CHAPTER 6

EVOLVING NETWORKS

Figure 6.0 (front cover)
Network representation by Mauro Martino

SECTION 6.0

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 been dominating the search market. Yet Google, the third mover, soon not only became the leading search engine, but acquired links at such an incredible rate that by 2000 became the most connected node of the Web as well [1]. But its status didn't last: in 2011 Facebook, with an even later start, took over as the Web's biggest hub.

This competition for the top spot is by no means unique to the online world: the history of business is full of companies whose consumers were hijacked by a more successful latecomer. Take Apple, whose ingenious Newton handheld, introduced in 1987, was wiped off the market by Palm. A decade later Apple engineered a dramatic comeback, creating the iPad, that changed the concept of a handheld computer. If we view the market as a bipartite network whose nodes are products and whose links are purchasing decisions, we can say that Apple's links in the 1990s were rewired to Palm, only to be re-captured by Apple again a decade later. This competitive landscape highlights an important limitation of our current modeling framework: the network models we encountered so far cannot 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 $k(t) \sim t^{1/2}$ Eq. 5.6. 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 late nodes can never turn into the largest hubs.

In reality a node's growth does not depend on the node's age only. Instead webpages, companies, or actors have intrinsic qualities that influence the rate at which they acquire links. Some show up late and nevertheless grab most 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, and other processes not captured by the Barabási-Albert model, like node and link deletion or aging, affect the network topology.

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 other nodes. We will call this property fitness. Fitness is an individual's skill to turn a random encounter into a lasting friendship; it is a company's competence in acquiring consumers relative to its competition; 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 management quality and innovativeness 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 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 system, where η_j is a random number chosen from a 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 a pre-existing node i is proportional to the product of node i's degree k_i and its fitness η_i

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}. \qquad (6.1)$$

In Eq. 6.1 the dependence of Π_i on k_i captures the fact that higher-degree nodes are easier to encounter, hence we are more likely to link to them. The dependence 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, Eq. 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

Movie 6.1

The evolution of the Bianconi-Barabási model

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

Video courtesy of D. Wang.

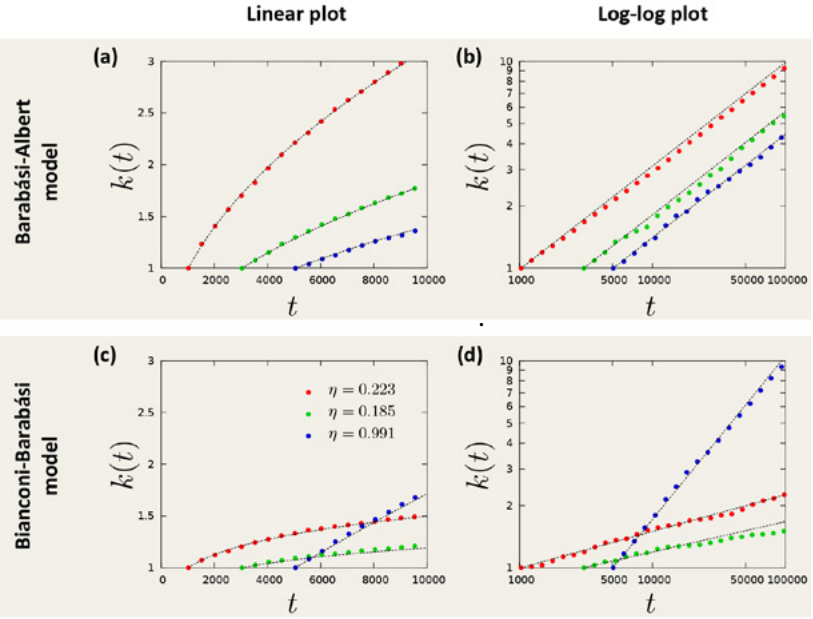

Figure 6.1
Competition in the Bianconi-Barabási model

(a) In the Barabási-Albert model all nodes increase their degree at the same rate, hence the earlier a node joins the network, the larger will be its degree. The figure shows the time dependence of the degree for nodes that arrived at different times, indicating that the later nodes are unable to pass the earlier nodes.

(b) Same as in (a) but in a log-log plot, demonstrating that each node follows the same growth law with identical dynamical exponents $\beta = 1/2$.

(c) In the Bianconi-Barabási model nodes increase their degree at a rate that is determined by their individual fitness. Hence a latecomer node (blue symbols) can overcome the earlier nodes.

(d) Same as in (c) but on a log-log plot, demonstrating that each node follows a growth curve with its own fitness-dependent dynamical exponent β, as predicted by **Eq. 6.3** and **Eq. 6.4**.

In (a)-(d) each curve corresponds to average over several independent runs using the same fitness sequence.

the nodes. We will call the model introduced above the *Bianconi-Barabási* model after the authors of the paper that introduced it [2, 3]. In the literature one may also account it as the *fitness model.*

DEGREE DYNAMICS

We can use the continuum theory to predict a node's temporal evolution in the model defined above. According to **Eq. 6.1**, the degree of node i changes at the rate

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

Let us assume that the time evolution of k_i follows a power law with a fitness-dependent exponent $\beta(\eta_i)$ **Fig. 6.1**,

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

Inserting **Eq. 6.3** into **Eq. 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 $\beta = 1/2$, indicating that the degree of each node increases as a square root of time. In contrast, according to **Eq. 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 each node that has a smaller fitness **BOX 6.1**. 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, predicting that

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

Eq. 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 apply Eq. 6.4 and Eq. 6.6 to calculate $\beta(\eta)$ and p_k for two different fitness distributions:

- Equal fitnesses

  When all fitnesses are equal, the Bianconi-Barabási model should reduce 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, Eq. 6.5 predicts $C = 2$. Using Eq. 6.4 we obtain $\beta = 1/2$ and Eq. 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 Eq. 6.5

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

  whose numerical solution is $C^* = 1.255$. According to Eq. 6.4, each node i has a different dynamic exponent, $\beta(\eta_i) = \eta_i / C^*$. Using Eq. 6.6 we obtain

  $$p_k \sim \int_1^0 d\eta \frac{C^*}{\eta} \frac{1}{k^{1+C^*/\eta}} \sim \frac{k^{-(1+C^*)}}{\ln k}, \tag{6.8}$$

  predicting that the degree distribution follows a power law with degree exponent $\gamma = 2.255$, affected by an inverse logarithmic correction $1/\ln k$.

Numerical support for these predictions is provided in Fig. 6.1 and Fig. 6.2. 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 Fig. 6.2 a indicates, the measured dynamical exponents are in excellent agreement with the prediction of Eq. 6.4. Fig. 6.2b also documents an agreement between Eq. 6.8 and the numerically obtained degree distribution.

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

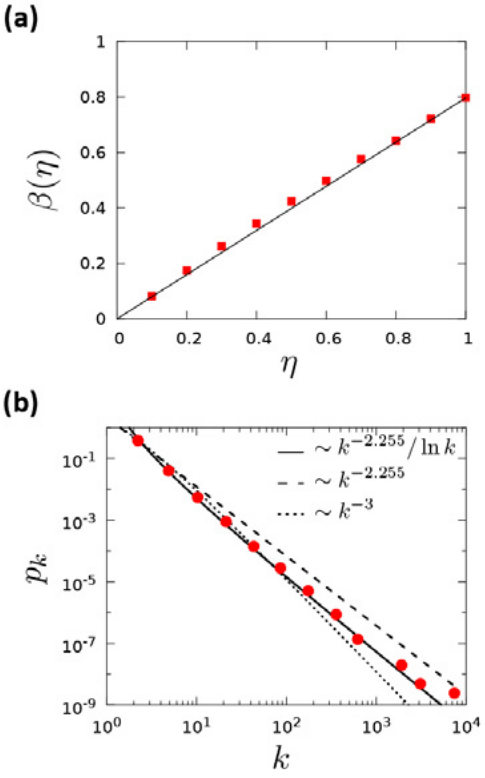

Figure 6.2
Characterizing the Bianconi-Barabási model

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

MEASURING FITNESS

Measuring the fitness of a node 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. Yet, our ability to determine the utility of a webpage is prone to errors: while a small segment of the population might find a webpage on sumo wrestling fascinating, most individuals are indifferent to it and some might even find it repulsive. Hence, different individuals will inevitably assign different fitnesses to the same node. Yet, according to Eq. 6.1 fitness reflects the network's collective perception of a node's importance relative to the other nodes. Thus, we can 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 conceptual 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 Eq. 6.3,

$$\log k_{n_i}(t,t_i) = \beta(n_i)\log t + \beta_i \,. \tag{6.9}$$

where $B_i = \log(m/t_i^{\beta(\eta_i)})$ is a time-independent parameter. Hence, the slope of $\log k_{n_i}(t, t_i)$ is a linear function of the dynamical exponent $\beta(\eta_i)$, which depends linearly on η_i according to Eq.6.4. Therefore, if we can track the time evolution of the degree for a large number of nodes, the distribution of the obtained growth exponent $\beta(\eta_i)$ will be identical with the fitness distribution $\rho(\eta)$. Such measurement were first carried out in the context of the WWW, relying on a dataset that crawled the links of about 22 million web documents per month 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 allow the determination of their growth exponent via Eq. 6.9. The obtained fitness distribution $\rho(\eta)$ has an exponential form Fig. 6.3, indicating that high fitness nodes are exponentially rare. This is somewhat unexpected, as one would be tempted to assume that on the web fitness varies widely: Google is probably significantly more interesting to Web

BOX 6.1

THE GENETIC ORIGINS OF FITNESS

Could fitness, an ability to acquire friends in a social network, have genetic origins? To answer this researchers examined the social network characteristics of 1,110 school-age twins [6, 7], using a technique previously developed to identify the heritability of a variety of traits and behaviors. The measurements indicate 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 as a friend).

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

This suggests that an individual's ability to acquire links, i.e. its fitness, is heritable. Hence, fitness may have genetic origins. This conclusion is also supported by research that associated a particular genetic variation with variation in popularity [8].

surfers than my personal webpage. Yet the exponential form of $\rho(\eta)$ indicates that most Web documents have comparable fitness. Consequently, the observed large differences in the degree of various web documents is the result of 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 Eq. 6.3 and Eq. 6.4, the relative difference between their degrees grows with time as

$$\frac{k_2 - k_1}{k_1} \sim t^{\frac{\eta_2 - \eta_1}{C}}. \tag{6.10}$$

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.

CASE STUDY: MEASURING THE FITNESS OF A SCIENTIFIC PUBLICATION

Eq. 6.9 assumes that Eq. 6.3 fully captures a Web document's temporal evolution. In some systems nodes follow a more complex dynamics, that we must account for when we try to measure their fitness. We illustrate this by determining the fitness of a research publication, allowing us to predict its future impact [11]. While most research papers acquire only a few citations, a small number of publications collect thousands and even tens of thousands of citations. These differences capture the considerable impact disparity characterizing the scientific enterprise BOX 6.3. These impact differences mirror differences in the novelty and the content of various publications. In general, we can write the probability that paper i is cited at time t after publication as [11]

$$\Pi_i \sim \eta_i c_i^t P_i(t), \tag{6.11}$$

where η_i is the paper's fitness, accounting for the perceived noveltyand importance of the reported discovery; c_t^i 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 again than less-cited contributions. The last term in Eq. 6.11 captures the fact that new ideas are integrated in subsequent work, hence the novelty of each paper fades with time [11, 12]. The measurements indicate that this decay has the log-normal form

$$P_i(t) = \frac{1}{\sqrt{2\pi\sigma_i t}} e^{-\frac{(\ln t - \mu_i)^2}{2\sigma_i^2}} \tag{6.12}$$

By solving the master equation behind Eq. 6.11, we obtain

$$c_i^t = m\left(e^{\frac{\beta\eta_i}{A}\Phi\left(\frac{\ln t - \mu_i}{\sigma_i}\right)}\right), \tag{6.13}$$

where

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

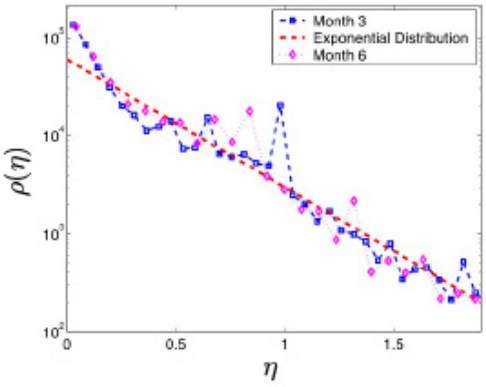

Figure 6.3
The fitness distribution of the WWW

The fitness distribution obtained from Eq. 6.9 by comparing the degree 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 Eq. 6.3. The slope of each curve is $\beta(\eta_j)$, which corresponds to the node's fitness η_i up to a multiplicative constant according to Eq. 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 indicates that the fitness distribution follows an exponential form. After [9].

BOX 6.2

ULTIMATE IMPACT

Citation counts offer only the momentary impact of a research paper. Therefore, they represent an inherently weak measure of long-term impact, as we do not know if a paper that acquired a hundred citations in two years has already had its run, or will continue to grow in impact, acquiring thousands more. Ideally we would like to predict how many citations will a paper acquire during its lifetime, or its ultimate impact. The citation model Eq. 6.11 and Eq. 6.14 allows us to determine the ultimate impact by taking the $t \to \infty$ limit in Eq. 6.13, finding [11]

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

is the cumulative normal distribution and m, β, and A are global parameters. Eq. 6.13 and Eq. 6.14 predict that the citation history of paper i is characterized by three fundamental parameters: the relative fitness $\eta_i' \equiv \eta_i \beta / A$, measuring a paper's importance relative to other papers; the immediacy μ_i, governing the time for a paper to reach its citation peak and the longevity σ_i, capturing the decay rate.

We fit Eq. 6.13 to the citation history of individual papers published by a given journal to obtain the journal's fitness distribution Fig. 6.4. We find that *Cell* has a fitness distribution shifted to the right, indicating that *Cell* papers tend to have high fitness. By comparison the fitness of papers published in *Physical Review* are shifted to the left, indicating that the journal publishes fewer high impact papers.

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

Eq. 6.15 predicts that despite the myriad of factors that contribute to the success and 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 Eq. 6.13 to a paper's existing citation history, we can use Eq. 6.15 to predict the ultimate impact of a publication.

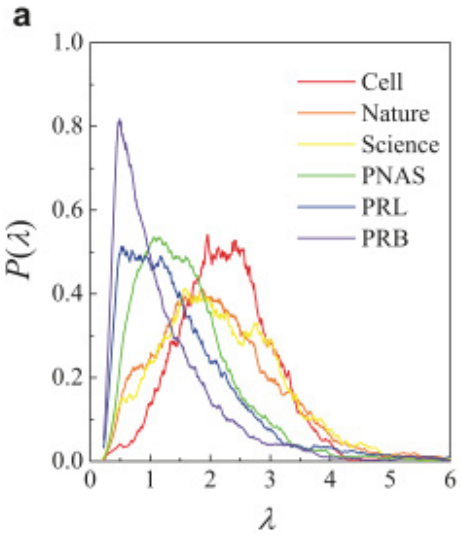

Figure 6.4
Fitness distribution of research papers

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

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

BOX 6.3

THE TOP ONE PERCENT

The "one percent" phrase has dominated the discourse during the 2012 US presidential election, reminding everyone that one percent of the population earns a disproportional 17.42% of the total US income. To those familiar with power laws this is hardly surprising: it is a consequence of the fat-tailed nature of the income distribution. Therefore, the "one percent" phenomenon is present in any quantity that follows a power law, from the links of the WWW to scientific impact [10].

The "one percent" debate is not as much about the magnitude of the income disparity, but its trends: income disparity dropped between 1940 and 1970, only to skyrocket again in the past decades (see black line). As models explaining the distribution of income predict a time-invariant distribution, the observed changes offer evidence of endogenous shifts in the share of the top one percent. As the red line indicates, the impact disparity in physical sciences has also been rising steadily over the past century. Indeed, while in 1930 a year after publication the top 1% of papers got only about 5% of the citations, today the magnitude of this impact disparity is comparable to the income disparity.

This shift of the bulk of the citations to a few of publications may reflect the fact that while the number of research papers exploded, the time we devote to reading them has not. Hence, we increasingly rely on crowdsourcing to discover relevant work, a process that favors the highly cited publications.

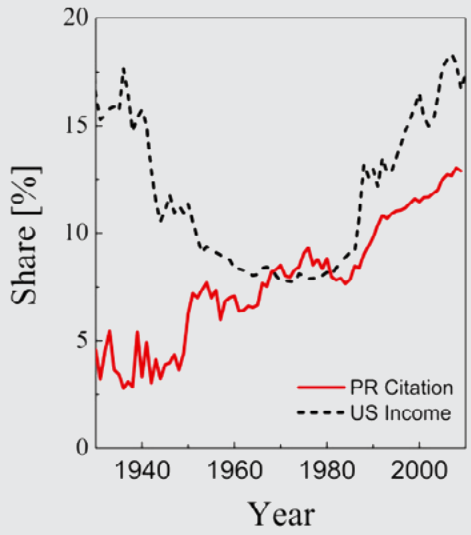

Figure 6.5
The 1% of Science

The share of citations in a given year received by the top 1% of all papers published during the previous year in Physical Review. The data captures the citation history of 463,348 papers published between 1893 and 2009 (red line). Also shown is the fraction of income earned by the top 1% of the population in US (black dashed line).

BOSE-EINSTEIN CONDENSATION

In the previous section we found that the Web's fitness distribution follows a simple exponential Fig. 6.3, while the fitness of research papers follows a peaked distribution Fig. 6.4. The diversity of the observed fitness distributions raises an important question: how does the network topology depend on the precise shape of $\rho(\eta)$? Technically, the answer is provided by Eq. 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 a Bose-Einstein condensation BOX 6.5, with significant consequences on the network topology [13]. We start by establishing a formal link between the Bianconi-Barabási model and a Bose gas, whose properties have been extensively studied in physics Fig. 6.5:

- **Fitness → Energy**: to each node with fitness η_i we assign an energy ε_i using

$$\varepsilon_i = \frac{1}{\beta_T} \log \eta_i \,. \tag{6.16}$$

  In physical systems β_T plays the role of the inverse temperature. Hence, we use the subscript T to distinguish β_T from the dynamic exponent β. According to Eq. 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 between nodes i and j we add a particle at the energy levels ε_i and ε_j, respectively.

- **Nodes → Energy levels:** the arrival of a new node with m links corresponds to adding a new energy level ε_j and $2m$ new particles to the Bose gas. Half of these particles land on level ε_j, corresponding to the links that start from the node j; the remaining m particles are distributed between the energy levels that correspond to the nodes to which the new node 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 fol-

Movie 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. Consequently this high fitness node attracts most links, forcing the system to undergo a Bose-Einstein condensation.

Video courtesy of D. Wang.

lows a Bose statistics, a formula derived by Satyendra Nath Bose in 1924, representing a fundamental result in quantum statistics BOX 6.6. Consequently, the links of the fitness model behave like subatomic particles in a quantum gas. This mapping to a Bose gas is exact and predicts the existence of two distinct phases [13, 14]:

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 will inevitably become the largest hub, in the scale-free phase the fittest node is not significantly bigger than the next fittest node.

BOX 6.4

BOSE-EINSTEIN CONDENSATION

In classical physics atoms can be distinguished and individually numbered, like the numbered balls used to pick the winning number in lottery. In the subatomic world particles differ in our ability to distinguish them: Fermi particles, like electrons, can be distinguished; in contrast Bose particles, like photons, are indistinguishable. Distinguishability impacts the energy of a particle. 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. In quantum mechanics energy is quantized, which means that it can only take up discrete (quantized) values. This is where distinguishability matters: the distinguishable Fermi particles are forbidden to have the same energy. Hence, only one electron can occupy a given energy level Fig. 6.7a. As Bose particles cannot be distinguished, many can crowd on the same energy level Fig. 6.7b.

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 Fig. 6.7a, 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 Fig. 6.7c. However, as any number of Bose particles can share the same energy level, they can all crowd at the lowest energy Fig. 6.7d. Hence, no matter how much "Bose water" 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.

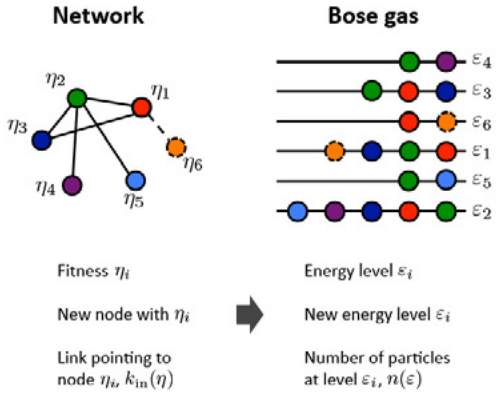

Figure 6.6
Mapping networks to a Bose gas

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

Right: The mapping assigns an energy level ε to each fitness η, resulting in a Bose gas with random energy levels. A link from node i to node j corresponds to two particles, one at level ε_i and the other at level ε_j.

Growth: The network grows by adding a new node, like the node with fitness η_6, at each time step. The new node connects to m=1 other nodes (dashed link), chosen randomly following Eq. 6.1. In the Bose gas this results in the addition of a new energy level ε_6 (dashed line), populated by two particles, and the deposition of another particle at ε_1, the energy level to which η_6 connects to.

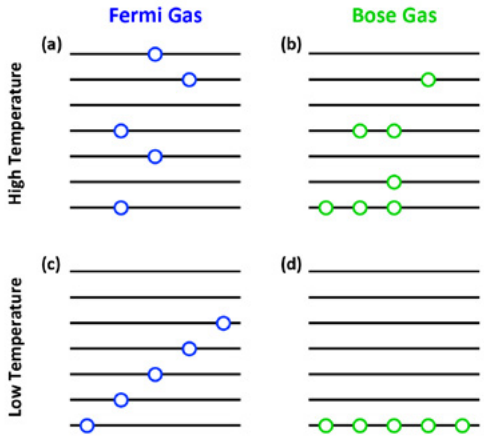

Figure 6.7
Bose and Fermi statistics

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 the difference between the two gases. At low temperatures, however, particles want to occupy the lowest possible energy and the difference between the two gasses becomes significant.

Indeed, at any moment the degree distribution follows a power law, indicating that the largest hub is closely followed by a few slightly smaller hubs, with almost as many links as the fittest node Fig. 6.8a. The uniform fitness distribution discussed in the previous section results in a scale-free network.

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 $\rho(\eta)$ BOX 6.7. In a Bose-Einstein condensate all particles crowd to the lowest energy level, leaving the rest of the energy levels unpopulated BOX 6.5. In a network this means that the fittest node grabs a finite fraction of the links, turning into a super-hub Fig. 6.8b, and the network develops a hub-and-spoke topology. In these networks the rich-gets-richer process is so dominant that becomes a winner takes-all phenomenon. Consequently, the network will loose its scale-free nature.

In summary, the precise shape of the fitness distribution, $\rho(\eta)$, plays an important role in shaping the topology of a growing network. While most fitness distributions (like the uniform distribution) lead to a power law degree distribution, 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 that generates 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, turning it into a star-like topology BOX 6.8.

BOX 6.5

FROM FITNESS TO A BOSE GAS

In the context of the Bose gas of Fig. 6.6 the probability that a particle lands on level i is given by

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

Hence, the rate at which the energy level ε_i accumulates particles is [13]

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

where $k_i(\varepsilon_i, t, t_i)$ is the occupation number of level i Fig. 6.6 and

$$Z_t \equiv \sum_{j=1}^{t} t e^{-\beta_T \varepsilon_j} k_j(\varepsilon_i,t,t_j) \qquad (6.18a)$$

is the partition function. The solution of Eq. 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, $\deg(\varepsilon)$ is the degeneracy of the energy level ε. Eq. 6.20 suggests that in the limit $t \to \infty$ the occupation number, representing the number of particles with energy ε, follows the well-known Bose statistics

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

This concludes the mapping of the fitness model to a Bose gas, indicating that the node degrees in the fitness model follow Bose statistics.

BOX 6.6

FITNESSES DISTRIBUTION LEADING TO BOSE-EINSTEIN CONDENSATION

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

(a) $\rho(\eta)$ must have a maximum $\eta_{\max}$.This means that η needs to have a clear upper bound.

(b) $\rho(\eta_{max})$=0, i.e. the system requires an infinite time to reach $\eta_{\max}$.

The uniform distribution, $\eta \in [0, 1]$ satisfies (a), as it is bounded, having $\eta_{\max}$=1. It fails, however, the criteria (b), as it can reach $\eta_{\max}$= 1 with a finite probability. Consequently, we cannot observe a Bose-Einstein condensation in this case. A fitness distribution that can lead to a Bose-Einstein condensation is

$$\rho(\eta) = (1-\eta)^{\xi} \qquad (6.22)$$

satisfying both (a) and (b). Indeed, $\eta_{\max} = 1$ and $\eta(1) = 0$, which is the reason why, upon varying ζ, we can observe Bose-Einstein condensation Fig. 6.8. Indeed, the existence of the solution of Eq. 6.20 depends on the functional form of the energy distribution, $g(\varepsilon)$, determined by the $\rho(\eta)$ fitness distribution. Specifically, if Eq. 6.22 has no non-negative solution for a given $g(\varepsilon)$, we observe a Bose-Einstein condensation, indicating that a finite fraction of the particles agglomerate at the lowest energy level.

BOX 6.7

MICROSOFT AND BOSE-EINSTEIN CONDENSATION

Think of the operating systems (OS) that run on each computer as nodes that compete for links in terms of users or computers. Each time a user installs Windows on his or her computer, a link is added to Microsoft. If a fit-gets-rich behavior of scale-free networks prevails in the marketplace, there should be a hierarchy of operating systems, such that the most popular node is followed closely by several less popular nodes.

In the OS market, however, such hierarchy is absent. True, Windows is not the only available operating system. All Apple products run Mac OS; DOS, the precursor of Windows, is still installed on some PCs; Linux, a free operating system, continues to gain market share and UNIX runs on many computers devoted to number crunching.

But all these operating systems are dwarfed by Windows, as in 2010 its different versions were humming on a whopping 86 percent of all personal computers. The second most popular operating system had only a 5 percent market share. Hence the OS market carries the signatures of a network that has undergone Bose-Einstein condensation [1].

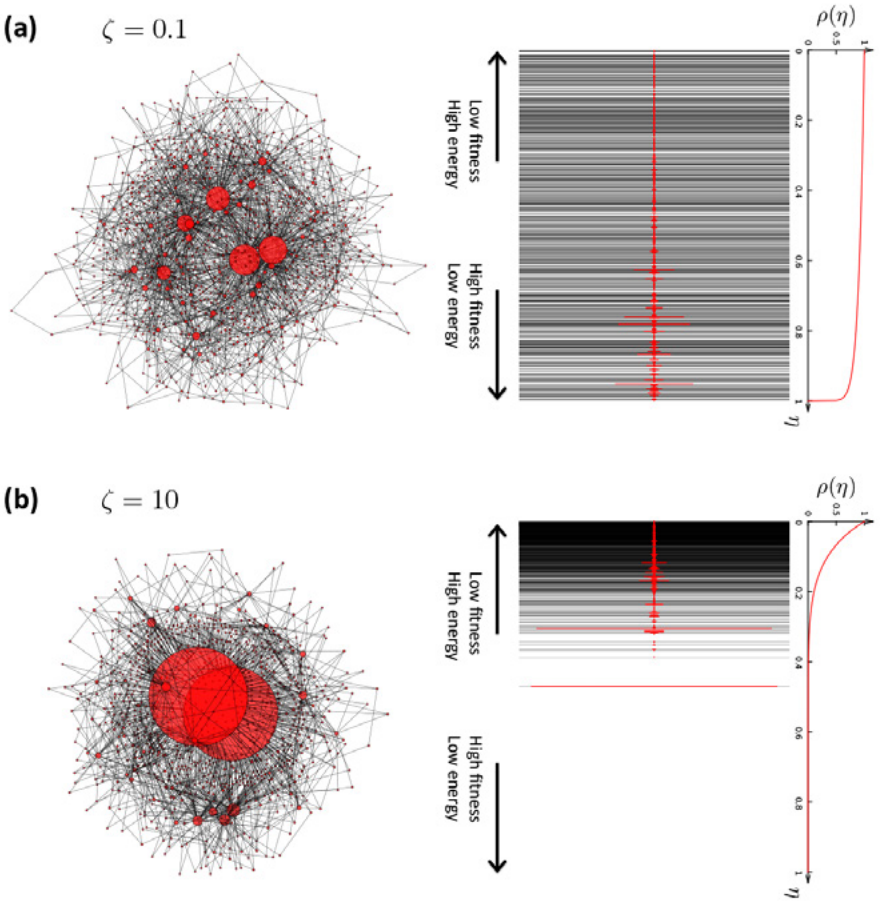

Figure 6.8
Bose Einstein Condensation in Networks

Left panels: (a) A scale-free network and (b) a network that has undergone a Bose-Einstein condensation generated by the fitness model with $\rho(\eta)$ following **Eq. 6.22**.

Middle panels: the energy levels (black lines) and the deposited particles (red dots) for a network with m=2 and N=1,000. Each energy level corresponds to the fitness of a node on the network shown in the left. Each link connected to a node is represented by a particle on the corresponding energy level.

Right panels: the fitness distribution $\rho(\eta)$, given by **Eq. 6.22**, illustrating the difference in the shape of the two $\rho(\eta)$ functions. The difference is determined by the parameter ζ.

EVOLVING NETWORKS

The Barabási-Albert model is a minimal model, its main purpose being to capture the core mechanisms responsible for the emergence of the scale-free property. Consequently, it has several well-known limitations:

(i) It predicts $\gamma = 3$ while the experimentally observed degree exponents vary between 2 and 4 Table 4.1.

(ii) It predicts a pure power-law degree distribution, while real systems are characterized by various deviations from a power law, like small-degree saturation or high-degree cutoff BOX 4.18.

(iii) It 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 in the network science community, clarifying how various elementary processes influence the network topology. The purpose of this section is to systematically extend the Barabási-Albert model to capture the wide range of phenomena shaping the structure of real networks.

INITIAL ATTRACTIVENESS

In the Barabási-Albert model an isolated node cannot acquire links, as according to preferential attachment Eq. 4.1 the likelihood that a new node attaches to a k=0 node is strictly zero. In real networks, however, even isolated nodes acquire links. Indeed, each new research paper has a finite probability of being cited or a person that moves to a new city will quickly acquire acquaintances. In growing networks zero-degree nodes can acquire links if we add a constant to the preferential attachment function Eq. 4.1, obtaining

$$\Pi(k) \sim A + k \,. \tag{6.23}$$

In Eq. 6.23 the parameter A is called initial attractiveness. As $\Pi(0) \sim A$,

initial attractiveness represents the probability that a node will acquire its first link. We can detect the presence of initial attractiveness in real networks by measuring $\Pi(k)$ Fig 6.9. Once present, initial attractiveness has two immediate consequences:

- **Increases the degree exponent:** If in the Barabási-Albert model we place Eq. 4.1 with Eq. 6.23, the degree exponent becomes [15, 16]

$$\gamma = 3 + \frac{A}{m} . \tag{6.24}$$

  By increasing γ, initial attractiveness makes a network more homogeneous and reduces 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, weakening the role of preferential attachment. For high-degree nodes the A term in Eq. 6.23 is negligible.

- **Generates a small-degree cutoff**: The solution of the continuum equation indicates that the degree distribution of a network governed by Eq. 6.23 does not follow a pure power-law, but has the form

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

Therefore, initial attractiveness induces a small-degree saturation at $k<A$. This saturation is again rooted in the fact that initial attractiveness enhances the probability that new nodes link to the small-degree nodes, which decreases the number of nodes with small k. For high degrees ($k \gg A$), where initial attractiveness loses its relevance, the degree distribution continues to follow a power law.

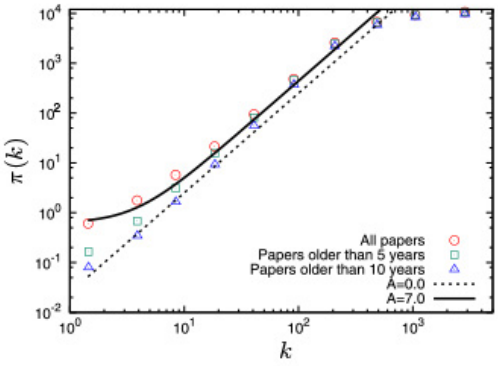

Figure 6.9
Initial Attractiveness

Cumulative preferential attachment function

$$\pi(k) = \sum_{k' \le k} \pi(k')$$

for the citation network, capturing the citation patterns of research papers published from 2007 to 2008. The $\pi(k)$ curve was measured using the methodology described in SECTION 5.7. The continuous line corresponds to $C(k+A)^{-\gamma}$ with initial attractiveness $A \sim 7.0$. The dashed line corresponds to $A = 0$, i.e. the case without attractiveness. After [17].

INTERNAL LINKS

In many networks most new links 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 existing web documents. Similarly, virtually all new social/friendship links form between individuals that already have other acquaintances and friends. Measurements show that in collaboration networks internal links follow double preferential attachment, i.e. the probability for a new internal link to connect two nodes with degree k and k' is [18]

$$\Pi(k,k') \sim (A + Bk)(A' + B'k') . \tag{6.26}$$

We explore several limiting cases to understand the impact of internal links:

- **Double preferential attachment (A=A′=0):** In this case both ends of a new link are chosen proportional to the degree of the nodes they connect. 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 Eq. 6.26 with A=A'=0. The

degree exponent of the resulting network is [19, 20]

$$\gamma = 2 + \frac{m}{m+2n}, \tag{6.27}$$

indicating that γ varies 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, it simultaneously helps both hubs to grow faster than they do in the Barabási-Albert model.

- **Random attachment (B=B′=0):** In this case the internal links are blind to the degree of the nodes they connect, implying that they are added between randomly chosen node pairs. Let us again consider the Barabási-Albert model, where after each new node we add n randomly selected links. In this case the degree exponent becomes [20]

$$\gamma = 3 + \frac{2n}{m}. \tag{6.28}$$

Hence, $\gamma \geq 3$ for any n, indicating that the resulting network will be more homogenous than the network generated by the Barabási-Albert model. 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 systematically disappear, leading to node or link deletion. 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 Fig. 6.10.

To explore the impact of node removal, let us start again from the Barabási-Albert model. In each time step we add a new node with m links and with probability r we remove a node. The observed topologies depend on the value of r [23, 24, 25, 26, 27, 28]:

- **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 degree exponent has the value

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

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

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

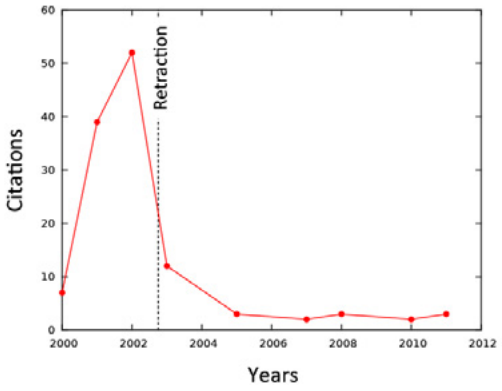

Figure 6.10
The impossibility of node removal

The citation history of a research paper by Jan Hendrik Schön published in *Science* [21]. Schön rose to prominence after a series of apparent breakthroughs in the area of semico ductors. Schön's findings were published by prominent scientific journals, like *Science* and *Nature*. His productivity was phenomenal: in 2001 he has coauthored one research paper every eight days. However, research groups around the world had difficulty reproducing some of his results.

Soon after Schön published a paper reporting a groundbreaking discovery on single-molecule semiconductors, researchers noticed that two experiments carried out at very different temperatures had identical noise [22].

The ensuing questions prompted Lucent Technologies, which ran the storied Bell Labs where Schön worked, to start a formal investigation. Eventually, Schön admitted falsifying some data to show more convincing evidence for the behavior that he observed. Several dozens of his papers, like the one whose citation pattern is shown in this figure, were retracted and the *University of Konstanz* revoked his PhD degree for "dishonorable conduct."

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

- **Declining networks:** For $r > 1$ the number of removed nodes exceeds the number of new nodes, hence the network declines BOX 6.10. Declining networks are important in several areas: Alzheimer's research focuses on the progressive loss of neurons with age and ecology focuses on the role of gradual habitat loss [29, 30, 31]. 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.

Note that node removal is not always random but can depend on the removed node's degree BOX 6.9. Furthermore, the behavior of a network can be rather complex if additional elementary processes are considered, inducing phase transitions between scale-free and exponential networks Box 6.10.

In summary, in most networks some 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 detail governing the node removal process.

ACCELERATED GROWTH

In the models discussed so far the number of links increases linearly with the number of nodes. In other words, we assumed that $L=\langle k\rangle N$, where $\langle k\rangle$ is independent of time. This is a reasonable assumption for many real networks. Yet, some real networks experience accelerated growth, meaning that the number of links grows faster than N, hence $\langle k\rangle$ increases. For example the average degree of the Internet increased from $\langle k\rangle$=3.42 in November 1997 to 3.96 by December 1998 [32]; the WWW increased its average degree from 7.22 to 7.86 during a five month interval [33, 34]; in metabolic networks the average degree of the metabolites grows approximately linearly with the number of metabolites [35]. To explore the consequence of such accelerated growth let us assume that in a growing network the number of links arriving with each new node follows [36, 37, 38, 39]

$$m(t) = m_0 t^{\theta}. \tag{6.30}$$

For θ=0 each node has the same number of links; for θ>0, however, the network follows accelerated growth. The degree exponent of the Barabási-Albert model with accelerated growth Eq. 6.30 is

$$\gamma = 3 + \frac{2\theta}{1-\theta}\,. \tag{6.31}$$

Hence, accelerated growth increases the degree exponent beyond γ=3,

making the network more homogenous. For θ=1 the degree exponent diverges, leading to hyper-accelerating growth [37]. 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, capturing the period when they still act in movies. So do scientists, whose professional lifespan corresponds to the time frame they continue to publish scientific papers. These nodes do not disappear abruptly, but fade away through a slow aging process, gradually reducing the rate at which they acquire new links [40, 41, 42, 43]. 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 [41].

To understand the impact of aging let us assume that the probability that a new node connects to node i is $\Pi(k, t-t_i)$, where ti is the time node i was added to the network. Hence, $t-t_i$ is the node's age. In analytical calculations slow aging is often modeled by choosing [40]

$$\Pi(k,t-t_i) \sim k(t-t_i)^{-\nu}, \tag{6.32}$$

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

- **For negative ν** the older is node i, the more likely that a new node will link to it. Hence, $\nu < 0$ enhances the role of preferential attachment. In the extreme case $\nu \to -\infty$, each new node will only connect to the oldest node, resulting in a hub-and-spoke topology Fig. 6.11a. The calculations show that the scale-free state persists in this regime, but the degree exponent drops under 3. Hence, $\nu < 0$ makes the network more heterogeneous

- **A positive ν** will encourage the new nodes to attach to younger nodes. In the extreme case $\nu \to \infty$ each node will connect to its immediate predecessor Fig. 6.11a. We do not need a very large ν to experience the impact on aging: the degree exponent diverges as we approach ν=1. Hence gradual aging homogenizes the network by shadowing the older hubs.

- **For $\nu > 1$** the aging effect overcomes the role of preferential attachment, leading to the loss of the scale-free property.

In summary, the results discussed in this section indicate that a wide range of elementary processes can affect the structure of a growing network Table 6.1. These results highlight the true power of the evolving network paradigm: it allows us to address, using a mathematically predictive framework, the impact of a wide range of elementary processes on the network topology and evolution.

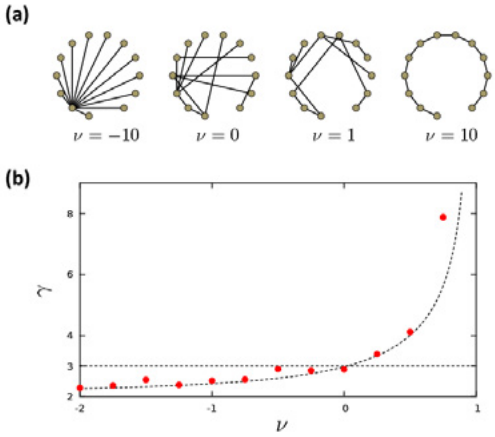

Figure 6.11
The impact of aging

(a) A schematic illustration of the expected network topologies for various aging exponents ν. In the context of a growing network model we assume that the probability to attach to a node is proportional to $k\tau^{-\nu}$, where τ is the age of the node. For negative ν nodes prefer the oldest nodes, turning the network into a hub-and-spoke topology. For positive ν the most recent nodes are the most attractive. Hence for large ν the network turns into a chain, as the last (hence the youngest) node is the most attractive for the new node. The network is shown for m=1 for clarity but note that the degree exponent is independent of m.

(b) The degree exponent γ vs the aging exponent ν, as predicted by the analytical solution of the rate equation. The red symbols are the result of simulations, each representing a single network of N=10,000 and m=1. The degree exponent is estimated using the method described in **CHAPTER 4**. Redrawn after Ref. [40].

BOX 6.8

DECLINING NETWORKS AND FASHION

The properties of declining networks is well illustrated by the New York City garment industry, whose 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. This is illustrated by the fate of the largest connected component, that collapsed from 3,249 nodes in 1985 to 190 nodes in 2003. Interestingly, the network's degree distribution remained relatively unchanged during this period. The analysis of the network's evolution allowed researchers to uncover several interesting properties of declining networks [23]:

- **Preferential Attachment:** While overall the network was shrinking with time, new nodes continued to arrive. The measurements indicate that the attachment probability of these new nodes follows $\Pi(k) \sim k^{\alpha}$ with α=1.20 ± 0.06 **PANEL A**, offering evidence of superlinear preferential attachment.

- **Link deletion:** The measurements also show that the probability that a firm looses a link decreased proportionally with the firms' degree, as $k(t)-\eta$ with η=0.41±0.04. This documents a weak-gets-weaker phenomenon where the less connected firms are more likely to loose their links.

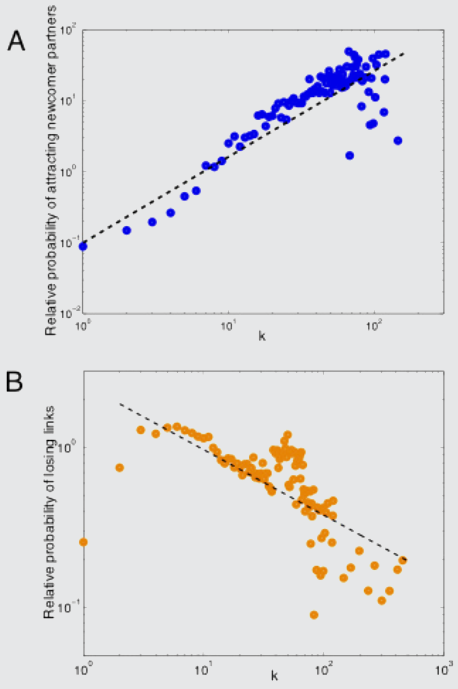

Figure 6.12
The decline of the garment industry

(a) Preferential attachment. The relative probability $\Pi(k)$ that a newcomer firm added at time t connects to an incumbent firm with k links. The dashed line has slope α=1.2.

(b) Link deletion. The relative probability, $R_{k(t)}$ of deleting a link from a degree node, compared with random link removal. The dashed line has slope η=0.41.

If link addition and removal were to be random, we would expect $\Pi(k) \sim 1$ and $R_{k(t)} \sim 1$ for all k. After [23].

Figure 6.13
Garment district

The Garment District is a Manhattan neighborhood located between Fifth Avenue 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 sculpture and a sculpture of a tailor, located in the heart of the district, pay tribute to the neighborhood's past and present.

PROCESS	PROCESS	γ	OBSERVATIONS
Preferential attachment	$\Pi(k) \sim k$	3	
Initial attractiveness	$\Pi(k) = A + k$	$3 + \frac{A}{m}$	Small degree cutoff $p_k \sim (k + A)^{-\gamma}$
Internal links	$\Pi(k, k') = (A + Bk)(A' + B'k')$	$2 + \frac{m}{m+2n}$	Double preferential attachment $A = A' = 0, B \quad 0, B' \quad 0$
		$3 + \frac{2n}{m}$	Random internal attachment $B = B' = 0$
Node (link) deletionN	ode removal rate r	$3 + \frac{2r}{1-r}$	Scale-free for $r < r^*$ Stretched exponential for $r = r^*$ Exponential for $r > r^*$
Accelerated growth	$m(t) = t^{\Theta}$	$\frac{3-\theta}{1-\theta}$	For $\theta = 1$ we have hyper-accelerated growth and the scale-free state disappears.
Aging	$\Pi(k) \sim (t - t_i)^{-\nu}$	See Figure 6.11	For $\nu > 1$ the network looses its scale-free topology.

Table 6.1 Elementary processes
A summary of the various elementary processes discussed in this section and their impact on the degree distribution.

BOX 6.9

NODE REMOVAL INDUCED PHASE TRANSITIONS

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 Eq. 6.23, i.e. preferential attachment with initial attractiveness, and we also remove nodes with rate r. The network displays three distinct phases, captured by the phase diagram shown below:

Subcritical node removal ($r < r^*(A)$): If the rate of node removal is under a critical value $r^*(A)$, 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 SECT. 4.A.

Exponential networks ($r > r^*(A)$): In this regime 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 discontinuous changes in the network topology. To be specific, a continuous increase in the node removal rate leads to a transition from a scale-free to an exponential network.

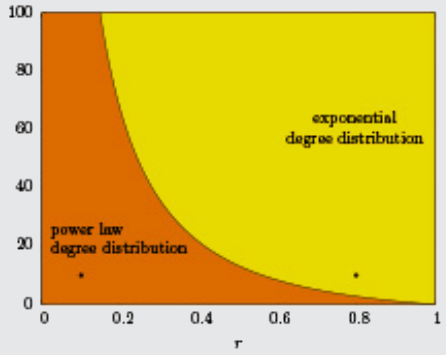

Figure 6.14
Scaling under node deletion

The degree distribution of a network whose growth is driven by preferential attachment with initial attractiveness A and node removal rate r. After [28].

SUMMARY

As we illustrated in this chapter, rather diverse processes, from node fitness to internal links or aging, can influence the topology of real networks. By exploring these processes, we came to see how to use the evolving network modeling framework to accurately predict the impact of various frequently encountered elementary events on a network's topology and evolution. The most important conclusion of the examples discussed in this chapter is that *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.*

In CHAPTER 4 we documented the difficulties we encounter when we attempt to fit a pure power law to the degree distribution of real networks. The roots of this problem were revealed in this chapter: if we account for the detailed dynamical processes that contribute to the evolution of real networks, we expect systematic deviations from a pure power law. Indeed, the previous sections predicted several analytical forms for the degree distribution:

- **Power law:** A pure power law emerges if a growing network is governed by preferential attachment only, lacking nonlinearities or initial attractiveness. In its pure form a power law is observed only in a few systems. Yet, it is the starting point for understanding the degree distribution of most real networks.

- **Stretched exponential:** If preferential attachment is sublinear, the degree distribution follows a stretched exponential SECTION 5.7. A stretched exponential degree-distribution can also appear under node removal at the critical point SECTION 6.5.

- **Fitness-induced corrections:** In the presence of fitnesses the precise form of p_k depends on the fitness distribution $\rho(\eta)$, which determines p_k via Eq. 6.6. For example, a uniform fitness distribution induces a logarithmic correction in p_k Eq. 6.8. Other forms of $\rho(\eta)$ can lead to rather exotic p_k.

- **Small-degree cutoffs:** Initial attractiveness adds a random component to preferential attachment. Consequently, the degree distribution develops a small-degree saturation.

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

In most networks several of the elementary processes discussed in this chapter appear together. For example, in scientific collaboration networks we have sublinear preferential attachment with initial attractiveness and links can be both external and internal. As researchers have different creativity, fitness also plays a role, requiring 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 $\rho(\eta)$. These findings indicate that if our goal is to obtain an accurate fit to the degree distribution, we first need to build a generative model that analytically predicts the expected functional form of p_k. Yet, in many systems developing an accurate theory for p_k may be an overkill. Hence, it is often sufficient to establish if we are dealing with a broad or a bounded degree distribution SECTION 4.9, as the system's properties will be primarily driven by this distinction.

The results of this chapter also allow us to reflect on the role of the various network models. We can categorize these models into three main classes Table 6.2:

Static Models: The random network model of Erdős and Rényi CHAPTER 3 and the small world network model of Watts and Strogatz Fig. 3.15 have a fixed number of nodes, prompting us to call them static. They both assume that the role of the network modeler is to cleverly place the links between the nodes. Both models predict a bounded degree distribution.

Generative Models: The configuration and the hidden parameter models discussed in SECTION 4.8 generate networks with some 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 aim to capture the mechanisms that govern the time evolution of a network. The most studied example in the Barabási-Albert model, but equally important are the extenions discussed in this chapter, from the Bianconi-Barabási mod-

el to models involving internal links, aging, node and link deletion, or accelerated growth. These models are motivated by the hypothesis that if we correctly capture all microscopic processes that contribute to a network's evolution, then the network's large-scale characteristics follow from that. There is an important role in network theory for each three modeling frameworks. 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 certain network property, we must resort to evolving network models, that capture the processes that have built the network in the first place.

Finally, the results of this chapter allow us to formulate our next network law:

The fifth law, the role of diversity.

With time the fittest nodes turn into the largest hubs.

A. Quantitative formulation

Eq. 6.4 offers the quantitative formulation of the fifth law, predicting that the dynamical exponent, capturing the rate at which a node acquires links, is proportional to the node's fitness. Hence the higher a node's fitness, the higher the rate it acquires links. Consequently, with time the nodes with the highest fitness will turn into the largest hubs.

B. Universality

In most networks nodes with different qualities and capabilities compete for links. Hence node fitness, capturing a node's ability to attract links, is present in most real networks.

C. Non-random origin

The dynamics of fitness-driven networks is quite different from the dynamics of the random network model, in which nodes acquire links at comparable rate. Hence, the properties of these networks cannot be explained within the random network framework.

Name	Static Models	Generative Models	Mechanistic Models
Example	Erdős-Rényi Watts-Strogatz	Configuration Model Hidden Parameter Model	Barabási-Albert Fitness Model
Characteristics	N fixed L variable p_k bounded	Pre-defined, arbitrary p_k.	p_k depends on the nature of the processes that contribute to the networks evolution.

Table 6.2
Models of network science

The table shows the three main modeling frameworks we encountered so far, together with their main distinguishing features.

HOMEWORK

1. Calculate the degree exponent and the dynamical exponent for a growing network with two distinct fitnesses. To be specific, let us assume that the fitnesses follow the double delta distribution

$$\rho(\eta) = \delta(\eta - a) + \delta(\eta - 1) \quad \text{with } 0 \le a \le 1. \tag{6.33}$$

   Discuss how the degree exponent depends on the parameter a.

2. Calculate the degree exponent of the directed Barabási-Albert model with accelerated growth, i.e. when $m(t)=t^{\odot}$.

3. Assume that a network is driven by a preferential attachment with additive fitness, $\pi(k_i) \sim A_i+k_i$, where A_i is chosen from a $\rho(A_i)$ distribution [44]. Calculate and discuss the degree distribution of the resulting network.

ADVANCED TOPICS 6.A
SOLVING THE FITNESS MODEL

The purpose of this section is to derive the degree distribution of the fitness model [2, 13, 14]. We start by calculating the mean of the sum $\sum_j \eta_j k_j$ over all possible realizations of the quenched fitnesses η. Since each node is born at a different time t_0, we can write the sum over j as an integral over t_0

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

By replacing $k_\eta(t, t_0)$ with Eq. 6.3 and performing the integral over t_0, we obtain

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

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 Eq. 6.35 the term $t^{\beta(\eta)}$ can be neglected compared to t, obtaining

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

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

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

Using Eq. 6.36, and the notation $k_\eta = k_{\eta i}(t, t_0)$, the dynamic equation Eq. 6.2 can be written as

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

which has a solution of the form Eq. 6.3, given that

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

confirming the self-consistent nature of the assumption Eq. 6.3. To complete the calculation we need to determine C from Eq. 6.37. After substituting $\beta(n)$ with η/C , we obtain

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

where η_{max} is the maximum possible fitness in the system. The integral Eq. 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

$$\sum_j \eta_j k_j \le \eta_{\max} \sum_j k_j = 2mt\eta_{\max} \tag{6.41}$$

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

If there is a single dynamic exponent β, the degree distribution should follow the power law $p_k \sim k^{-\gamma}$, where the degree exponent is given by $\gamma=1/\beta+1$. However, in the fitness model we have a spectrum of dynamic exponents $\beta(\eta)$, thus p_k is given by a weighted sum over different power-laws. To find p_k we need to calculate the cumulative probability that a randomly chosen node's degree satisfies $k_\eta(t)>k$. This cumulative probability is given by

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

Thus, the degree distribution is given by the integral

$$P_k = \int_{\eta_{max}}^{0} d\eta \frac{\partial P(k_\eta(t) > k)}{\partial t} \propto \int d\eta \rho(\eta) \frac{C}{\eta} \left(\frac{m}{k}\right)^{\frac{C}{\eta}+1}. \tag{6.43}$$

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, preprint, 2013.

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

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

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

[14] 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.

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

[16] 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.

[17] Y.-H. Eom and S. Fortunato. Characterizing and Modeling Citation Dynamics, PLoS ONE 6(9): e24926, 2011.

[18] 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.

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

[20] G. Goshal, L. Ping, and A.-L Barabási, preprint, 2013.

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

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

[23] S. Saavedra, F. Reed-Tsochas, and B. Uzzi. Asymmetric disassembly and robustness in declining networks, PNAS105:16466–16471, 2008.

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

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

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

[27] 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.

[28] 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.

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

[30] R. Sole and J. Bascompte. Self-Organization in Complex Ecosystems (Princeton Univ Press, Princeton), 2006.

[31] 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.

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

[33] 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.

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

[35] 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.

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

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

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

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

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

[41] 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.

[42] K. Klemm and V. M. Eguiluz. Highly clustered scale free networks,

Phys. Rev. E 65: 036123, 2002.

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

[44] G. Ergun and G. J. Rodgers, Growing random networks with fitness, Physica A 303: 261-272, 2002.