantifying their coupling. Furthermore, a novel spatial-temporal approach is recommended to apply on infrastructure systems so as to reveal their coupling structure and inform models as well as practical design or mitigation recommendations. Utilizing previous efforts and developments in network theory and space-time methods, the approach presented in this chapter applies the Product-Sum method to quantify variability in interdependence in both time and space realms. In addition, analyzing through slices in time and space allows for the depiction of interdependence trends that exist in multiple dimensions so as to calibrate emerging models of networks of networks and to define future strategies for the control of interdependencies, including siting and staging of resources as well as strategies for interface decoupling.

A methodological application was provided with representations of power and water system restoration curves derived from the 2010 moment magnitude (M_W) 8.8 Chilean earthquake in the Talcahuano-Concepción region, resulting in the creation of a spatial-temporal variogram surface that served as the foundation for incorporating both time and space lag dependencies. Graphical depictions of interdependencies represented by Pearson's correlation coefficients revealed the heterogeneous nature of local infrastructure interdependencies and uncovered an inherent southwest to northeast directionality in the global plot of interdependencies when shifting network restoration maps. Finally, analyses from average global interdependency plots (overall correlations) demonstrated the changing radial extent of restoration correlation influence with time as a measure of coupling between networked systems which in some cases changed from 2.0 km to less than 0.5 km. These findings, enriched with global correlation plots and local coupling insights, imply that restoration is interdependent in the perpendicular direction of the studied systems, which highlights the need to ensure operation of the systems in their main direction through back-up systems of pertinent capacities or uncouplings of specific geographical extents related to the radial distances to which interdependencies matter.

Future research will include the creation of surface or volumetric representations of local correlation values in time and space to better track their evolution, and further build upon the understanding of interdependencies given spatial-temporal lags for different components and systems. At the same time, other spatial-temporal methodologies such as the graph wavelet-based approach will be studied for further research. This chapter provides a basic methodology and structure for quantifying spatial-temporal coupling across infrastructure networks and enables validation, calibration, and expansion of emerging interdependent infrastructure models. By demonstrating the applicability of spatial-temporal modeling among infrastructure networks, research and design in future complex network interdependence can yield more accurate and realistic results for protecting utility networks and their users, particularly after episodes of abnormal operation.

References

1. Amin M (2001) Toward Self-Healing Infrastructure Systems. Computer Applications in Power, IEEE 14.1:20–28, doi:10.1109/67.893351

2. Barrat A, Boccaletti S, Caldarelli G, Chessa A, Latora V (2008) Complex Networks: from Biology to information technology. Journal of Physics A 41 220301

3. Barthélemy M. (2010) Spatial networks. ARXIV 2010arXiv1010.0302B.

4. Bashan A, Bartsch R, Kantelhardt J W, Havlin S, and Ivanov P (2012) Network physiology reveals relations between network topology and physiological function. Nature Communications 3.702, doi:10.1038/ncomms1705

5. Brummitt C, D'Souza R, and Leicht E (2012) Suppressing Cascades of Load in Interdependent Networks. PNAS 109.12:E680–E689, doi:10.1073/pnas.1110586109

6. Buldyrev S V, Parshani R, Paul G, Stanley H E, and Havlin S (2010) Catastrophic Cascade of Failures in Interdependent Networks. Nature 464:1025–1028

7. Chang S, McDaniels T, and Beaubien C (2009) Societal impacts of infrastructure failure interdependencies: building an empirical knowledge based. Proc. of the 2009 TCLEE Conference, Oakland, CA, 693-702

8. Clinton, W J (1996) Executive order 13010, establishing the president's commission on critical infrastructure protection (PCCIP), U.S.Government Printing Office, Washington, D.C.

9. Cressie N, Huang H (1999) Classes of Nonseparable, Spatio-temporal Stationary Covariance Functions. JASA 94:1330–1340

10. Crovella M, Kolaczyk E (2002) Graph Wavelets for Spatial Traffic Analysis. In IEEE INFOCOM

11. De Cesare L, Myers D, Posa D (2001) Estimating and modeling space-time correlation structures. Statistics and Probability Letters, 51.1:9–14

12. De Iaco S, Myers D, Posa D (2001) Spacetime analysis using a general productsum model, Statistics and Probability Letters. 52.1:21–28, doi:10.1016/S0167-7152(00)00200-5

13. Dueñas-Osorio L, Craig J, Goodno B, and Bostrom A (2007) Interdependent response of networked systems. Journal of Infrastructure Systems, 13.3: 185–194.

14. Dueñas-Osorio L, Craig J, and Goodno B (2007) Seismic response of critical independent networks. Earthquake Engineering and Structural Dynamics, 36.2: 285–306

15. Dueñas-Osorio L and Vemuru S M (2009) Cascading failures in complex infrastructure systems. Structural Safety, 31.2: 157–167

16. Dueñas-Osorio L, Kwasinski A (2012), Quantification of Lifeline System Interdependencies after the 27 February 2010 M_W 8.8 Offshore Maule, Chile, Earthquake. Earthquake Spectra 28.S1:S581–S603, doi:10.1193/1.4000054

17. Gao J, Buldyrev S V, Havlin S, and Stanley H E (2011) Robustness of a Network of Networks. Phys. Rev. Lett. 107, 195701

18. Gao J, Buldyrev S V, Havlin S, and Stanley H E (2012) Networks Formed from Interdependent Networks. Nature Physics 8:40–48

19. Gneiting T, Genton M, Guttorp P (2007) Geostatistical space-time models, stationarity, separability and full symmetry. In Statistical Methods for Spatio-Temporal Systems, Chapman and Hall/CRC, Boca Raton, 151–175.

20. Pidd H (2012) India blackouts leave 700 million without power. The Guardian. http://www.guardian.co.uk/world/2012/jul/31/india-blackout-electricity-power-cuts

21. Haimes Y and Jiang P (2001) Leontief-based Model of Risk in Complex Interconnected Infrastructures. Journal of Infrastructure Systems 7.1: 1–12

22. Havlin S (2009) Phone Infections. Science 324:1023–1024

23. Hernández-Fajardo I and Dueñas-Osorio L (2013). Probabilistic study of cascading failures in complex interdependent lifeline systems. Reliability Engineering and System Safety, 111: 260–272.

24. Kröger W (2008) Critical infrastructures at risk: A need for a new conceptual approach and extended analytical tools. Reliability Engineering and System Safety, 93.12:1781–1787, ISSN 0951–8320, doi:10.1016/j.ress.2008.03.005

25. Lee E, Mitchell J, and Wallace W (2007), Restoration of Services in Interdependent Infrastructure Systems: A Network Flows Approach, IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews 37:13031317

26. Mendonça D, Lee E E, and Wallace W A (2006) Impact of the 2001 World Trade Center Attack on Critical Interdependent Infrastructures. J. Infrastruct. Syst., 12.4: 260–270, doi:10.1061/(ASCE)1076-0342

27. Lee R and Kiremidjian A S (2007) Uncertainty and Correlation for Loss Assessment of Spatially Distributed Systems. Earthquake Spectra 23.4: 753–770.

28. Ouyang M and Dueñas-Osorio L (2011) Efficient Approach to Compute Generalized Interdependent Effects between Infrastructure Systems. J. Comput. Civ. Eng., 25.5:394–406, doi:10.1061/(ASCE)CP.1943-5487.0000103

29. Quill E (2012) When Networks Network: Once studied solo, systems display surprising behavior when they interact. Science News 182.6:18

30. President's Commission on Critical Infrastructure Protection (PCCIP). (1997). Critical foundations: Protecting Americas infrastructures. Rep., U.S. Government Printing Office, Washington, D.C.

31. Rahnamay-Naeini M, Pezoa J, et al. (2011) Modeling Stochastic Correlated Failures and their Effects on Network Reliability. Computer Communications and Networks (ICCCN), Proceedings of 20th International Conference on 1–6, doi:10.1109/ICCCN.2011.6005789

32. Rinaldi S, et al. (2001) Indentifying, understanding, and analyzing critical infrastructure interdependencies. IEEE Control Systems Magazine 21:11–25

33. Rosato V, Issacharoff L, Tiriticco F, and Meloni S (2008), Modeling interdependent infrastructures using interacting dynamical models. Int. J. Crit. Infrastruct., 4(1–2), 6379.

34. Satumtira G and Dueñas-Osorio L (2010). Chapter 1: Synthesis of modeling and simulation methods on critical infrastructure interdependencies research. Sustainable Infrastructure Systems: Simulation, Imaging, and Intelligent Engineering. Eds. K. Gopalakrishnan and S. Peeta. New York: Springer-Verlag.

35. Svendsen N K and Wolthusen S D (2007), Connectivity models of interdependency in mixed-type critical infrastructure networks. Inform. Sec. Tech. Rep., 12(1), 4455.

36. Sydney A, Scoglio C, Youssef M, Schumm P (2010), Characterizing the Robustness of Complex Networks. Int. J. Internet Technology and Secured, Transactions. 2.3/4:291–320

37. Technical Council on Lifeline Earthquake Engineering (TCLEE), (2010) Preliminary report on lifeline system performance after the M_W 8.8 offshore Maule, Chile, earthquake of February 27, 2010, American Society of Civil Engineers (ASCE)

38. Trauth M H (2010) MATLAB Recipes for Earth Sciences. Springer.

39. Sarma H and Russell R (2012) Second day of India's electricity outage hits 620 million. USA Today. http://usatoday30.usatoday.com/news/world/story/2012-07-31/india-power-outage/56600520/1

40. Wu J, Dueñas-Osorio L, and Villagran M (2012) Spatial Quantification of Lifeline System Interdependencies. Proceedings of the 15th world conference in earthquake engineering (15WCEE), Lisbon, Portugal, September 24–28, 2012

41. Wu J and Dueñas-Osorio L (2013) Calibration and validation of a seismic damage propagation model for interdependent infrastructure systems. Earthquake Spectra, 29.3:1021–1041

42. Zio E and Sansavini G (2011) Modeling Interdependent Network Systems for Identifying Cascade-Safe Operating Margins. Reliability, IEEE Transactions on 60.1:94–101, doi:10.1109/TR.2010.2104211