6

ALBERT-LÁSZLÓ BARABÁSI

NETWORK SCIENCE
EVOLVING NETWORKS

INDEX

Figure 6.0 (cover image)
Cubism and Abstract Art by Alfred Barr

Alfred H. Barr, Jr. (1902-1981), an American art historian and the first director of the Museum of Modern Art (MoMA) in New York City, was one of the most influential forces in the development of popular attitudes towards modern art. This 1936 chart by Barr uses a network framework to illustrate the development and crosscurrents of modern art. It appeared on the dust jacket of the catalogue for *Cubism and Abstract Art*, the movement's first major exhibit at MoMA.

INTRODUCTION

Founded six years after birth of the World Wide Web, Google was a latecomer to search. By the late 1990s Alta Vista and Inktomi, two search engines with an early start, have dominated the search market. Yet, the third mover Google soon not only became the leading search engine, but acquired links at such an incredible rate that by 2000 became the biggest hub of the Web as well [1]. But it didn't last: In 2011 Facebook, a youngster with Google's standards, took over as the Web's biggest node.

The Web's competitive landscape highlights an important limitation of our modeling framework: None of the network models we encountered so far are able to account for it. Indeed, in the Erdős-Rényi model the identity of the biggest node is driven entirely by chance. The Barabási-Albert model offers a more realistic picture, predicting that each node increases its degree following $k(t) \sim t^{1/2}$. This means that the oldest node always has the most links, a phenomena called the *first mover's advantage* in the business literature. It also means that a late node can never become the largest hubs.

In reality the growth rate of a node does not depend on its age alone. Instead webpages, companies, or actors have intrinsic qualities that influence the rate at which they acquire links. Some show up late and nevertheless grab an extraordinary number of links within a short timeframe. Others rise early yet never quite make it. The goal of this chapter is to understand how the differences in the node's ability to acquire links affect the network topology. Going beyond this competitive landscape, we also explore how other processes, like node and link deletion (Figure 6.1) or the aging of nodes, phenomena frequently observed in real networks, change the way networks evolve and alter their topology. Our goal is to develop a self-consistent theory of evolving networks that can be adjusted at will to predict the dynamics and the topology of a wide range of real networks.

Figure 6.1
Garment District

The Garment District is a Manhattan neighborhood located between Fifth and Ninth Avenue, from 34th to 42nd Street. Since the early 20th century it has been the center for fashion manufacturing and design in the United States. The *Needle threading a button* and the *Jewish Tailor*, two sculptures located in the heart of the district, pay tribute to the neighborhood's past.

The garment industry of New York City offers a prominent example of a declining network, helping us understand how the loss of nodes shapes a network's topology (BOX 6.5). Uncovering the impact of processes like node and link loss on the network topology is one of the goals of this chapter.

THE BIANCONI-BARABÁSI MODEL

Some people have a knack for turning each random encounter into a lasting social link; some companies turn each consumer into a loyal partner; some webpages turn visitors into addicts. A common feature of these successful nodes is some intrinsic property that propels them ahead of the pack. We will call this property *fitness*.

Fitness is an individual's gift to turn a random encounter into a lasting friendship; it is a company's knack to acquire consumers relative to its competition; it is a webpage's ability to bring us back on a daily basis despite the many other pages that compete for our attention. Fitness may have genetic roots in people, it may be related to innovativeness and management quality in companies and may depend on the content offered by a website.

In the Barabási-Albert model we assumed that a node's growth rate is determined solely by its degree. To incorporate the role of fitness we assume that preferential attachment is driven by the product of a node's fitness, η, and its degree k. The resulting model, called the *Bianconi-Barabási* or the *fitness model*, consists of the following two steps [2, 3]:

- **Growth**
  In each timestep a new node j with m links and fitness η_j is added to the network, where η_j is a random number chosen from a *fitness distribution $\rho(\eta)$*. Once assigned, a node's fitness does not change.

- **Preferential Attachment**
  The probability that a link of a new node connects to node i is proportional to the product of node i's degree k_i and its fitness η_i,

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \cdot \tag{6.1}$$

In (6.1) the dependence of Π_i on k_i captures the fact that higher-degree nodes have more visibility, hence we are more likely to link to them. The de-
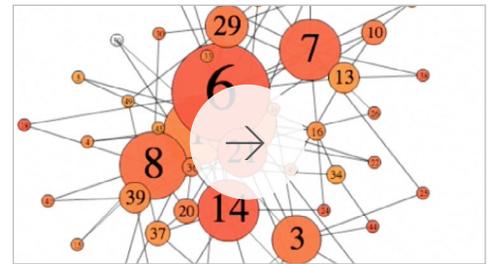
The Bianconi-Barabási Model

The movie shows a growing network in which each new node acquires a randomly chosen fitness parameter at birth, indicated by the color of the node. Each new node chooses the nodes it links to following generalized preferential attachment (6.1), making a node's growth rate proportional to its fitness. The node size is proportional to its degree, illustrating that with time the nodes with the highest fitness turn into the largest hubs. *Video courtesy of Dashun Wang.*

LINEAR PLOT   LOG-LOG PLOT

(a) BARABÁSI–ALBERT MODEL
(b)
(c) BIANCONI-BARABÁSI MODEL
(d)

$\eta = 0.223$
$\eta = 0.185$
$\eta = 0.991$

pendence of Π_i on η_i implies that between two nodes with the same degree, the one with higher fitness is selected with a higher probability. Hence, (6.1) assures that even a relatively young node, with initially only a few links, can acquire links rapidly if it has larger fitness than the rest of the nodes.

DEGREE DYNAMICS

We can use the continuum theory to predict each node's temporal evolution. According to (6.1), the degree of node i changes at the rate

$$\frac{\partial k_i}{\partial t} = m\frac{\eta_i k_i}{\sum_j \eta_j k_j} . \tag{6.2}$$

Let us assume that the time evolution of k_i follows a power law with a fitness-dependent exponent $\beta(\eta_i)$ (Figure 6.2),

$$k(t, t_i, \eta_i) = m\left(\frac{t}{t_i}\right)^{\beta(\eta_i)} . \tag{6.3}$$

Inserting (6.3) into (6.2) we find that the *dynamic exponent* satisfies (ADVANCED TOPICS 6.A)

$$\beta(\eta) = \frac{\eta}{C} \tag{6.4}$$

with

$$C = \int \rho(\eta)\frac{\eta}{1 - \beta(\eta)}d\eta . \tag{6.5}$$

In the Barabási-Albert model we have $\beta = 1/2$, hence the degree of each node increases as a square root of time. According to (6.4), in the Bianconi-Barabási model the dynamic exponent is proportional to the node's fitness, η, hence each node has its own dynamic exponent. Consequently, a node with a higher fitness will increase its degree faster. Given sufficient time, the fitter node will leave behind nodes with a smaller fitness (Figure 6.2). Facebook is a poster child of this phenomenon: a latecomer with an addictive product, it acquired links faster than its competitors,

eventually becoming the Web's biggest hub.

DEGREE DISTRIBUTION

The degree distribution of the network generated by the Bianconi-Barabási model can be calculated using the continuum theory (ADVANCED TOPICS 6.A), obtaining

$$p_k \approx C \int d\eta \frac{\rho(\eta)}{\eta} \left(\frac{m}{k}\right)^{\frac{C}{\eta}+1}. \tag{6.6}$$

Equation (6.6) is a weighted sum of multiple power-laws, indicating that p_k depends on the precise form of the fitness distribution, $\rho(\eta)$. To illustrate the properties of the model we use (6.4) and (6.6) to calculate $\beta(\eta)$ and p_k for two fitness distributions:

- **Equal Fitnesses**
  When all fitnesses are equal, the Bianconi-Barabási model reduces to the Barabási-Albert model. Indeed, let us use $\rho(\eta) = \delta(\eta - 1)$, capturing the fact that each node has the same fitness $\eta = 1$. In this case (6.5) yields $C = 2$. Using (6.4) we obtain $\beta = 1/2$ and (6.6) predicts $p_k \sim k^{-3}$, the known scaling of the degree distribution in the Barabási-Albert model.

- **Uniform Fitness Distribution**
  The model's behavior is more interesting when nodes have different fitnesses. Let us choose η to be uniformly distributed in the [0,1] interval. In this case C is the solution of the transcendental equation (6.5)

$$\exp(-2/C) = 1 - 1/C, \tag{6.7}$$

  whose numerical solution is $C^* = 1.255$. Consequently, (6.4) predicts that each node i has a different dynamic exponent, $\beta(\eta_i) = \eta_i / C^*$.

  Using (6.6) we obtain

$$p_k \sim \int_0^1 d\eta \frac{C^*}{\eta} \frac{1}{k^{1+C^*/\eta}} \sim \frac{k^{-(1+C^*)}}{\ln k}, \tag{6.8}$$

  predicting that the degree distribution follows a power law with degree exponent $\gamma = 2.255$. Yet, we do not expect a perfect power law, but the scaling is affected by an inverse logarithmic correction $1/\ln k$.

  Numerical support for the above predictions is provided in Figures 6.2 and 6.3. The simulations confirm that $k_i(t)$ follows a power law for each η and that the dynamical exponent $\beta(\eta)$ increases with the fitness η. As Figure 6.3a indicates, the measured dynamical exponents are in excellent agreement with the prediction (6.4). Figure 6.3b also documents an agreement between (6.8) and the numerically obtained degree distribution.

(a)

(b)
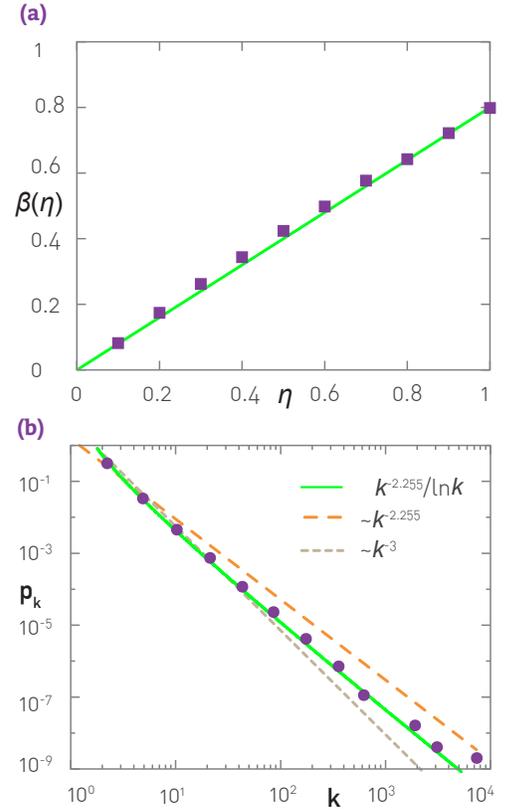
Figure 6.3
Characterizing the Bianconi-Barabási Model

(a) The measured dynamic exponent $\beta(\eta)$ shown in function of η for a uniform $\rho(\eta)$ distribution. The squares were obtained from numerical simulations while the solid line corresponds to the analytical prediction $\beta(\eta) = \eta/1.255$.

(b) Degree distribution of the model obtained numerically for a network with $m=2$ and $N = 10^6$ and fitnesses chosen uniformly from the $\eta \in [0, 1]$ interval. The green solid line corresponds to the prediction (6.8) with $\gamma = 2.255$. The long-dashed line is $p_k \sim k^{-2.255}$ without the logarithmic correction, while the short-dashed curve correspond to $p_k \sim k^{-3}$, expected if all fitness are equal. Note that the best fit is provided by (6.8).

In summary, the Bianconi-Barabási model can account for the fact that nodes with different internal characteristics acquire links at different rates. It predicts that a node's growth rate is determined by its fitness η and allows us to calculate the dependence of the degree distribution on the fitness distribution $\rho(\eta)$.

MEASURING FITNESS

Measuring a node's fitness could help us identify web sites that are poised to grow in visibility, research papers that will become influential, or actors on their way to stardom (BOX 6.1). Yet, our ability to determine the fitness is prone to errors. Consider the challenge of assigning fitness to a webpage on sumo wrestling: While a small segment of the population might find sumo wrestling fascinating, most individuals are indifferent to it and some might even find it odd. Hence, different individuals will inevitably assign different fitnesses to the same node.

According to (6.1) fitness is not assigned by any individual, but reflects the network's *collective perception of a node's importance relative to the other nodes*. We can, therefore, determine a node's fitness by comparing its time evolution to the time evolution of other nodes in the network. In this section we show that if we have dynamical information about the evolution of the individual nodes, the quantitative framework of the Bianconi-Barabási model allows us to determine the fitness of each node.

To relate a node's growth rate to its fitness we take the logarithm of (6.3),

$$\ln k(t, t_i, \eta_i) = \beta(\eta_i) \ln t + B_i \qquad (6.9)$$

where $B_i = \ln (m/t_i^{\beta(\eta_i)})$ is a time-independent parameter. Hence, the slope of $\ln k(t, t_i, \eta_i)$ is a linear function of the dynamical exponent $\beta(\eta_i)$. In turn $\beta(\eta_i)$ depends linearly on η_i according to (6.4). Therefore, if we can track the time evolution of the degree for a large number of nodes, the distribution of the dynamical exponent $\beta(\eta_i)$ will be identical with the fitness distribution $\rho(\eta)$.

The Fitness of a Web Document
Node fitnesses were systematically measured in the context of the WWW, relying on a dataset that crawled monthly the links of about 22 million web documents for 13 months [9]. While most nodes (documents) did not change their degree during this time frame, 6.5% of nodes showed sufficient changes to determine their dynamical exponent via (6.9). The obtained fitness distribution $\rho(\eta)$ has an exponential form (Figure 6.4), indicat-

Could fitness, an ability to acquire friends in a social network, have genetic origins? To answer this researchers examined the social network of 1,110 school-age twins [6, 7], using a technique developed to identify the heritability of traits and behaviors. They found that:

• Genetic factors account for 46% of the variation in a student's in-degree (i.e. the number of students that name a given student a friend).

• Generic factors are not significant for out-degrees (i.e. the number of students a particular student names as friends).

This suggests that an individual's ability to acquire links, or its fitness, is heritable. In other words, in social networks fitness has genetic origins. This conclusion is also supported by research that associated a particular genetic trait with variations in popularity [8].

ing that high fitness nodes are rare.

The shape of the obtained fitness distribution is somewhat unexpected, as one would be tempted to assume that on the web fitness varies widely: For example Google is far more attrative to Web surfers than my personal webpage. Yet the exponential form of $\rho(\eta)$ indicates that the fitness of Web documents is bounded, varying in a relativelly narrow range. Consequently, the observed large differences in the degree of two web documents is generated by the system's dynamics: Growth and preferential attachment amplifies the small fitness differences, turning nodes with slightly higher fitness into much bigger nodes.

To illustrate this amplification, consider two nodes that arrived at the same time, but have different fitnesses $\eta_2 > \eta_1$. According to (6.3) and (6.4), the relative difference between their degrees grows for large t as

$$\frac{k_2 - k_1}{k_1} \sim t^{\frac{\eta_2 - \eta_1}{C}}.$$

(6.10)

While the difference between η_2 and η_1 may be small, far into the future (large t) the relative difference between their degrees can become quite significant.

The Fitness of a Scientific Publication

In some networks the nodes follow a more complex dynamics than the one predicted by (6.3). To measure their fitness we must first account for their precise growth law. We illustrate this procedure by determining the fitness of a research publication, allowing us to predict its future impact.

While most research papers acquire only a few citations, a small number of publications collect thousands and even tens of thousands of citations [10]. These impact differences mirror differences in the novelty and the relevance of various publications. In general, the probability that a research paper i is cited at time t after publication is [11]

$$\Pi_i \sim \eta_i c_i^t P_i(t),$$

(6.11)

where the paper's fitness η_i accounts for the perceived novelty and importance of the reported discovery; c_i^t is the cumulative number of citations acquired by paper i at time t after publication, accounting for the fact that well-cited papers are more likely to be cited than less-cited contributions (preferential attachment). The last term in (6.11) captures the fact that new ideas are integrated into subsequent work, hence the novelty of each paper fades with time [11, 12]. Measurements indicate that this decay has the log-normal form

$$P_i(t) = \frac{1}{\sqrt{2\pi t}\,\sigma_i} e^{-\frac{(\ln t - \mu_i)^2}{2\sigma_i^2}}.$$

(6.12)

By solving the master equation behind (6.11) we obtain the time-dependent growth of a paper's citations

$$c_i^t = m\left(e^{\frac{\beta \eta_i}{A}\Phi\left(\frac{\ln t - \mu_i}{\sigma_i}\right)} - 1 \right),$$

(6.13)
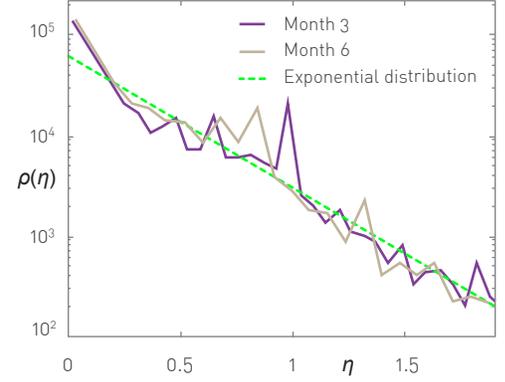
Figure 6.4
The Fitness Distribution of the WWW

The fitness distribution obtained by measuring the time evolution of a large number of Web documents. The measurements indicate that each node's degree has a power law time dependence, as predicted by (6.3). The slope of each curve is $\beta(\eta_i)$, which corresponds to the node's fitness η_i up to a multiplicative constant according to (6.4). The plot shows the result of two measurements based on datasets recorded three months apart, demonstrating that the fitness distribution is time independent. The dashed line suggests that the fitness distribution is well approximated by an exponential. After [9].

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy \qquad (6.14)$$

is the cumulative normal distribution and m, β, and A are global parameters.

Equation (6.13) predicts that the citation history of paper i is characterized by three parameters: the immediacy μ_i, governing the time for a paper to reach its citation peak and the longevity σ_i, capturing the decay rate. The most important is the relative fitness $\eta_i' \equiv \eta_i \beta/A$, which measures a paper's importance relative to other papers and determines its ultimate impact (BOX 6.2).

We fit (6.13) to the citation history of individual papers published by a journal to obtain the journal's fitness distribution (Figure 6.5). The measurements indicate that the fitness distribution of the top cell biology journal, *Cell*, is shifted to the right, indicating that *Cell* papers tend to have high fitness. Not surprisingly, the journal has one of the highest impact factors of all journals. By comparison the fitness of papers published in *Physical Review* are shifted to the left, indicating that the journal publishes fewer high fitness papers.

In summary, the framework offered by the Bianconi-Barabási model allows us to determine the fitness of individual nodes and the shape of the fitness distribution $\rho(\eta)$. The measurements show that the fitness distribution is typically exponentially bounded, meaning that fitness differences between different nodes are small. With time these differences are magnified, resulting in a power law degree distribution of incoming links in the case of the WWW or a broad citation distribution in citation networks.
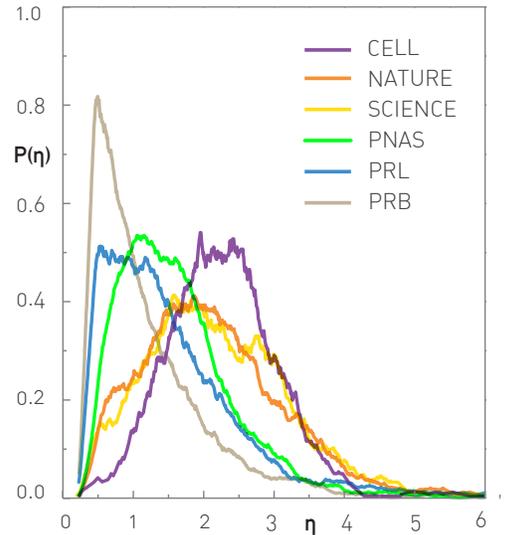
Figure 6.5
Fitness Distribution of Research Papers

The fitness distribution of papers published in six journals in 1990. Each paper's fitness was obtained by fitting (6.13) to the paper's citation history for a decade long time interval. Two journals are from physics (*Physical Review B* and *Physical Review Letters*), one from biology (*Cell*) and three are interdisciplinary (*Nature*, *Science*, and *PNAS*).

The obtained fitness distributions are shifted relative to each other, indicating that *Cell* publishes the highest fitness papers, followed by *Nature*, *Science*, *PNAS*, *Physical Reviews Letters* and *Physical Review B*. After [11].

BOX 6.2

ULTIMATE IMPACT
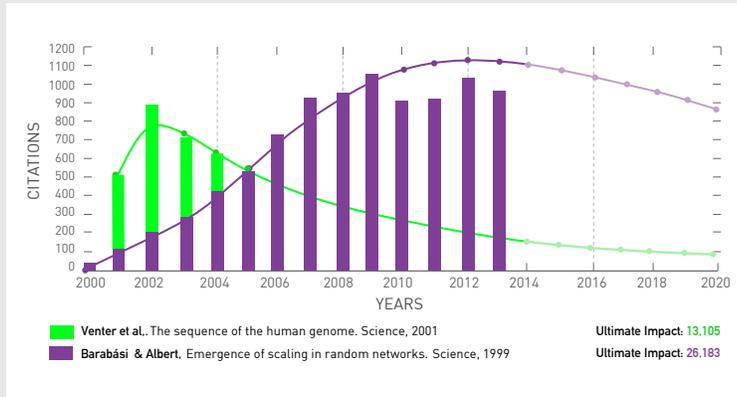
Figure 6.6
Predicting Ultimate Impact

The yearly citation history of the paper reporting the first draft of the human genome (Venter et al. [13]) and the one reporting the discovery of scale-free networks (Barabási and Albert [14]). The early impact of the two papers cannot be more different: According to the Web of Science, two years after publication the much anticipated human genome paper collected over 1,400 citations; in contrast the scale-free network paper was cited only 120 times. Their long-term citation dynamics is also remarkably different: The citations of the human genome paper peaked after year two, a pattern shared by more than 85% of all research papers. In contrast the yearly citations to the scale-free paper continued to increase for about a decade.

The continuous curves corresponding to the fit (6.13) to the respective citation history, allowing us to determine the paper's future citations and its *ultimate impact*. The ultimate impact corresponds to the total area under each curve for t → ∞. According to (6.15) the ultimate impact of the human genome paper is 13,105, while that of the scale-free paper is 26,183. Therefore the early citation count of a paper is not a strong indicator of its ultimate impact.

Citation counts offer only the historical impact of a research paper. They do not tell us, however, if the paper has already had its run, or its impact will continue to grow. To gauge a paper's true impact we need to determine how many citations a paper acquires *during its lifetime*. The citation model (6.11) and (6.14) allows us to predict this *ultimate impact*. Indeed, by taking the t → ∞ limit of (6.13), we obtain [11]

$$c_i^{\infty} = m(e^{\eta_i} - 1) \ . \qquad (6.15)$$

Consequently, despite the myriad of factors that contribute to the citation history of a research paper, its ultimate impact is determined only by its fitness η_i. As fitness can be determined by fitting (6.13) to a paper's previous citation history, we can use (6.15) to predict the ultimate impact of a publication (Figure 6.6).

BOSE-EINSTEIN CONDENSATION

In the previous section we found that the Web's fitness distribution follows a simple exponential (Figure 6.4), while the fitness of research papers follows a peaked distribution (Figure 6.5). The diversity of the observed fitness distributions raises an important question: How does the network topology depend on the shape of $\rho(\eta)$?

Technically, the answer is provided by (6.6) that links p_k to $\rho(\eta)$. Yet, the true impact of the fitness distribution was realized only after the discovery that some networks can undergo Bose-Einstein condensation. In the section we discuss the mapping that lead to this discovery and its consequences for the network topology [15].

We start by establishing a formal link between the Bianconi-Barabási model and a Bose gas, whose properties have been extensively studied in physics (Figure 6.7):

- **Fitness → Energy**

  We assign to each node with fitness η_i an energy ε_i via

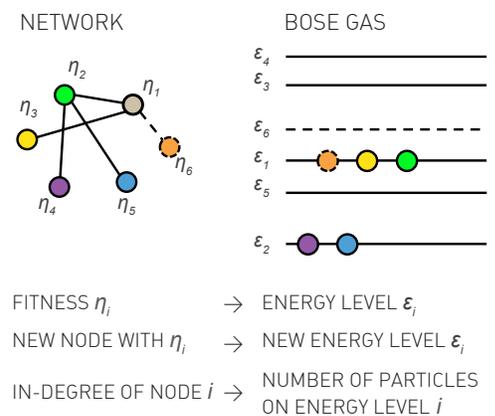 $$\varepsilon_i = \frac{1}{\beta_T} \log \eta_i. \qquad (6.16)$$

  In physical systems β_T plays the role of the inverse temperature. We use the subscript T to distinguish β_T from the dynamic exponent β. According to (6.16), each node in a network corresponds to an energy level in a Bose gas. The larger the node's fitness, the lower is its energy.

- **Links → Particles**

  For each link from node i to node j we add a particle at the energy level ε_j.

- **Nodes → Energy levels**

  The arrival of a new node with m links corresponds to adding a new energy level ε_j and m new particles to the Bose gas, placed on the energy levels of the nodes to which the new node links to.

NETWORK   BOSE GAS

FITNESS η_i → ENERGY LEVEL ε_i

NEW NODE WITH η_i → NEW ENERGY LEVEL ε_i

IN-DEGREE OF NODE i → NUMBER OF PARTICLES ON ENERGY LEVEL i

Figure 6.7
Mapping Networks to a Bose Gas

Network
A network of six nodes, where each node is characterized by a unique fitness η_i indicated by the node color. The fitnesses are chosen from the fitness distribution $\rho(\eta)$.

Bose Gas
The mapping assigns an energy level ε to each fitness η, resulting in a Bose gas with random energy levels. A link going from a new node i to node j corresponds to one particle at level ε_j.

Growth
The network grows by adding a new node, like the orange node with fitness η_6. For $m=1$ the new node connects to the grey node (dashed link), chosen following (6.1). In the Bose gas this corresponds to the addition of a new energy level ε_6 (dashed line), and the deposition of a particle at ε_1, the energy level of node 1 to which node η_6 links to.

If we follow the mathematical consequences of this mapping, we find that in the resulting gas the number of particles on each energy level follows Bose statistics, a formula derived by Satyendra Nath Bose in 1924 (BOX 6.3). Consequently, the links of the fitness model behave like subatomic particles in a quantum gas.

The mapping to a Bose gas is exact and predicts the existence of two distinct phases [15, 16]:

Scale-free Phase

For most fitness distributions the network displays a fit-gets-rich dynamics, meaning that the degree of each node is ultimately determined by its fitness. While the fittest node inevitably becomes the largest hub, in this phase at any moment the degree distribution follows a power law, indicating that the generated network has a scale-free topology. Consequently the largest hub follows (4.18), growing only sublinearly. This hub is closely trailed by a few slightly smaller hubs, with almost as many links as the fittest node (Figure 6.9a). The model with uniform fitness distribution discussed in SECTION 6.2 is in this scale-free phase.

Bose-Einstein Condensation

The unexpected outcome of the mapping to a Bose gas is the possibility of a Bose-Einstein condensation for some fitness distributions. In a Bose-Einstein condensate all particles crowd to the lowest energy level, leaving the rest of the energy levels unpopulated (BOX 6.4).

In a network Bose-Einstein condensation means that the fittest node grabs a finite fraction of the links, turning into a super-hub (Figure 6.9b). The resulting network is not scale-free but has a hub-and-spoke topology. In this phase the rich-gets-richer process is so dominant that it becomes a qualitatively different *winner takes-all phenomenon*. Consequently, the network will loose its scale-free nature.

In physical systems Bose-Einstein condensation is induced by lowering the temperature of the Bose gas below some critical temperature (BOX 6.4). In networks, the temperature β_T in (6.16) is a dummy variable, disappearing from all topologically relevant quantities, like the degree distribution p_k. Hence, the presence or the absence of Bose-Einstein condensation depends only on the form of the fitness distribution $\rho(\eta)$. For a network to undergo Bose-Einstein condensation, the fitness distribution needs to satisfy the condition

$$\int_{\eta_{\min}}^{\eta_{\max}} \frac{\eta\rho(\eta)}{1-\eta}d\eta < 1$$

A fitness distribution that leads to a Bose-Einstein condensation is

$$\rho(\eta) = (1-\zeta)(1-\eta)^\xi, \qquad (6.22)$$

In the context of the Bose gas (Figure 6.7) the probability that a particle lands on level *i* is

$$\Pi_i = \frac{e^{-\beta_T \varepsilon_i} k_i}{\sum_j e^{-\beta_T \varepsilon_j} k_j} . \qquad (6.17)$$

Hence, the rate at which the energy level ε_i accumulates particles is [15]

$$\frac{\partial k_i(\varepsilon_i, t, t_i)}{\partial t} = m \frac{e^{-\beta_T \varepsilon_i} k_i(\varepsilon_i, t, t_i)}{Z_t} \qquad (6.18)$$

where $k_i(\varepsilon_i, t, t_i)$ is the occupation number of level *i* and

$$Z_t = \sum_{j=1}^{t} te^{-\beta_T \varepsilon_j} k_j(\varepsilon_i, t, t_j)$$

is the partition function. The solution of (6.18) is

$$k_i(\varepsilon_i, t, t_i) = m\left(\frac{t}{t_i}\right)^{f(\varepsilon_i)} \qquad (6.19)$$

where $f(\varepsilon) = e^{-\beta_T (\varepsilon - \mu)}$ and μ is the *chemical potential* satisfying

$$\int \deg(\varepsilon) \frac{1}{e^{\beta_T(\varepsilon-\mu)} - 1} = 1 . \qquad (6.20)$$

Here, deg(ε) is the degeneracy of the energy level ε. Equation (6.20) suggests that in the limit $t \rightarrow \infty$ the occupation number, representing the number of particles with energy ε, follows the *Bose statistics*

$$n(\varepsilon) = \frac{1}{e^{\beta_T(\varepsilon-\mu)} - 1} . \qquad (6.21)$$

This mapping of the fitness model to a Bose gas proves that the node degrees in the Bianconi-Barabási model follow Bose statistics.

BOX 6.4

In classical physics the kinetic energy of a moving particle, $E = mv^2/2$, can have any value between zero (at rest) and an arbitrarily large E, when it moves very fast. Furthermore, an arbitrary number of particles can have the same energy E, if they have the same velocity v. In quantum mechanics energy is quantized, meaning that it can only take up discrete (quantized) values. Furthermore, in quantum mechanics we encounter two different classes of particles. Fermi particles, like electrons, are forbidden to have the same energy within the same system. Hence, only one electron can occupy a given energy level (Figure 6.8a). In contrast Bose particles, like photons, are allowed to crowd in arbitrary numbers on the same energy level (Figure 6.8b).

At high temperatures, when thermal agitation forces the particles to take up different energies, the difference between a Fermi and a Bose gas is negligible (Figure 6.8a,b). The difference becomes significant at low temperatures when all particles are forced to take up their lowest allowed energy. In a Fermi gas at low temperatures the particles fill the energy levels from bottom up, just like pouring water fills up a vase (Figure 6.8c). However, as any number of Bose particles can share the same energy, they can all crowd at the lowest energy level (Figure 6.8d). In other words, no matter how much "Bose liquid" we pour into the vase, it will stay at the bottom of the vessel, never filling it up. This phenomenon is called a Bose-Einstein condensation and it was first proposed by Einstein in 1924. Experimental evidence for Bose-Einstein condensation emerged only in 1995 and was recognized with the 2001 Nobel prize in physics.

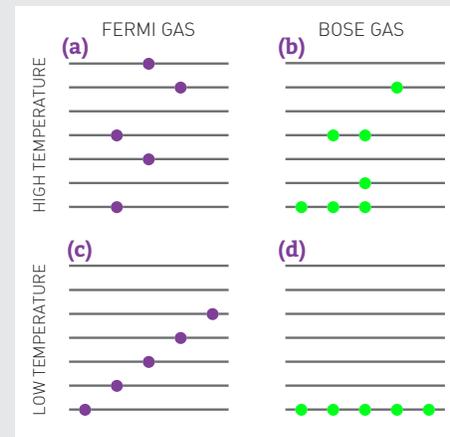**Figure 6.8**
Bose and Fermi Statistics

In a Fermi gas (a,c) only one particle is allowed on each energy level, while in a Bose gas (b,d) there is no such a restriction. At high temperatures it is hard to notice any difference between the two gases. At low temperatures, however, each particle wants to occupy the lowest possible energy and the difference between the two gases becomes significant.

whereby varying ζ we can induce a Bose-Einstein condensation (Figure 6.9). Indeed, whether (6.20) has a solution depends on the functional form of the energy distribution, $g(\varepsilon)$, which is determined by the shape of $\rho(\eta)$. Specifically, if (6.22) has no non-negative solution for a given $g(\varepsilon)$, a Bose-Einstein condensation emerges, and a finite fraction of the particles agglomerate at the lowest energy level.

In summary, the precise shape of the fitness distribution determines the topology of a growing network. While fitness distributions like the uniform distribution lead to a scale-free topology, some $\rho(\eta)$ allow for Bose-Einstein condensation. If a network undergoes a Bose-Einstein condensation, then one or a few nodes grab most of the links. Hence, the rich-gets-richer process responsible for the scale-free state turns into a winner-takes-all phenomenon. The Bose-Einstein condensation has such an obvious impact on a network's structure that, if present, it is hard to miss: it destroys the hierarchy of hubs characterizing a scale-free network, forcing the network into a star-like hub-and-spoke topology (Figure 6.9).
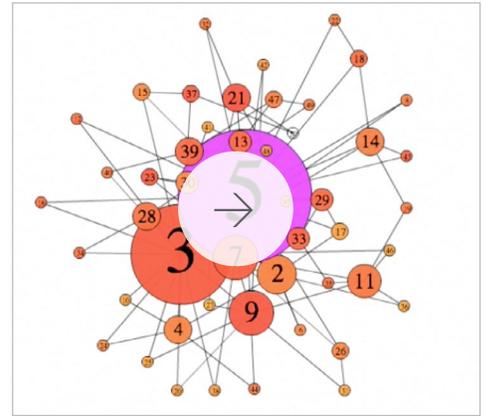
Online Resource 6.2
Bose-Einstein Condensation in Networks

The movie shows the time evolution of a growing network in which one node (purple) has a much higher fitness than the rest of the nodes. This high fitness node attracts most links, forcing the system to undergo a Bose-Einstein condensation. *Video courtesy of Dashun Wang.*
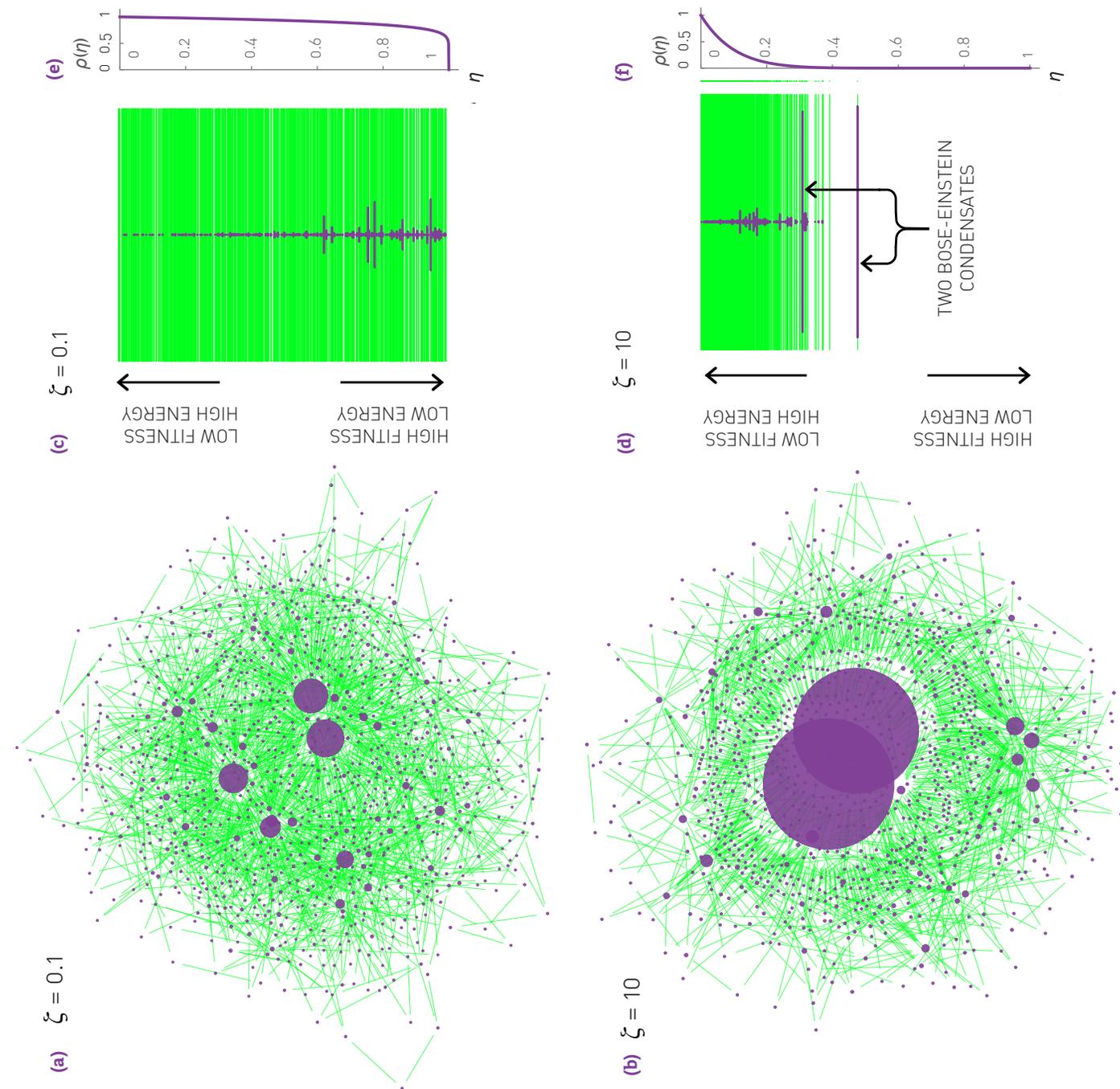
Figure 6.9
Bose-Einstein Condensation in Networks

(a,b) A scale-free network (a) and a network that has undergone a Bose-Einstein condensation (b). Both networks were generated by the Bianconi-Barabási model with $\rho(\eta)$ following **(6.22)**, but with different exponent ζ. Note that in the condensed phase (b) we have two large hubs with comparable size.

(c,d) The energy levels (green lines) and the deposited particles (purple dots) for a network with $m=2$ and $N=1,000$. Each energy level corresponds to the fitness of a node of the network shown in (a,b). Each link connected to a node is represented by a particle on the corresponding energy level. As we did not allow multi-links, two highly populated energy levels emerged in (d), indicating that we have two condensates, corresponding to the two hubs seen in (b).

(e,f) The fitness distribution $\rho(\eta)$, given by **(6.22)**, illustrates the difference in the shape of the two $\rho(\eta)$ functions. The difference is determined by the parameter ζ, which is (e) $\zeta = 0.1$ and (f) $\zeta = 10$.

(a) $\zeta = 0.1$

(b) $\zeta = 10$

(c) $\zeta = 0.1$

(d) $\zeta = 10$

HIGH FITNESS
LOW ENERGY

LOW FITNESS
HIGH ENERGY

TWO BOSE-EINSTEIN CONDENSATES

(e) $\rho(\eta)$

(f) $\rho(\eta)$

EVOLVING NETWORKS

The Barabási-Albert model is a minimal model, designed to capture the mechanisms responsible for the emergence of the scale-free property. Consequently, it has several well-known limitations (see also SECTION 5.10):

(i) The model predicts $\gamma = 3$, while the experimentally observed degree exponents vary between 2 and 5 (Table 4.1).

(ii) The model predicts a power-law degree distribution, while in real systems we observe systematic deviations from a pure power-law function, like small-degree saturation or high-degree cutoff (BOX 4.8).

(iii) The model ignores a number of elementary processes that are obviously present in many real networks, like the addition of internal links and node or link removal.

These limitations have inspired considerable research, clarifying the role of the numerous elementary processes that influence the network topology. In this section we systematically extend the Barabási-Albert model, arriving to a family of evolving network models that can capture the wide range of phenomena known to shape the topology of real networks.

INITIAL ATTRACTIVENESS

In the Barabási-Albert model an isolated node cannot acquire links, as according to preferential attachment (4.1) the likelihood that a new node attaches to a $k=0$ node is strictly zero. In real networks even isolated nodes acquire links. Indeed, each new research paper has a finite probability of being cited for the first time; a person that moves to a new city quickly acquires acquaintances. To allow unconnected nodes to acquire links we add a constant to the preferential attachment function (4.1),

$$\Pi(k) \sim A + k .$$

(6.23)

Here the constant A is called *initial attractiveness*. As $\Pi(0) \sim A$, initial

attractiveness is proportional to the probability that a node acquires its first link in the next time step.

Direct measurement of *Π(k)* shows that initial attractiveness is present in real networks (Figure 6.10). Once present, it has two consequences:

- **Increases the Degree Exponent**
  If in the Barabási-Albert model we replace (4.1) with (6.23), the degree exponent becomes [17, 18]

$$\gamma = 3 + \frac{A}{m} \cdot$$
(6.24)

Consequently initial attractiveness increases γ, making the network more homogeneous and reducing the size of the hubs. Indeed, initial attractiveness adds a random component to the probability of attaching to a node. This random component favors the numerous small-degree nodes and weakens the role of preferential attachment. For high-degree nodes the initial attractiveness term *A* in (6.23) is negligible.

- **Generates a Small-degree Saturation**
  The solution of the continuum equation indicates that the degree distribution of a network governed by (6.23) does not follow a pure power-law, but has the form

$$p_k = C(k + A)^{-\gamma}.$$
(6.25)

Therefore, initial attractiveness induces a small-degree saturation for $k<A$, playing the role of k_{sat} in (4.39). This saturation is rooted in the fact that initial attractiveness enhances the probability that new nodes link to the small-degree nodes, which pushes the small-*k* nodes towards higher degrees. For high degrees ($k \gg A$) the degree distribution continues to follow a power law, as in this range initial attractiveness does not alter the attachment probability.

INTERNAL LINKS

In many networks new links do not only arrive with new nodes but are added between pre-existing nodes. For example, the vast majority of new links on the WWW are *internal links*, corresponding to newly added URLs between pre-existing web documents. Similarly, virtually all new social/friendship links form between individuals that already have other friends and acquaintances.

Measurements show that in collaboration networks the internal links follow double preferential attachment, i.e. the probability for a new internal link to connect nodes with degrees *k* and *k'* is [20]

$$\Pi(k,k') \sim (A + Bk)(A + Bk') \cdot$$
(6.26)
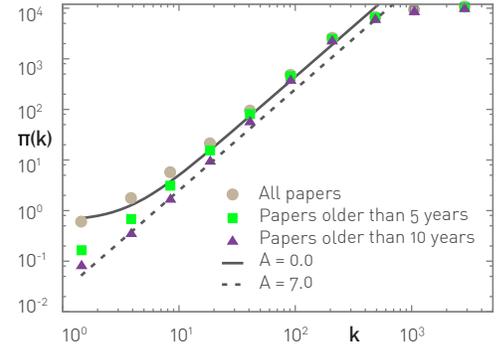
Figure 6.10
Initial Attractiveness

The cumulative preferential attachment function (5.21) for the citation network, capturing the citation patterns of research papers published from 2007 to 2008. The *π(k)* curve was measured using the methodology described in SECTION 5.6. The continuous line corresponds to initial attractiveness $A \sim 7.0$, while the dashed line corresponds to $A = 0$, i.e. the case without initial attractiveness. $A = 7$ implies that the probability of a new paper to be cited for the first time is comparable to the citation probability of a paper with seven citations. After [19].

To understand the impact of internal links we explore the limiting cases of (6.26):

- **Double Preferential Attachment (A=0)**

  Consider an extension of the Barabási-Albert model, where in each time step we add a new node with m links, followed by n internal links, each selected with probability (6.26) with A=0. Consequently the likelihood that a new link emerges is proportional to the degree of the nodes it connects. The degree exponent of the resulting network is [21, 22]

$$\gamma = 2 + \frac{m}{m + 2n} \, , \qquad (6.27)$$

  indicating that γ is between 2 and 3. This means that double preferential attachment *lowers the degree exponent* from 3 to 2, hence increasing the network's heterogeneity. Indeed, by preferentially connecting the hubs to each other, internal links make both hubs larger at the expense of the smaller nodes.

- **Random Attachment (B=0)**

  In this case the internal links are blind to the degree of the nodes they connect. Consequently the internal links are added between randomly chosen node pairs. Let us again consider the Barabási-Albert model, where after each new node we add n links between randomly selected nodes. The degree exponent of the obtained network is [22]

$$\gamma = 3 + \frac{2n}{m} \, . \qquad (6.28)$$

  Hence we have $\gamma \geq 3$ for any n and m combination, indicating that the resulting network will be more homogenous than the network without internal links. Indeed, randomly added internal links mimic the process observed in random networks, making the node degrees more similar to each other.

NODE DELETION

In many real systems nodes and links can disappear. For example, nodes are deleted from an organizational network when employees leave the company or from the WWW when web documents are removed. At the same time in some networks node removal is virtually impossible (Figure 6.11).

To explore the impact of node removal, we start from the Barabási-Albert model. In each time step we add a new node with m links and with rate r we remove a node. Depending on r, we observe three distinct scaling regimes [25-30]:

- **Scale-free Phase**

  For $r<1$ the number of removed nodes is smaller than the number of new nodes, hence the network continues to grow. In this case the network is scale-free with degree exponent
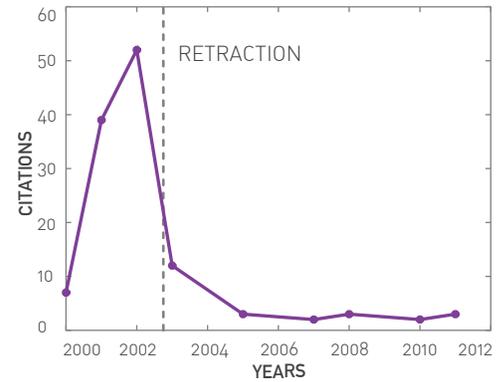

Figure 6.11
The Impossibility of Node Deletion

The citation history of a research paper by Jan Hendrik Schön published in *Science* [23] illustrates how difficult it is to remove a node from the citation network. Schön rose to prominence after a series of breakthroughs in the area of semiconductors. His productivity was phenomenal: In 2001 he has coauthored one research paper every eight days, published by the most prominent scientific journals, like *Science* and *Nature*.

Soon after Schön published a paper reporting a groundbreaking discovery on single-molecule semiconductors, researchers noticed that he reported for two experiments, carried out at different temperatures, identical noise [24]. The ensuing questions prompted Lucent Technologies, which ran Bell Labs where Schön worked, to start a formal investigation. Eventually Schön admitted falsifying data. Several dozens of his papers, like the one whose citation pattern is shown in this figure, were retracted.

While the papers' formal retraction lead to a dramatic drop in citations, the papers continue to be cited after their official "deletion" from the literature, as seen in the figure above. This indicates that it is virtually impossible to remove a node from the citation network.

$$\gamma = 3 + \frac{2r}{1-r} \, . \tag{6.29}$$

Hence, random node removal increases γ, homogenizing the network.

- **Exponential Phase**

  For r=1 nodes arrive and are removed at the same rate, hence the network has a fixed size (N=constant). In this case the network will loose its scale-free nature. Indeed, for $r \to 1$ we have $\gamma \to \infty$ in (6.29).

- **Declining Networks**

  For $r > 1$ the number of removed nodes exceeds the number of new nodes, hence the network declines (BOX 6.5). Declining networks emerge in several areas. For example, Alzheimer's research focuses on the progressive loss of neurons with age and ecology explores the role of gradual habitat loss [31-33]. A classical example of a declining network is the telegraph, that dominated long distance communication in the second part of the 19th century and early 20th century. It was once a growing network: In the United States the length of the telegraph lines grew from 40 miles in 1846 to 23,000 in 1852. Yet, following the second World War, the telegraph gradually disappeared.

The behavior of a network can be rather complex if node removal coexists with other elementary processes. This is illustrated in Figure 6.12, indicating that the joint presence of initial attractiveness and node deletion induces phase transitions between scale-free and exponential networks. Finally, note that node removal is not always random, but can depend on the removed node's degree (BOX 6.5).

In summary, in most networks nodes can disappear. Yet as long as the network continues to grow, its scale-free nature can persist. The degree exponent depends, however, on the details governing the node removal process.

ACCELERATED GROWTH

In the models discussed so far the number of links increased linearly with the number of nodes. In other words, we assumed that $L=\langle k \rangle N/2$, where $\langle k \rangle$ is independent of time. This is a reasonable assumption for many real networks. Yet, for some real networks the number of links grows faster than N, a phenomena called *accelerated growth*. For example the average degree of the Internet increased from $\langle k \rangle$=3.42 in November 1997 to 3.96 by December 1998 [34]; the WWW increased its average degree from 7.22 to 7.86 during a five month interval [35, 36]; in metabolic networks the average degree of the metabolites grows approximately linearly with the number of metabolites [37].
To explore the consequences of accelerated growth let us assume that in a growing network the number of links arriving with each new node follows [38-41]
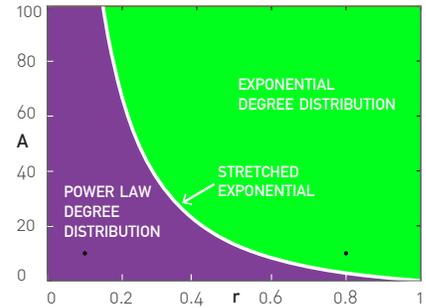
Figure 6.12
Phase Transitions Induced by Node Removal

The coexistence of node removal with other elementary processes can lead to interesting topological phase transitions. This is illustrated by a simple model in which the network's growth is governed by (6.23), and we also remove nodes with rate r [30]. The network displays three distinct phases, captured by the phase diagram shown above, whose axes are the node removal rate r and initial attractiveness A:

Subcritical Node Removal: $r < r^*(A)$
If the rate of node removal is under a critical value $r^*(A)$, shown as the white line on the figure, the network will be scale-free.

Critical Node Removal: $r=r^*(A)$
Once r reaches a critical value $r^*(A)$, the degree distribution turns into a stretched exponential (SECTION 4.A).

Exponential Networks: $r> r^*(A)$
The network looses its scale-free nature, developing an exponential degree distribution.

Therefore, the coexistence of multiple elementary processes in a network can lead to sudden changes in the network topology. To be specific, a continuous increase in the node removal rate leads to a phase transition from a scale-free to an exponential network.

BOX 6.5

The New York City garment industry offers a prominent example of a declining network (Figure 6.1). Its nodes are designers and contractors that are connected to each other by the annual coproduction of lines of clothing. As the industry decayed, the network has persistently shrunk: The network's largest connected component collapsed from 3,249 nodes in 1985 to 190 nodes in 2003. Interestingly, the network's degree distribution remained unchanged during this period. The analysis of the network's evolution uncovered several properties of declining networks [25]:

• **Preferential Attachment**
While overall the network was shrinking, new nodes continued to arrive. The measurements indicate that the attachment probability of these new nodes follows $\Pi(k) \sim k^{\alpha}$ with $\alpha=1.20 \pm 0.06$ (Figure 6.13a), offering evidence of superlinear preferential attachment (SECTION 5.7).

• **Link Deletion**
The probability that a firm lost a link follows $k(t)^{-\eta}$ with $\eta = 0.41 \pm 0.04$, i.e. it decreased with the firms' degree (Figure 6.13b). This documents a *weak-gets-weaker* phenomenon, when the less connected firms are more likely to loose links.
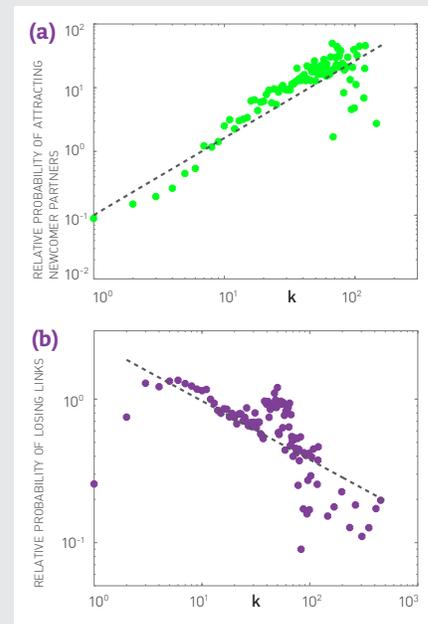
Figure 6.13
The Decline of the Garment Industry

(a) *Preferential attachment*. The probability that a newcomer firm added at time *t* connects to an incumbent firm with *k* links, relative to a random link addition. The dashed line has slope $\alpha=1.2$. If link addition were to be random, we would expect this quantity to be ≈1.

(b) *Link deletion*. The probability of deleting a link from a degree-*k* node, relative to random link removal. The dashed line has slope $\eta=0.41$. If link loss were to be random the relative probability should be ≈1 for any *k*.

$$m(t) = m_0 t^\theta. \qquad (6.30)$$

For $\theta=0$ each new node has the same number of links; for $\theta>0$, however-er, the network follows accelerated growth.

The degree exponent of the Barabási-Albert model with accelerated growth (6.30) is

$$\gamma = 3 + \frac{2\theta}{1-\theta} . \qquad (6.31)$$

Hence, accelerated growth pushes the degree exponent beyond $\gamma=3$, making the network more homogenous. For $\theta=1$ the degree exponent diverges, leading to *hyper-accelerating growth* [39]. In this case $\langle k \rangle$ grows linearly with time and the network looses its scale-free nature.

AGING

In many real systems nodes have a limited lifetime. For example, actors have a finite professional life span, defined as the period when they act in movies. So do scientists, whose professional lifespan typically cor-responds to the time frame during which they continue to publish sci-entific papers. In these networks nodes do not disappear abruptly, but fade away through a slow aging process, gradually reducing the rate at which they acquire new links [42-45]. Capacity limitations can induce a similar phenomena: If nodes have finite resources to handle links, once they approach their limit, they will stop accepting new links [43].

To understand the impact of aging we assume that the probability that a new node connects to node i is $\Pi(k_i, t-t_i)$, where t_i is the time node i was added to the network. Hence, $t-t_i$ is the node's age. Aging is often mod-eled by choosing [42]

$$\Pi(k_i, t - t_i) \sim k(t - t_i)^{-v} , \qquad (6.32)$$

where v is a tunable parameter governing the dependence of the at-tachment probability on the node's age. Depending on the value of v we can distinguish three scaling regimes:

- **Negative v**

  If $v < 0$, new nodes will link to older nodes. Hence, a negative v en-*hances* the role of preferential attachment. In the extreme case $v \to -\infty$ each new node connects to the oldest node, resulting in a hub-and-spoke topology (Figure 6.14a). The calculations show that the scale-free state persists in this regime, but the degree exponent drops under 3 (Figure 6.14e). Hence $v < 0$ makes the network more heterogeneous.

- **Positive v**

  In this case new nodes are encouraged to attach to younger nodes. In the extreme case $v \to \infty$ each node will connect to its immedi-

(a)

(b)

$v = -10$

$v = 0$

(c)

(d)

$v = 1$

$v = 10$

(e)
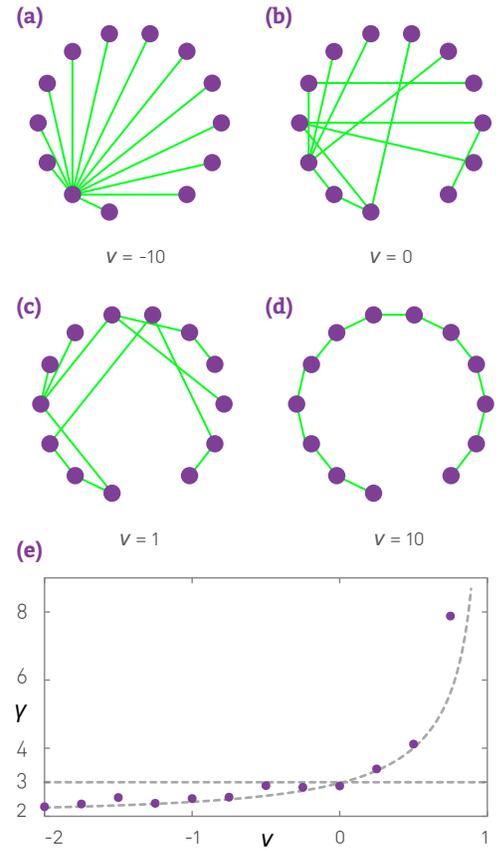
Figure 6.14
The Impact of Aging

(a-d) A schematic illustration of the expected network topologies for various aging expo-nents v in (6.32). In the context of a growing network we assume that the probability to attach to a node is proportional to $k\tau^{-v}$, where τ is the *age* of the node. For negative v nodes prefer to link to the oldest nodes, turning the network into a hub-and-spoke topology. For positive v the most recent nodes are the most attractive. For large v the network turns into a chain, as the last (i.e. the youngest) node is always the most attractive for the new node. The network is shown for $m=1$ for clarity but the degree ex-ponent is independent of m.

(e) The degree exponent γ vs. the aging expo-nent v predicted by the analytical solution of the aging model. The purple symbols are the result of simulations, each represent-ing a single network with $N=10,000$ and $m=1$. Redrawn after Ref. [42].

ate predecessor (Figure 6.14d). We do not need a very large v to experience the impact on aging: The degree exponent diverges as we approach $v=1$ (Figure 6.14e). Hence gradual aging homogenizes the network by shadowing the older hubs.

- **$v > 1$**

    In this case the aging effect overcomes the role of preferential attachment, leading to the loss of the scale-free property (Figure 6.14d).

In summary, the results discussed in this section indicate that a wide range of elementary processes can affect the structure and the dynamics of a growing network (Figure 6.15). These results highlight the true power of the evolving network paradigm: It allows us to address, using a mathematically self-consistent and predictive framework, the impact of various processes on the network topology and evolution.

FIG. 6.15

Elementary Processes Affecting the Network Topology

A summary of the elementary processes discussed in this section and their impact on the degree distribution. Each model is defined as extensions of the Barabási-Albert model.

ACCELERATED GROWTH

$m(t) = t^\theta$

$$\gamma = 3 + \frac{2\theta}{1-\theta}$$

AGING

$$\Pi(k) \sim (t - t_t)^{-\nu}$$

NODE DELETION

Node Removal Rate r

$r < r^*$
Scale-free
$$\gamma = 3 + \frac{2r}{1-r}$$

$r = r^*$
Stretched
Exponential

$r > r^*$
Exponential

n INTERNAL LINKS

$$\Pi(k, k') = (A + Bk)(A + Bk')$$

Double Preferential Attachment
$A = 0$

Random Internal Attachment
$B = 0$

$$\gamma = 2 + \frac{m}{m + 2n}$$

$\gamma < 3$

$$\gamma = 3 + \frac{2n}{m}$$

$\gamma > 3$

INITIAL ATTRACTIVENESS

$$\Pi(k) = A + k$$

$$\gamma = 3 + \frac{A}{m}$$

$$p_k \sim (k + A)^{-\gamma}$$

BARABÁSI-ALBERT MODEL

$$\Pi(k) \sim k$$

$$\gamma = 3$$
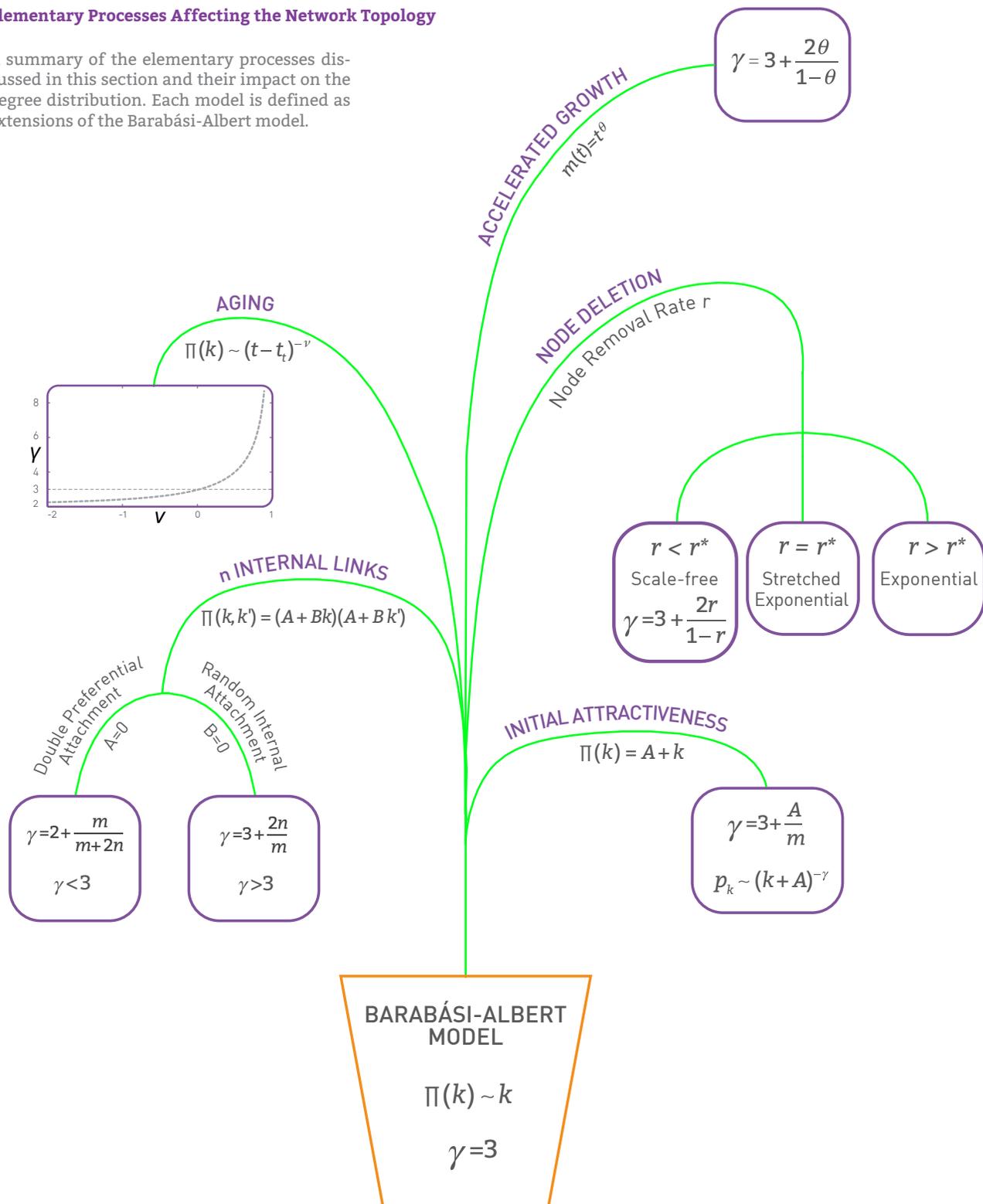
SUMMARY

As we showed in this chapter, rather diverse processes, from fitness to internal links and aging, can influence the structure of real networks. Through them we learned how to use the theory of evolving networks to predict the impact of various elementary events on a network's topology and evolution. The discussed examples allow us to draw a key conclusion: *if we want to understand the structure of a network we must first get its dynamics right. The topology is the bonus of this approach*.

The developed tools allow us to reflect on a number of issues that we encountered in the past chapters, from the correct fit to the degree distribution to the role of the different modeling frameworks. Next we briefly discuss some of these issues.

TOPOLOGICAL DIVERSITY

In CHAPTER 4 we discussed the difficulties we encounter when we attempt to fit a pure power law to the degree distribution of a real network. The roots of this problem became obvious in this chapter: If we account for the real dynamical processes that contribute to the evolution of a network, we expect systematic deviations from a pure power law. Indeed, we predicted several analytical forms for the degree distribution:

- **Power-Law**

  A pure power-law emerges if a growing network is governed by linear preferential attachment only, as predicted by the Barabási-Albert model. It is rare to observe such a pure power law in real systems. This idealized model represents the starting point for understanding the degree distribution of real networks.

- **Stretched Exponential**

  If preferential attachment is sublinear, the degree distribution follows a stretched exponential (SECTION 5.7). A similar degree-distribution can also appear under node removal at the critical point (Figure 6.12).

- **Fitness-induced Corrections**

  In the presence of fitness the precise form of p_k depends on the fitness distribution $\rho(\eta)$, which determines p_k via (6.6). For example, a uniform fitness distribution induces a logarithmic correction in p_k as predicted by (6.8). Other forms of $\rho(\eta)$ can lead to rather exotic forms for p_k.

- **Small-degree Saturation**

  Initial attractiveness adds a random component to preferential attachment. Consequently, the degree distribution develops a small-degree saturation, as seen in (6.24).

- **High-degree Cutoffs**

  Node and link removal, present in many real systems, can induce exponential high-degree cutoffs in the degree distribution. Furthermore, random node-removal can deplete the small-degree nodes, inducing a peak in p_k.

In most real networks several of the elementary processes discussed in this chapter appear together. For example, in the scientific collaboration network we have sublinear preferential attachment with initial attractiveness and the links can be both external and internal. As researchers have different creativity, fitness also plays a role, hence an accurate model requires us to know the appropriate fitness distribution. Therefore, the degree distribution is expected to display small degree saturation (thanks to initial attractiveness), stretched exponential cutoff at high degrees (thanks to sublinear preferential attachment), and some unknown corrections due to the particular form of the fitness distribution $\rho(\eta)$.

In general if wish to obtain an accurate fit to the degree distribution, we first need to build a generative model that analytically predicts the functional form of p_k. Yet, in many systems developing an accurate theory for p_k may be an overkill. It is often sufficient, instead, to establish if we are dealing with an exponentially bounded or a heavy tailed degree distribution (SECTION 4.9), as the system's properties will be primarily driven by this distinction.

MODELING DIVERSITY

The results of this chapter also allow us to reflect on the role of the network models encountered so far. We can categorize these models into three main classes (Table 6.1):

Static Models

The random network model of Erdős and Rényi (CHAPTER 3) and the small-world network model of Watts and Strogatz (BOX 3.8) have a fixed number of nodes, prompting us to call them *static models*. They both assume that the role of the network modeler is to place the links between the nodes using some random algorithm. To explore their properties we need to rely on combinatorial graph theory, developed by Erdős and Rényi. Both models predict a bounded degree distribution.

Generative Models

The configuration and the hidden parameter models discussed in SECTION 4.8 generate networks with a predefined degree distribution. Hence, these models are not mechanistic, in the sense that they do not tell us why a network develops a particular degree distribution. Rather, they help us understand how various network properties, from clustering to path lengths, depend on the degree distribution.

Evolving Network Models

These models capture the mechanisms that govern the time evolution of a network. The most studied example is the Barabási-Albert model, but equally insightful are the extensions discussed in this chapter, from the Bianconi-Barabási model to models involving internal links, aging, node and link deletion, or accelerated growth. These models are motivated by the observation that if we correctly capture all microscopic processes that contribute to a network's evolution, then the network's topological characteristics follow from that. To explore the properties of the networks generated by them, we need to use dynamical methods like the continuum theory and the rate equation approach.

Each of these modeling frameworks have their important role in network theory. The Erdős-Rényi model allows us to check if a certain network property could be explained by a pure random connectivity

MODEL CLASS	EXAMPLES	CHARACTERISTICS
Static Models	Erdős–Rényi Watts-Strogatz	• N fixed • p_k exponentially bounded • Static, time independent topologies
Generative Models	Configuration Model Hidden Parameter Model	• Arbitrary pre-defined p_k • Static, time independent topologies
Evolving Network Models	Barabási–Albert Model Bianconi-Barabási Model Initial Attractiveness Model Internal Links Model Node Deletion Model Accelerated Growth Model Aging Model	• p_k is determined by the processes that contribute to the network's evolution. • Time-varying network topologies

Table 6.1

Classes of Models in Network Science

The table summarizes the three main modeling frameworks used in network science, together with their distinguishing features.

pattern. If our interest is limited to the role of the network environment on some phenomena, like spreading processes or network robustness, the generative models offer an excellent starting point. If, however, we want to understand the origin of a network property, we must resort to evolving network models, that capture the processes that built the network in the first place.

HOMEWORK

6.1. Accelerated Growth

Calculate the degree exponent of the directed Barabási-Albert model with accelerated growth, assuming that the degree of the newly arriving nodes increases in time as $m(t) = t^\Theta$.

6.2. The *t*-Party Evolving Network Model

In the *t*-party gender play no role, hence each newcomer is allowed to invite only one other participants to a dance. However, attractiveness plays a role: More attractive participants are more likely to be invited to a dance by a new participant. The party evolves following these rules:

- Every participant corresponds to a node *i* and is assigned a time-independent attractiveness coefficient η_i.

- At each time step a new node joins the *t*-party.

- This new node then invites one already partying node to a dance, establishing a new link with it.

- The new node chooses its dance partner with probability proportional to the potential partner's attractiveness. If there are *t* nodes already in the party, the probability that node *i* receives a dance invitation is

$$\Pi_i = \frac{\eta_i}{\sum_j \eta_j} = \frac{\eta_i}{t\langle\eta\rangle},$$

where $\langle\eta\rangle$ is the average attractiveness.

(a) Derive the time evolution of the node degrees, telling us how many dances a node had.

(b) Derive the degree distribution of nodes with attractiveness η.

(c) If half of the nodes have $\eta = 2$, and the other half $\eta = 1$, what is the degree distribution of the network after a sufficiently long time?

6.3. Bianconi-Barabási Model

Consider the Bianconi-Barabási model with two distinct fitnesses,

$\eta = a$ and $\eta = 1$. To be specific, let us assume that the fitness follows the double delta distribution

$$\rho(\eta) = \frac{1}{2}\delta(\eta - a) + \frac{1}{2}\delta(\eta - 1) \text{ with } 0 \le a \le 1$$

- **(a)** Calculate the degree exponent, and its dependence on the parameter a.

- **(b)** Calculate the stationary degree distribution of the network.

6.4. Additive Fitness

Assume that the growth of a network is governed by preferential attachment with additive fitness

$$\Pi(k_i) \sim \eta_i + k_i,$$

where a different η_i is assigned to each node, chosen from a $\rho(\eta)$ fitness distribution. Calculate and discuss the degree distribution of the resulting network.

ADVANCED TOPICS 6.A
ANALYTICAL SOLUTION OF THE BIANCONI-BARABÁSI MODEL

The purpose of this section is to derive the degree distribution of the Bianconi-Barabási model [2, 15, 16,17]. We start by calculating

$$\left\langle \sum_j \eta_j k_j \right\rangle$$

over all possible realizations of the quenched fitnesses η. Since each node is born at a different time t_o, we can write the sum over j as an integral over t_o

$$\left\langle \sum_j \eta_j k_j \right\rangle = \int d\eta \rho(\eta) \eta \int_1^t dt_0 k_\eta(t,t_0).$$ (6.34)

By replacing $k_\eta(t, t_o)$ with (6.3) and performing the integral over t_o, we obtain

$$\left\langle \sum_j \eta_j k_j \right\rangle = \int d\eta \rho(\eta) \eta m \frac{t - t^{\beta(\eta)}}{1 - \beta(\eta)} .$$ (6.35)

The dynamic exponent $\beta(\eta)$ is bounded, i.e. $0 < \beta(\eta) < 1$, because a node can only increase its degree with time $(\beta(\eta) > 0)$ and $k_i(t)$ cannot increase faster than t $(\beta(\eta) < 1)$. Therefore in the limit $t \to \infty$ in (6.35) the term $t^{\beta(\eta)}$ can be neglected compared to t, obtaining

$$\left\langle \sum_j \eta_j k_j \right\rangle \overset{t \to \infty}{=} Cmt(1 - O(t^{-\varepsilon})) ,$$ (6.36)

where $\varepsilon = (1 - \max_\eta \beta(\eta)) > 0$ and

$$C = \int d\eta \rho(\eta) \frac{\eta}{1 - \beta(\eta)} .$$ (6.37)

Using (6.36) and the notation $k_\eta = k_\eta(t, t_o, \eta)$ we write the dynamic equation (6.2) as

$$\frac{\partial k_\eta}{\partial t} = \frac{\eta k_\eta}{Ct} ,$$ (6.38)

which has a solution of the form (6.3), given that

$$\beta(\eta) = \frac{\eta}{C} ,$$ (6.39)

confirming the self-consistent nature of the assumption (6.3).

To complete the calculation we need to determine C from (6.37). After substituting $\beta(n)$ with η/C, we obtain

$$1 = \int_0^{\eta_{max}} d\eta \rho(\eta) \frac{1}{\frac{C}{\eta} - 1},$$ (6.40)

where η_{max} is the maximum possible fitness in the system. The integral (6.40) is singular. However, since $\beta(\eta) = \eta/C < 1$ for any η, we have $C > \eta_{max}$, thus the integration limit never reaches the singularity. Note also that since

$$Cmt = \sum_j \eta_j k_j \le \eta_{max} \sum_j k_j = 2mt\eta_{max}$$ (6.41)

we have $C \le 2\eta_{max}$.

If there is a single dynamic exponent β, the degree distribution follows the power law $p_k \sim k^{-\gamma}$ with degree exponent $\gamma = 1/\beta + 1$. In the Bianconi-Barabási model we have a spectrum of dynamic exponents $\beta(\eta)$, thus p_k is a weighted sum over different power-laws.

To determine the degree distribution in the large N limit, we first calculate the number of nodes with fitness η and with degree greater than k, i.e. those that satisfy $k_\eta(t) > k$. Using (6.3) we find that this condition implies

$$t_0 < t \left(\frac{m}{k}\right)^{C/\eta}.$$ (6.42)

Exactly one node is added at each time step and each node has probability $\rho(\eta) d\eta$ to have fitness η. Therefore $t \left(\frac{m}{k}\right)^{C/\eta} \rho(\eta) d\eta$ nodes satisfy condition (6.42). To obtain the cumulative distribution function (the probability that a random node i has degree smaller or equal to k), we write

$$P(k) = P(k_i \le k) = 1 - P(k_i > k) \approx 1 - \frac{\int_0^{\eta_{max}} t \left(\frac{m}{k}\right)^{C/\eta} \rho(\eta) d\eta}{m_0 + t} \approx 1 - \int_0^{\eta_{max}} \left(\frac{m}{k}\right)^{C/\eta} \rho(\eta) d\eta,$$ (6.43)

where the last equation is valid asymptotically, for large t. The probability density function for the degree distribution is

$$p(k) = P'(k) = \int_0^{\eta_{max}} \frac{C}{\eta} m^{C/\eta} k^{-(C/\eta + 1)} \rho(\eta) d\eta,$$

recovering (6.6).

BIBLIOGRAPHY

[1] A.L. Barabási. *Linked: The New Science of Networks*. Perseus, Boston, 2001.

[2] G. Bianconi and A.-L. Barabási. Competition and multiscaling in evolving networks. Europhysics Letters, 54: 436-442, 2001.

[3] A.-L. Barabási, R. Albert, H. Jeong, and G. Bianconi. Power-law distribution of the world wide web. Science, 287: 2115, 2000.

[4] P.L. Krapivsky and S. Redner. Statistics of changes in lead node in connectivity-driven networks. Phys. Rev. Lett., 89:258703, 2002.

[5] C. Godreche and J. M. Luck. On leaders and condensates in a growing network. J. Stat. Mech., P07031, 2010.

[6] J. H. Fowler, C. T. Dawes, and N. A. Christakis. Model of Genetic Variation in Human Social Networks. PNAS, 106: 1720-1724, 2009.

[7] M. O. Jackson. Genetic influences on social network characteristics. PNAS, 106:1687–1688, 2009.

[8] S.A. Burt. Genes and popularity: Evidence of an evocative gene environment correlation. Psychol. Sci., 19:112–113, 2008.

[9] J. S. Kong, N. Sarshar, and V. P. Roychowdhury. Experience versus talent shapes the structure of the Web. PNAS, 105:13724-9, 2008.

[10] A.-L. Barabási, C. Song, and D. Wang. Handful of papers dominates citation. Nature, 491:40, 2012.

[11] D. Wang, C. Song, and A.-L. Barabási. Quantifying Long term scientific impact. Science, 342:127-131, 2013.

[12] M. Medo, G. Cimini, and S. Gualdi. Temporal effects in the growth of

networks. Phys. Rev. Lett., 107:238701, 2011.

[13] C. Venter et al. The sequence of the human genome. Science, 291:1304-1351, 2001.

[14] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. Science, 286:509-512, 1999.

[15] G. Bianconi and A.-L. Barabási. Bose-Einstein condensation in complex networks. Phys. Rev. Lett., 86: 5632–5635, 2001.

[16] C. Borgs, J. Chayes, C. Daskalakis, and S. Roch. First to market is not everything: analysis of preferential attachment with fitness. STOC'07, San Diego, California, 2007.

[17] S. N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin. Structure of growing networks with preferential linking. Phys. Rev. Lett., 85: 4633, 2000.

[18] C. Godreche, H. Grandclaude, and J.M. Luck. Finite-time fluctuations in the degree statistics of growing networks. J. of Stat. Phys., 137:1117-1146, 2009.

[19] Y.-H. Eom and S. Fortunato. Characterizing and Modeling Citation Dynamics. PLoS ONE, 6: e24926, 2011.

[20] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. Physica A, 311: 590-614, 2002.

[21] R. Albert, and A.-L. Barabási. Topology of evolving networks: local events and universality. Phys. Rev. Lett., 85:5234-5237, 2000.

[22] G. Goshal, L. Chi, and A.-L Barabási. Uncovering the role of elementary processes in network evolution. Scientific Reports, 3:1-8, 2013.

[23] J.H. Schön, Ch. Kloc, R.C. Haddon, and B. Batlogg. A superconducting field-effect switch. Science, 288: 656–8. 2000.

[24] D. Agin. *Junk Science: An Overdue Indictment of Government, Industry, and Faith Groups That Twist Science for Their Own Gain.* Macmillan, New York, 2007.

[25] S. Saavedra, F. Reed-Tsochas, and B. Uzzi. Asymmetric disassembly and robustness in declining networks. PNAS, 105:16466–16471, 2008.

[26] F. Chung and L. Lu. Coupling on-line and off-line analyses for random power-law graphs. Int. Math., 1: 409-461, 2004.

[27] C. Cooper, A. Frieze, and J. Vera. Random deletion in a scalefree random graph process. Int. Math. 1, 463-483, 2004.

[28] S. N. Dorogovtsev and J. Mendes. Scaling behavior of developing and decaying networks. Europhys. Lett., 52: 33-39, 2000.

[29] C. Moore, G. Ghoshal, and M. E. J. Newman. Exact solutions for models of evolving networks with addition and deletion of nodes. Phys. Rev. E, 74: 036121, 2006.

[30] H. Bauke, C. Moore, J. Rouquier, and D. Sherrington. Topological phase transition in a network model with preferential attachment and node removal. The European Physical Journal B, 83: 519-524, 2011.

[31] M. Pascual and J. Dunne, (eds). *Ecological Networks: Linking Structure to Dynamics in Food Webs*. Oxford Univ Press, Oxford, 2005.

[32] R. Sole and J. Bascompte. *Self-Organization in Complex Ecosystems*. Princeton University Press, Princeton, 2006.

[33] U. T. Srinivasan, J. A. Dunne, J. Harte, and N. D. Martinez. Response of complex food webs to realistic extinction sequencesm. Ecology, 88:671–682, 2007.

[34] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. ACM SIGCOMM Computer Communication Review, 29: 251-262, 1999.

[35] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, and A. Tomkins. Graph structure in the web. Computer Networks, 33: 309-320, 2000.

[36] J. Leskovec, J. Kleinberg, and C. Faloutsos, Graph evolution: Densification and shrinking diameters. ACM TKDD07, ACM Transactions on Knowledge Discovery from Data, 1:1, 2007.

[37] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. Nature, 407: 651–655, 2000.

[38] S. Dorogovtsev and J. Mendes. Effect of the accelerating growth of communications networks on their structure. Phys. Rev. E, 63: 025101(R), 2001.

[39] M. J. Gagen and J. S. Mattick. Accelerating, hyperaccelerating, and decelerating networks. Phys. Rev. E, 72: 016123, 2005.

[40] C. Cooper and P. Prałat. Scale-free graphs of increasing degree. Random Structures & Algorithms, 38: 396–421, 2011.

[41] N. Deo and A. Cami. Preferential deletion in dynamic models of web-like networks. Inf. Proc. Lett., 102: 156-162, 2007.

[42] S.N. Dorogovtsev and J.F.F. Mendes. Evolution of networks with aging of sites. Phys. Rev. E, 62:1842, 2000.

[43] A.N. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley. Classes of small-world networks. Proc. National Academy of Sciences USA, 97: 11149, 2000.

[44] K. Klemm and V. M. Eguiluz. Highly clustered scale free networks. Phys. Rev. E, 65: 036123, 2002.

[45] X. Zhu, R. Wang, and J.-Y. Zhu. The effect of aging on network structure. Phys. Rev. E, 68: 056121, 2003.