

CHAPTER 4

THE SCALE-FREE PROPERTY

Introduction

Power laws and scale-free networks 1

Hubs 2

The meaning of scale-free 3

Universality 4

Ultra-small property 5

The role of the degree exponent 6

Generating networks with a pre-defined degree distribution 7

Summary

ADVANCED TOPICS 4.A

Power laws

ADVANCED TOPICS 4.B

Plotting a power-law degree distribution

ADVANCED TOPICS 4.C

Estimating the degree exponent

Bibliography

Figure 4.0 (next page)

Art and Networks: Tomás Saraceno

Tomás Saraceno creates art work inspired by spider webs and neural networks. Trained as an architect, Saraceno deploys theoretical frameworks and insights from engineering, physics, chemistry, aeronautics, and materials science using networks as a source of inspiration and metaphor. The image shows his work displayed in the Miami Art Museum, an example the artist's take on a complex network.



INTRODUCTION

As difficult it is to overstate the importance of the World Wide Web in our daily life, it is equally hard to exaggerate the role the Web played in the development of network theory. It aided the discovery of a number of fundamental network properties and became a standard testbed for many network measures. As its name states, the WWW is a “web” whose nodes are documents and the links are the uniform resource locators (URLs) that allow us to move with a click from one web document to the other. With an estimated size of over one trillion documents ($N \approx 10^{12}$), the Web is the largest network humanity has ever built. It exceeds in size even the human brain ($N \approx 10^{11}$ neurons).

We can use a software called a crawler to map out the Web’s wiring diagram. A crawler can start from any web document, identifying the links (URLs) on it. Next it downloads the documents these links point to and identifies the links on these documents, and so on. This process iteratively returns a local map of the Web. Search engines like Google or Bing operate such crawlers that constantly index new documents, along the way providing a detailed map of the WWW.

The first map of the WWW obtained with the explicit goal of understanding the structure of the network behind it was generated by Hawoong Jeong at University of Notre Dame. He mapped out the nd.edu domain [1], consisting of about 300,000 documents and 1.5 million links. The purpose of the map in Fig. 4.1 was to compare the properties of the Web graph to the random network model. Indeed, in 1998 there were reasons to believe that the WWW could be well approximated by a random network. The content of each document reflects the personal and professional interests of its creator, from individuals to organizations. Given the diversity of these interests, the links on these documents might appear to point to randomly chosen documents. A quick look at the map in Fig. 4.1 supports this view: there is a high degree of randomness behind the Web’s wiring diagram. Yet, a closer inspection reveals some puzzling differences between this map and a random network. In a random network highly connected nodes, or hubs, are effectively forbidden.

In contrast in Fig. 4.1 numerous small-degree nodes coexist with a few hubs, nodes with an exceptionally large number of links. The purpose of this chapter is to show that these hubs are not unique to the Web, but we encounter them in many real networks. They represent a signature of a deeper organizing principle that we call the scale-free property.

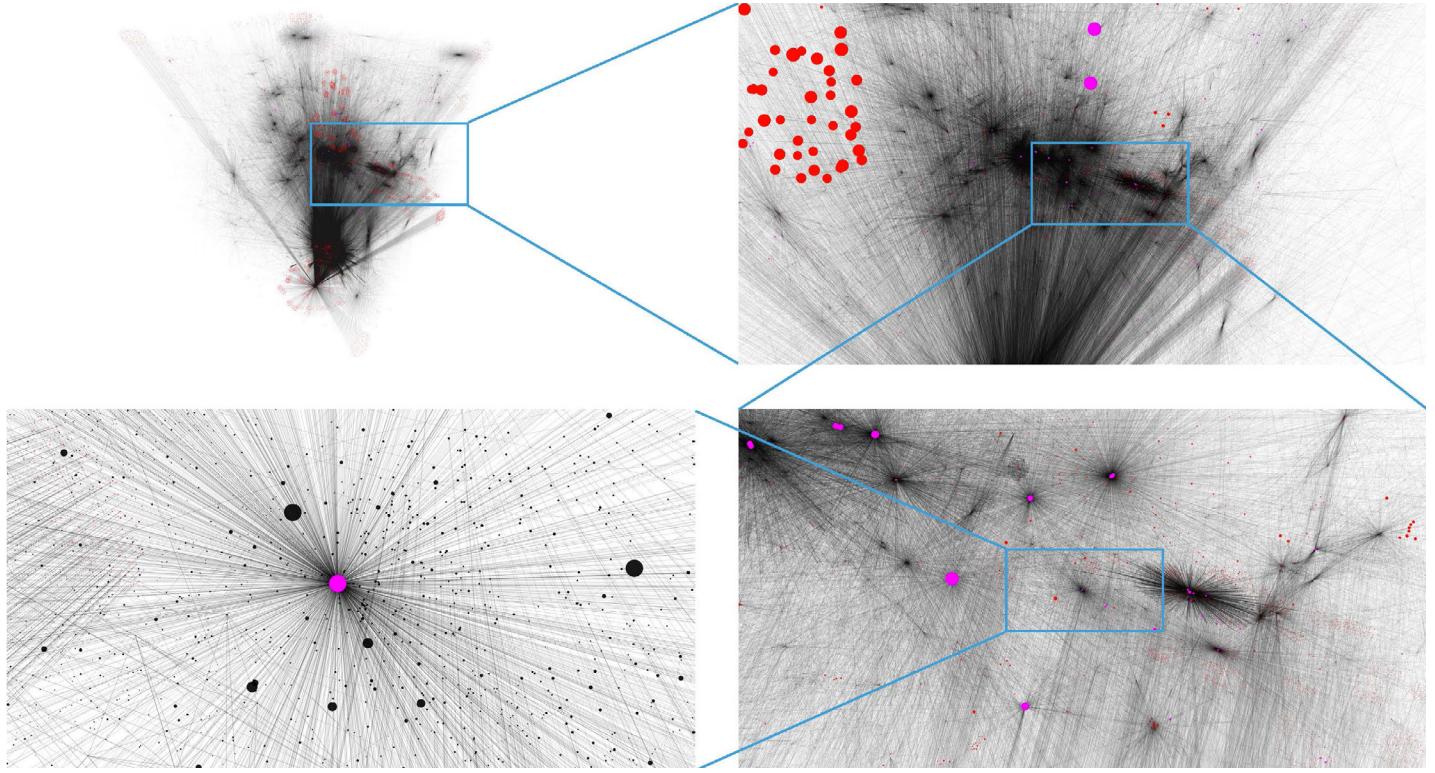


Figure 4.1
The topology of the WWW

A visualization of the web sample that led to the discovery of the scale-free property. The sequence of images shows an increasingly magnified local region of the network. The first panel displays all 325,725 nodes, offering a global view of the full dataset. Nodes with more than 50 links are shown in red and nodes with more than 500 links in purple. The increasingly magnified closeups reveal the presence of a few highly connected nodes, called hubs, that accompany scale-free networks (Image by M. Martino).

POWER LAWS AND SCALE-FREE NETWORKS

If the WWW were to be a random network, its degrees should follow a Poisson distribution. Yet, as Fig. 4.1 indicates, the Poisson form offers a poor fit for the WWW's degree distribution. Instead we find that on a log-log scale the data points form an approximate straight line, suggesting that the degree distribution of the WWW is best approximated with

$$p_k \sim k^{-\gamma}. \quad (4.1)$$

Eq. 4.1 is called a *power law* distribution and the exponent γ is its *degree exponent*. If we take a logarithm of Eq. 4.1, we obtain

$$\log p_k \sim -\gamma \log k. \quad (4.2)$$

Therefore, if Eq. 4.1 holds, $\log p_k$ is expected to depend linearly on $\log k$, the slope of this line being the degree exponent γ , as observed in Fig. 4.2.

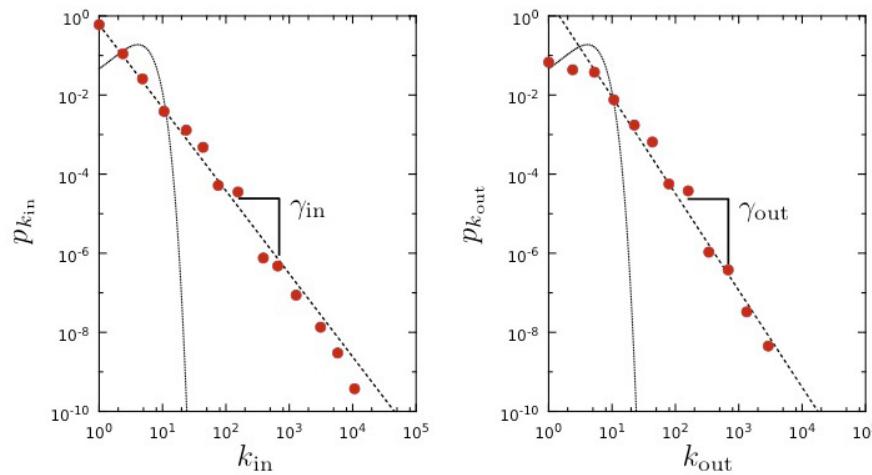


Figure 4.2
The degree distribution of the WWW

The incoming (left panel) and outgoing (right panel) degree distribution of the WWW sample mapped in the 1999 study of Albert *et al.* [1]. The degree distribution is shown on double logarithmic axis (log-log plot), in which a power law is expected to follow a straight line. The symbols correspond to the empirical data and the dotted line corresponds to the power-law fit, with degree exponents $\gamma_{in} = 2.1$ and $\gamma_{out} = 2.45$. The degree distribution predicted by a Poisson function with average degree $\langle k_{in} \rangle = \langle k_{out} \rangle = 4.60$, representing the observed values for the WWW sample, is shown as a dotted line.

As the WWW is a directed network, each document is characterized by an out-degree k_{out} , representing the number of links that point from a document to other documents, and an in-degree k_{in} , representing the number of other documents that point to a given document. We must therefore distinguish two different degree distributions: the probability that a randomly chosen document points to k_{out} other web documents, or $p_{k_{out}}$, and the probability that a randomly chosen node has k_{in} other web documents pointing to it, or $p_{k_{in}}$. In the case of the WWW both $p_{k_{in}}$ and $p_{k_{out}}$ can be approximated by a power law

$$p_{k_{in}} \propto k^{-\gamma_{in}} \quad (4.3)$$

$$p_{k_{out}} \propto k^{-\gamma_{out}} \quad (4.4)$$

where γ_{in} and γ_{out} are the degree exponents for the in- and out-degrees, respectively Fig. 4.2. In general γ_{in} can differ from γ_{out} . For example, for the WWW sample of Fig. 4.1 we have $\gamma_{in} \approx 2.1$ and $\gamma_{out} \approx 2.45$. The empirical evidence discussed above leads to the concept of a scale-free network [2]: *Networks whose degree distribution follows a power law are called scale-free networks*. As Fig. 4.2 indicates, for the WWW the power law persists for almost four orders of magnitude, prompting us to call the network behind the Web scale-free. In this case the scale-free property applies to both in and out-degrees. To explore the consequences of the scale-free property, we have to define the power-law distribution in more precise terms. For this we introduce the discrete and the continuum formalisms used throughout this book.

DISCRETE FORMALISM

As node degrees are always positive integers, $k = 0, 1, 2, 3, \dots, N$, the discrete formalism captures the probability p_k that a node has exactly k links

$$p_k = Ck^{-\gamma}. \quad (4.5)$$

The constant C is determined by the normalization condition

$$\sum_{k=1}^{\infty} p_k = 1. \quad (4.6)$$

Using Eq. 4.4 we obtain, $C \sum_{k=1}^{\infty} k^{-\gamma} = 1$, hence

$$C = \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}} = \frac{1}{\zeta(\gamma)}, \quad (4.7)$$

where $\zeta(\gamma)$ is the Riemann-zeta function. Thus for $k > 0$ the discrete power-law distribution has the form

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}. \quad (4.8)$$

Note that Eq. 4.8 diverges at $k=0$. We therefore need to separately specify p_0 , representing the fraction of nodes that have no links to other nodes (isolated nodes).

CONTINUUM FORMALISM

In analytical calculations it is often convenient to assume that the degrees can take up any positive real value. In this case the power-law degree distribution is written as:

$$p(k) = Ck^{-\gamma}. \quad (4.9)$$

Using the normalization condition:

$$\int_{k_{min}}^{\infty} p(k)dk = 1 \quad (4.10)$$

we obtain the constant:

$$C = \frac{1}{\int_{K_{min}}^{\infty} k^{-\gamma} dk} = (\gamma - 1)K_{min}^{\gamma-1}. \quad (4.11)$$

Therefore in the continuum formalism the degree distribution has the form:

$$p(k) = (\gamma - 1)k_{min}^{\gamma-1} k^{-\gamma}. \quad (4.12)$$

Here k_{min} is the smallest degree for which the power law Eq. 4.8 holds. Note that p_k encountered in the discrete formalism has a precise meaning: it provides the probability that a randomly selected node has degree k . In contrast, only the integral of $p(k)$ encountered in the continuum formalism has a physical interpretation:

$$\int_{k_1}^{k_2} p(k)dk \quad (4.13)$$

provides the probability that a randomly chosen node has degree between k_1 and k_2 . In summary, networks whose degree distribution follows a power law are called scale-free networks. If a network is directed, the scale-free property can apply separately to the in- and the out-degrees.

To mathematically study the properties of scale-free networks, we can use the discrete or the continuum formalism. Note, however, that the scale-free property is independent of the formalism we use to describe the degree distribution.

BOX 4.1

The 80/20 rule and the top one percent

Vilfredo Pareto, a 19th century economist, noticed that in Italy a few wealthy individuals earned most of the money, while the majority of the population earned rather small amounts. He connected this disparity to the observation that incomes follow a power law, representing the first known report of a power-law distribution [3]. His finding entered the popular literature as the 80/20 rule: roughly 80 percent of money is earned by only 20 percent of the population.

The 80/20 emerges in many areas, like management, stating that 80 percent of profits are produced by only 20 percent of the employees or that 80 percent of decisions are made during 20 percent of meeting time.

They are present in networks as well: 80 percent of links on the Web point to only 15 percent of webpages; 80 percent of citations go to only 38 percent of scientists; 80 percent of links in Hollywood are connected to 30 percent of actors [4]. Typically all quantities obeying the 80/20 rule follow a power law distribution.

During the 2009 economic crisis power laws have gained a new meaning: the Occupy Wall Street Movement highlighted the fact that in the US 1% of the population earns a disproportionate 15% of the total US income. This 1% effect, a signature of a profound income disparity, is again a natural consequence of the power-law nature of the income distribution.

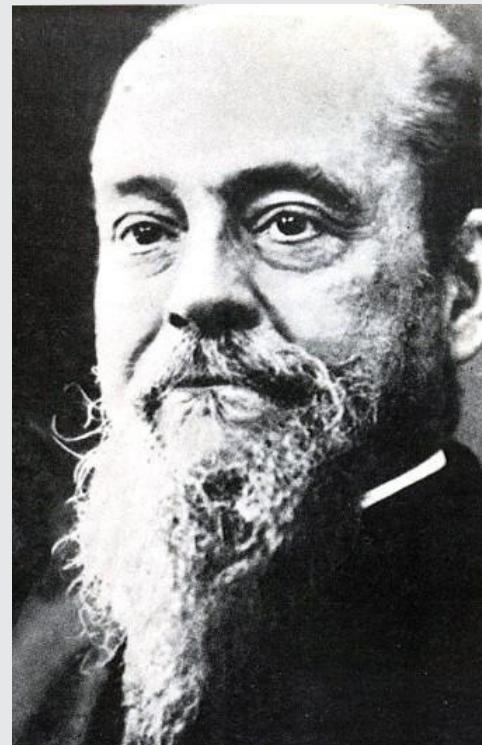


Figure 4.3
Vilfredo Federico Damaso Pareto (1848 – 1923)

Italian economist, political scientist, and philosopher, who had important contributions to our understanding of income distribution and to the analysis of individuals choices. A number of fundamental principles are named after him, like Pareto efficiency, Pareto distribution (another name for a power-law distribution), the *Pareto principle* (or 80/20 law).

HUBS

The main difference between a random and a scale-free network comes in the tail of the degree distribution, representing the high- k region of p_k .

Fig. 4.4 compares a power law with a Poisson function, indicating that:

- For small k the power law is above the Poisson function, hence a scale-free network has a large number of small degree nodes that are virtually absent in a random network.
- For k the vicinity of $\langle k \rangle$ the Poisson distribution is above the power law, indicating that in a random network most nodes have degree $k \approx \langle k \rangle$.
- For large k the power law is again above the Poisson curve. The difference is particularly visible if we show p_k on a log-log plot **Fig. 4.4b**, indicating that the probability of observing a high-degree node, or hub, is several orders of magnitudes higher in a scale-free than in a random network.

Let us use the WWW to illustrate the properties of the high- k regime. The probability to have a node with $k \approx 100$ is about $p_{100} \approx 10^{-30}$ in a Poisson distribution while it is about $p_{100} \approx 10^{-4}$ if p_k follows a power law. Consequently, if the WWW were to be a random network with

$$N_{k>100} = \sum_{k=101}^{\infty} \frac{(4.6)^k}{k!} e^{-4.6} \approx 10^{18}, \langle k \rangle \approx 4.6 \quad (4.14)$$

and $N \approx 10^{12}$ **Table 4.1**, we would expect nodes with more than 100 links, or effectively none. In contrast, given the WWW's power law degree distribution, with $\gamma_{in} = 2.1$, we have $N_{k>100} = 10^9$ nodes with degree $k > 100$.

HUBS

All real networks are finite. The size of the WWW is estimated to be $N \approx 10^{12}$ nodes; the size of the social network is the Earth's population, about $N \approx 7 \times 10^9$. These numbers are huge, but finite. Other networks pale in comparison: the genetic network in a human cell has approximately 20,000 genes while the metabolic network of the *E. Coli* bacteria has only about a

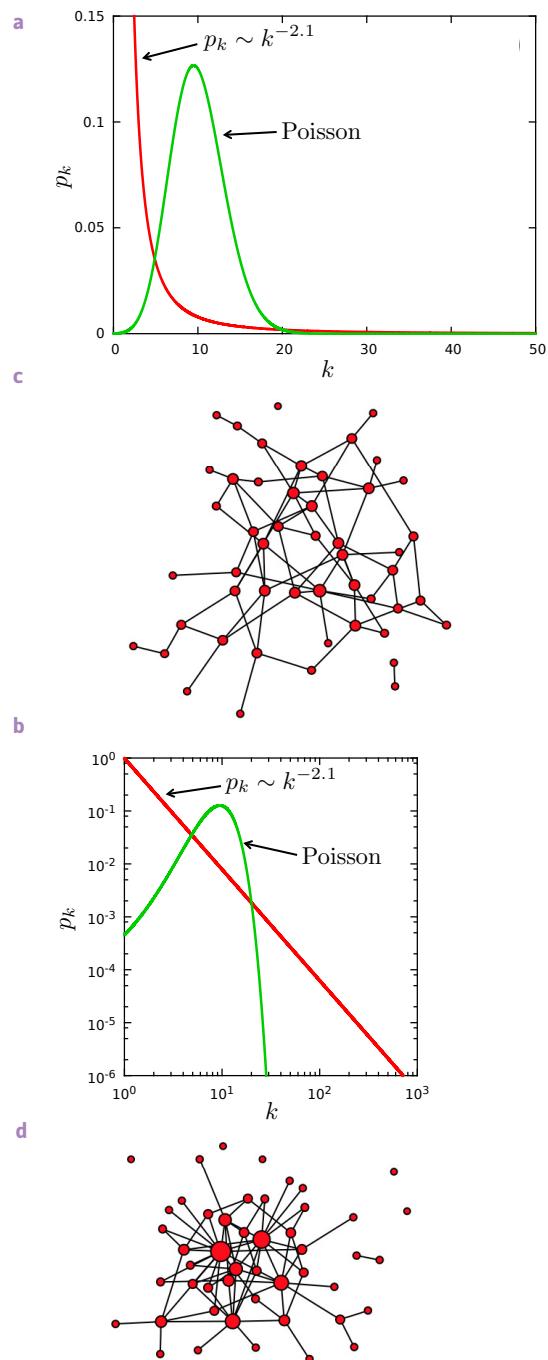


Figure 4.4
Poisson vs. power-law distributions

(a) A Poisson function and a power-law function with $\gamma = 2.1$. Both distributions have $\langle k \rangle = 10$.

(b) The curves in (a) shown on a log-log plot, offering a better view of the difference between the two functions in the high- k regime.

(c) A random network with $\langle k \rangle = 3$ and $N = 50$, illustrating that most nodes have comparable degree $k \approx \langle k \rangle$.

(d) A scale-free network with $\langle k \rangle = 3$, illustrating that numerous small-degree nodes coexist with a few highly connected hubs.

thousand metabolites. This prompts us to ask: how does the network size affect the size of its hubs?

For an arbitrary degree distribution p_k we can calculate the expected maximum degree, k_{max} , often called natural cutoff. It represents the expected size of the largest hub.

It is instructive to perform the calculation first for the exponential distribution $p_k = Ce^{-\lambda k}$. Assuming that the network's minimum degree is k , the normalization condition

$$\int_{k_{min}}^{\infty} p(k)dk = 1 \quad (4.15)$$

provides $C = \lambda e^{\lambda k_{min}}$. To calculate k_{max} we assume that in a network of N nodes we expect at most one node in the (k_{max}, ∞) regime. In other words the probability to observe a node whose degree exceeds k_{max} is $1/N$:

$$\int_{k_{max}}^{\infty} p(k)dk = \frac{1}{N}. \quad (4.16)$$

Equation Eq. 4.14 yields

$$k_{max} = k_{min} + \frac{\ln N}{\lambda}. \quad (4.17)$$

As $\ln N$ is a slow function of the system size, Eq. 4.17 tells us that the maximum degree will not be very different from k_{min} . For a Poisson degree distribution the calculation is a bit more involved, but the obtained dependence of k_{max} on N is even slower than the logarithmic dependence predicted by Eq. 4.17.

For a scale-free network, according to Eq. 4.16 and Eq. 4.17 the natural cut-off follows

$$k_{max} \sim k_{min} N^{\frac{1}{\gamma-1}}. \quad (4.18)$$

Hence the larger a network, the larger is the degree of its biggest hub. The polynomial dependence of k_{max} on N implies that in a large scale-free network there can be orders of magnitude differences in size between the smallest node, k_{min} , and the biggest hub, k_{max} Fig. 4.5.

To illustrate the difference in the maximum degree of an exponential and a scale-free network let us return to the WWW sample of Fig. 4.1 consisting of $N \approx 3 \times 10^5$ nodes. As $k_{min} = 1$, if the degree distribution were to follow an exponential, Eq. 4.17 predicts that the maximum degree should be $k_{max} \approx 13$. In a scale-free network of similar size and $\gamma=2.1$, Eq. 4.18 predicts $k_{max} \approx 85,000$, a remarkable difference. Note that the largest in-degree of this WWW map of Fig. 4.1 is 10,721, which is comparable to the predicted k_{max} .

This reinforces our conclusion that in a random network hubs are forbidden, while in scale-free networks they occur naturally.

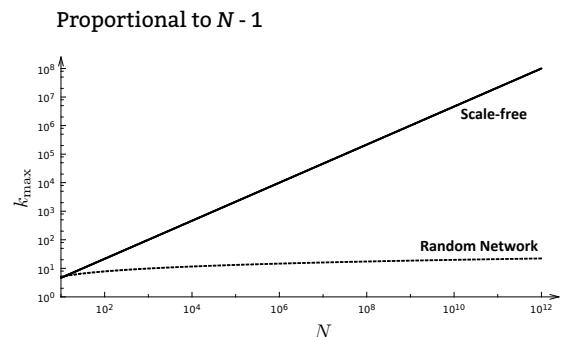


Figure 4.5
Hubs are large in scale-free networks

The expected degree of the largest node (natural cutoff) in scale-free and random networks with the same average degree $\langle k \rangle = 3$. For the scale-free network we chose $\gamma = 2.5$. For comparison, we also show the linear behavior, $k_{max} \sim N - 1$, expected for a complete network. Overall, hubs in a scale-free network are several orders of magnitude larger than the biggest node in a random network with the same N and $\langle k \rangle$.

In summary the key difference between a random and a scale-free network comes in the different shape of the Poisson and of the power-law function: in a random network most nodes have comparable degrees and hence hubs are forbidden. Hubs are not only tolerated, but are expected in scale-free networks Fig. 4.5.

The more nodes a scale-free network has, the larger are its hubs. The hubs grow polynomially with the network size, hence their size can be considerable in large networks. In contrast in a random network the size of the largest node grows logarithmically or slower with N , implying that hubs will be tiny even in a very large network.

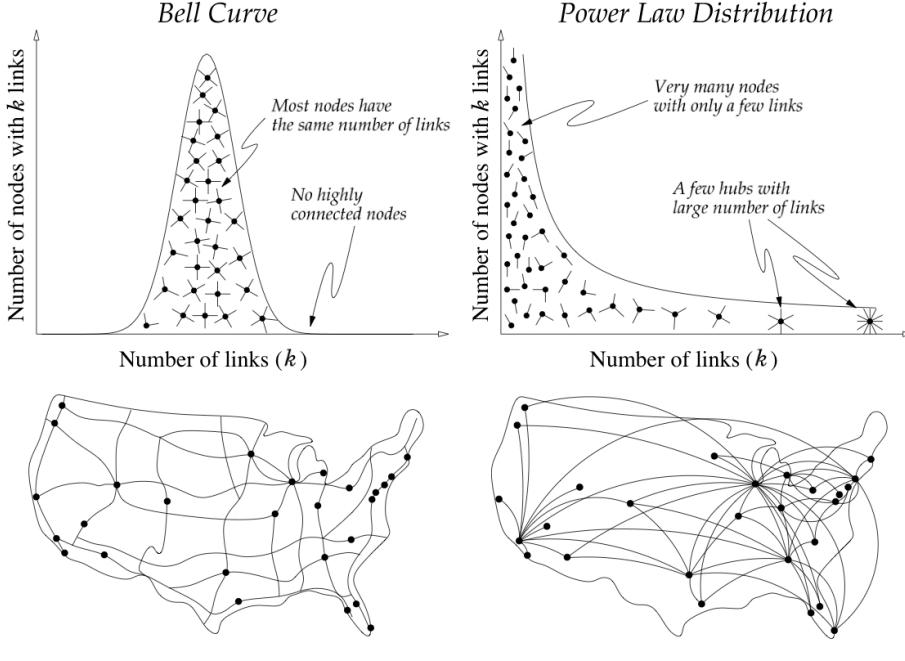


Figure 4.6
Random versus scale-free networks

Left column: the degrees of a random network follow a Poisson distribution, which is rather similar to the Bell curve shown in the figure. This indicates that most nodes have comparable degree. Hence nodes with a large number of links are absent (top panel). Consequently a random network looks a bit like a national highway network in which nodes are cities and links are the major highways connecting them (bottom panel). Indeed, there are no major cities with hundreds of highways and no city is disconnected from the highway system.

Right column: In a network with a power-law degree distribution most nodes have only a few links. These numerous small nodes are held together by a few highly connected hubs (top panel). Consequently a scale-free network looks a bit like the air-traffic network, whose nodes are airports and links are direct flights between them. Most airports are tiny, with only a few flights linking them to other airports. Yet, we can also have few very large airports, like Chicago or Atlanta, that hold hundreds of airports together, acting as major hubs (bottom panel).

Once hubs are present, they change the way we navigate the network. For example, if we travel from Boston to Los Angeles by car, we must drive through many cities (nodes). On the airplane network, however, we can reach most destinations via a single hub, like Chicago.

After [4].

THE MEANING OF SCALE-FREE

What is behind the “scale-free” name? The term is rooted in a branch of statistical physics called the theory of phase transitions SECTION 3.F, that extensively explored power laws in the 1960s and 1970s. To best understand the meaning of the scale-free term, we need to familiarize ourselves with the moments of the degree distribution. The n^{th} moment of the degree distribution is defined as:

$$k^n = \sum_{k_{\min}}^{\infty} k^n p_k = \int_{k_{\min}}^{\infty} k^n p(k) dk. \quad (4.19)$$

The lower moments have important interpretation:

- $n=1$: the first moment is the average degree, $\langle k \rangle$.
- $n=2$: the second moment, $\langle k^2 \rangle$, provides the variance $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$, measuring the spread in the degrees. Its square root, σ , is the standard deviation.
- $n=3$: the third moment, $\langle k^3 \rangle$, determines the skewness of a distribution, telling us how symmetric is p_k around the average $\langle k \rangle$. Symmetric distributions have zero skewness. For a scale-free network the n^{th} moment of the degree distribution is

$$k^n = \int_{k_{\min}}^{k_{\max}} k^n p(k) dk = C \frac{k_{\max}^{n-\gamma+1} - k_{\min}^{n-\gamma+1}}{n - \gamma + 1}. \quad (4.20)$$

While typically k_{\max} is fixed, the degree of the largest hub, k_{\max} , increases with the system size, following Eq. 4.18.

Hence to understand the behavior of $\langle k^n \rangle$ we need to take the asymptotic limit $k_{\max} \rightarrow \infty$ in Eq. 4.20, probing the properties of very large networks. In this limit Eq. 4.20 predicts that the value of $\langle k^n \rangle$ depends on the interplay between n and γ :

- If $n - \gamma + 1 \leq 0$ then the first term on the r.h.s. of Eq. 4.20, $k_{\max}^{n-\gamma+1}$, goes to zero as k_{\max} increases. Therefore all moments that satisfy $n \leq \gamma - 1$ will be finite.

- If $n-\gamma+1 \geq 0$ then $\langle k_n \rangle$ goes to infinity as $k_{max} \rightarrow \infty$. Therefore all moments satisfying $n \geq \gamma-1$ diverge.

For most real scale-free networks the degree exponent γ is between 2 and 3 [Table 4.1](#). Hence for these in the $N \rightarrow \infty$ limit the first moment $\langle k \rangle$ is finite, but the second and higher moments, $\langle k^2 \rangle, \langle k^3 \rangle$, go to infinity. This divergence helps us understand the origin of the “scale-free” term:

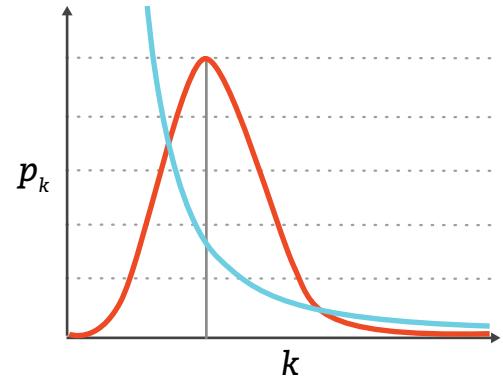
- If the degrees follow a normal distribution, then the degree of a randomly chosen node is

$$k = \langle k \rangle \pm \sigma_k \quad (4.21)$$

For a random network with a Poisson degree distribution $\sigma_k = \sqrt{\langle k \rangle}$, which is always smaller than $\langle k \rangle$. Hence the degrees are in the range $k = \langle k \rangle \pm \langle k \rangle^{1/2}$, indicating that nodes in a random network have comparable degrees. Therefore the average degree $\langle k \rangle$ serves as the “scale” of a random network.

- For a network with a power-law degree distribution and $\gamma < 3$ the first moment is finite but the second moment is infinite. The divergence of $\langle k^2 \rangle$, and hence of σ_k for large N indicates that the fluctuations around the average could be arbitrary large. That is, when we randomly choose a node, we do not know what to expect, as the chosen node’s degree could be tiny or arbitrarily large. Hence networks with $\gamma < 3$ do not have a meaningful internal scale. They are “scale-free” [Fig. 4.7](#). For example the average degree of the WWW sample is $\langle k \rangle = 4.60$ [Table 4.1](#). Given that $\gamma \approx 2.1$, the second moment diverges, which means that our expectation for the in-degree of a randomly chosen WWW document is $\langle k \rangle = 4.60 \pm \infty$ in the $N \rightarrow \infty$ limit. That is, a randomly chosen webpage could easily yield a document of degree one or two, as 74.02% of nodes have in-degree less than $\langle k \rangle$. Yet, it could also yield a node with hundreds of millions of links, like google.com or facebook.com.

Strictly speaking $\langle k^2 \rangle$ diverges only in the $N \rightarrow \infty$ limit. Yet, the divergence is relevant for finite networks as well. To illustrate this, [Table 4.1](#) and [Figure 4.8](#) show the standard deviation $\sigma_k = \sqrt{\langle k^2 \rangle - \langle k \rangle^2}$ for ten real networks. For most of these networks σ_k is significantly larger than $\langle k \rangle$, documenting large variations in node degrees. For example, the degree of a randomly chosen node in the studied WWW sample is $k_{in} = 4.60 \pm 39.05$, indicating once again that the average is not informative in this case. In summary, the scale-free name captures the lack of an internal scale, a consequence of the fact that nodes with widely different degrees coexist. This feature distinguishes scale-free networks from lattices, in which all nodes have exactly the same degree ($\sigma_k = 0$), or from random networks, whose degrees vary in a narrow range ($\sigma_k = \langle k \rangle^{1/2}$). As we will see in the coming chapters, this divergence is the origin of some of the most interesting properties of scale-free networks, from their robustness to random failures to the anomalous spread of viruses.



Random network

Randomly chosen node: $k = \langle k \rangle \pm \langle k \rangle^{1/2}$
Scale: $\langle k \rangle$

Scale-free network

Randomly chosen node: $k = \langle k \rangle \pm \infty$
 $\langle k \rangle$ is meaningless as ‘scale’

Figure 4.7
Scale-free networks lack an internal scale

For any bounded distribution (e.g. a Poisson or a Gaussian distribution) the degree of a randomly chosen node will be in the vicinity of $\langle k \rangle$. Hence $\langle k \rangle$ serves as the network’s scale. In a scale-free network the second moment diverges, hence the degree of a randomly chosen node can be arbitrarily different from $\langle k \rangle$. As a scale-free network lacks an intrinsic scale, is it scale-free.

NETWORK	NL		$\langle k \rangle$ $\langle k_{in} \rangle = \langle k_{out} \rangle$	σ_{in}	σ_{out}	σ	γ_{in}	γ_{out}	γ
Internet	192,244	609,066	6.34	-	-	14.14	-	-	3.42*
WWW	325,729	1,497,134	4.60	39.05	21.48	-	2.31	2.00	-
Power Grid	4,941	6,594	2.67	-	-	1.79	-	-	Exp.
Mobile Phone Calls	36,595	91,826	2.51	2.39	2.32	-	4.69*	5.01*	-
Email	57,194	103,731	1.81	9.56	34.07	-	3.43*	2.03	-
Science Collaboration	23,133	93,439	8.08	-	-	10.63	-	-	3.35
Actor Network	702,388	29,397,908	83.71	-	-	200.86	-	-	2.12
Citation Network	449,673	4,689,479	10.43	29.37	9.49	-	3.03**	4.00	-
E. Coli Metabolism	1,039	5,802	5.58	22.46	19.12	-	2.43	2.90	-
Yeast Protein Interactions	2,018	2,930	2.90	-	-	4.88	-	-	2.89*

Table 4.1

The characteristics of several real network

The table shows the standard deviation of the degree distribution $\sigma = \sqrt{\langle k^2 \rangle - \langle k \rangle^2}$ (σ_{in} and σ_{out} for directed networks) for our ten reference networks. It indicates that for most networks σ is much larger than $\langle k \rangle$, consequence of their scale-free nature. It also lists the estimated degree exponent, γ , for each network, determined using the procedure discussed in ADVANCED TOPICS 4.A. The stars next to the reported values indicate the statistical confidence for a particular fit to the degree distribution. That is, * means that the fit shows statistical confidence for a power-law $k^{-\gamma}$ fit; while ** marks datasets that display statistical confidence for a $\sigma_k = \sqrt{\langle k^2 \rangle - \langle k \rangle^2}$ fit. Those with no stars do not show statistical confidence for any of the two forms; the reasons for this are discussed later in the next chapter and in ADVANCED TOPICS 4.C. Note that the power grid is not considered scale-free. For this network a degree distribution of the form $e^{-\lambda k}$ offers a statically significant fit.

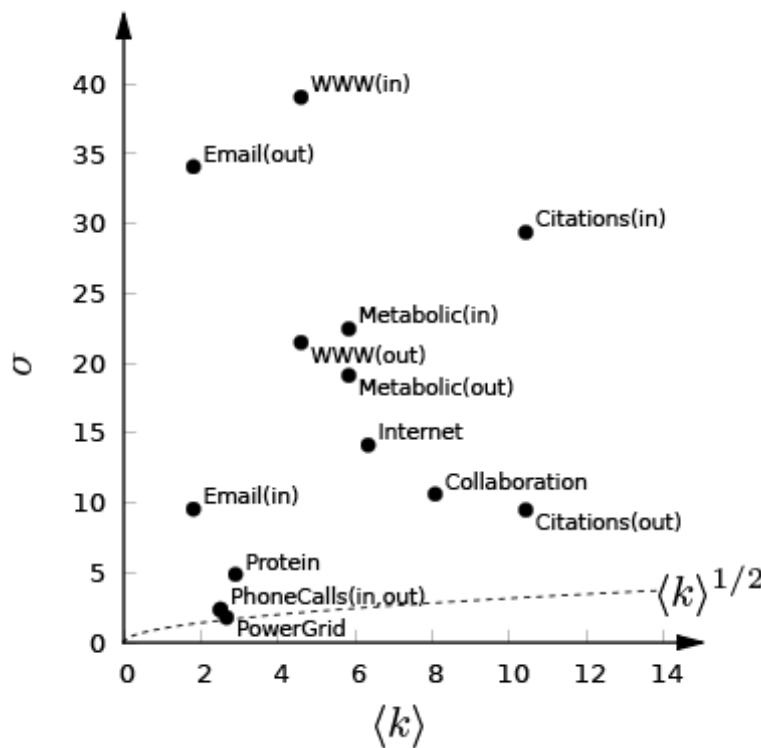


Figure 4.8
Standard deviation is large in real networks

For a random network the standard deviation follows $\sigma_k = \sqrt{\langle k \rangle}$, shown as a dashed line on the figure. The symbols show σ for ten reference networks Table 4.1, indicating that for each σ is larger than expected for a random network with similar $\langle k \rangle$. The only exception is the power grid, which is not scale-free. While the phone call network is scale-free, it has a large γ , hence it behaves like a random network.

UNIVERSALITY

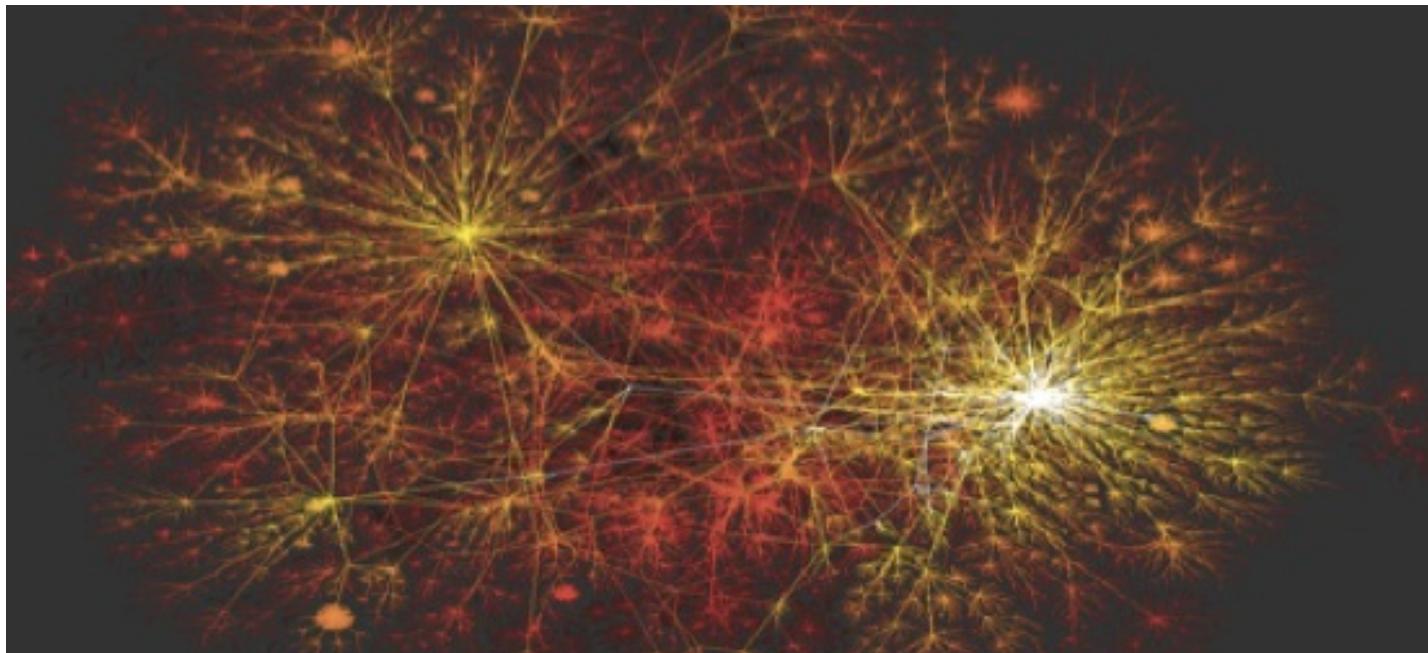


Figure 4.9
The topology of the Internet

While the terms ‘WWW’ and ‘Internet’ are often used interchangeably in the popular press, they refer to rather different systems. The WWW is an information network, with Web documents as nodes and URLs as links. In contrast the Internet is an infrastructural network, whose nodes are routers and links correspond to physical connections, like copper or optical cables.

This difference has important consequences: while the cost of linking to a web document residing on the same computer or on a different continent is the same, establishing a direct Internet link between routers in Boston and Budapest would require us to lay a new cable between the two continents, which would be prohibitively expensive. Despite these differences, the degree distribution of both networks is well approximated by a

An iconic representation of the Internet topology at the beginning of the 21st century. The image was produced by CAIDA, an organization based at University of California in San Diego, devoted to collect, analyze, and visualize Internet data. The map offers a visual demonstration of the Internet’s scale-free nature: a few highly connected hubs hold together numerous small nodes.

power law [1, 5, 6]. We have discussed the scale-free property of the WWW in the previous sections. The signatures of the Internet's scale-free nature are visible in Fig. 4.9, showing that a few high-degree routers hold together a large number of routers with only a few links.

In the past decade many real networks of major scientific, technological and societal importance were found to display the scale-free property. This is illustrated in Fig. 4.10, where we show the degree distribution of an infrastructural network (Internet), a biological network (protein-protein interactions) and a professional affiliation network (Hollywood actors). For each network the degree distribution significantly deviates from a Poisson distribution, being better approximated with a power law.

The diversity of the systems that share the scale-free property is remarkable. Indeed, the WWW is a man-made network with a history of little more than two decades, while the protein interaction network is the product of four billion years of evolution. In some of these networks the nodes are molecules, in others they are computers. It is this diversity that prompts us to call the scale-free property a universal network characteristics.

From the perspective of a researcher, a crucial question is the following: how do we establish the scale-free nature of a network? One one end, a quick look at the degree distribution will immediately reveal whether the network could be scale-free: in scale-free networks we observe orders of magnitude differences between the degrees of the smallest and the largest nodes. In contrast most nodes have comparable degrees in a random network. Yet, as the value of the degree exponent plays an important role in predicting various network properties, we need tools to fit the p_k distribution and to estimate γ . This prompts us to address several issues:

PLOTTING THE DEGREE DISTRIBUTION

The degree distributions shown in this chapter are all plotted on a double logarithmic scale, often called a log-log plot. The main reason is that when nodes with widely different degrees coexist, a linear plot is unable to display them all. We also use logarithmic binning to obtain the clean-looking degree distributions shown throughout this book, ensuring that each datapoint has proper statistical significance. The practical tips for plotting a network's degree distribution are discussed in ADVANCED TOPICS 4.B.

MEASURING THE DEGREE EXPONENT

A quick estimate of the degree exponent is often obtained by fitting a straight line to p_k on a log-log plot. Yet, this approach can be affected by systematic biases, resulting in an incorrect γ . The statistical tools available to estimate γ are discussed in ADVANCED TOPICS 4.C. We used these tools to determine the degree exponents listed in Table 4.1.

THE SHAPE OF p_k FOR REAL NETWORKS

Most degree distributions observed in real networks display clear devi-

ations from a pure power law. These can be attributed to data incompleteness or data collection biases, but the deviations also carry important information about processes that contribute to the emergence of a particular network. In **ADVANCED TOPICS 4.B** we discuss some of these deviations, and in **CHAPTER 6** we explore their origins.

Since the discovery of the scale-free nature of the WWW, an amazing number of real networks of major scientific and technological interest have been found to be scale-free [Fig. 4.10](#) from biological to social and even linguistic networks. This does not mean that all networks are scale-free. Indeed, many important networks, from the power grid to networks observed in materials science [BOX 4.2](#) do not display the scale-free property.

Yet, the prevalence of the scale-free property have prompted the research community to devote special attention to this class of networks. Uncovering the reasons why some networks are scale-free while others are not, and understanding the consequences of the scale-free property, help us better understand real networks.

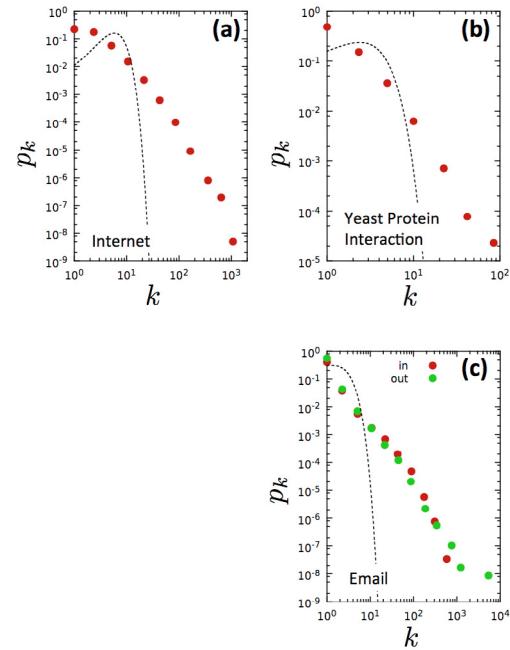


Figure 4.9b
Many real networks are scale-free

The degree distribution of three of the networks listed in [Table 4.1](#).

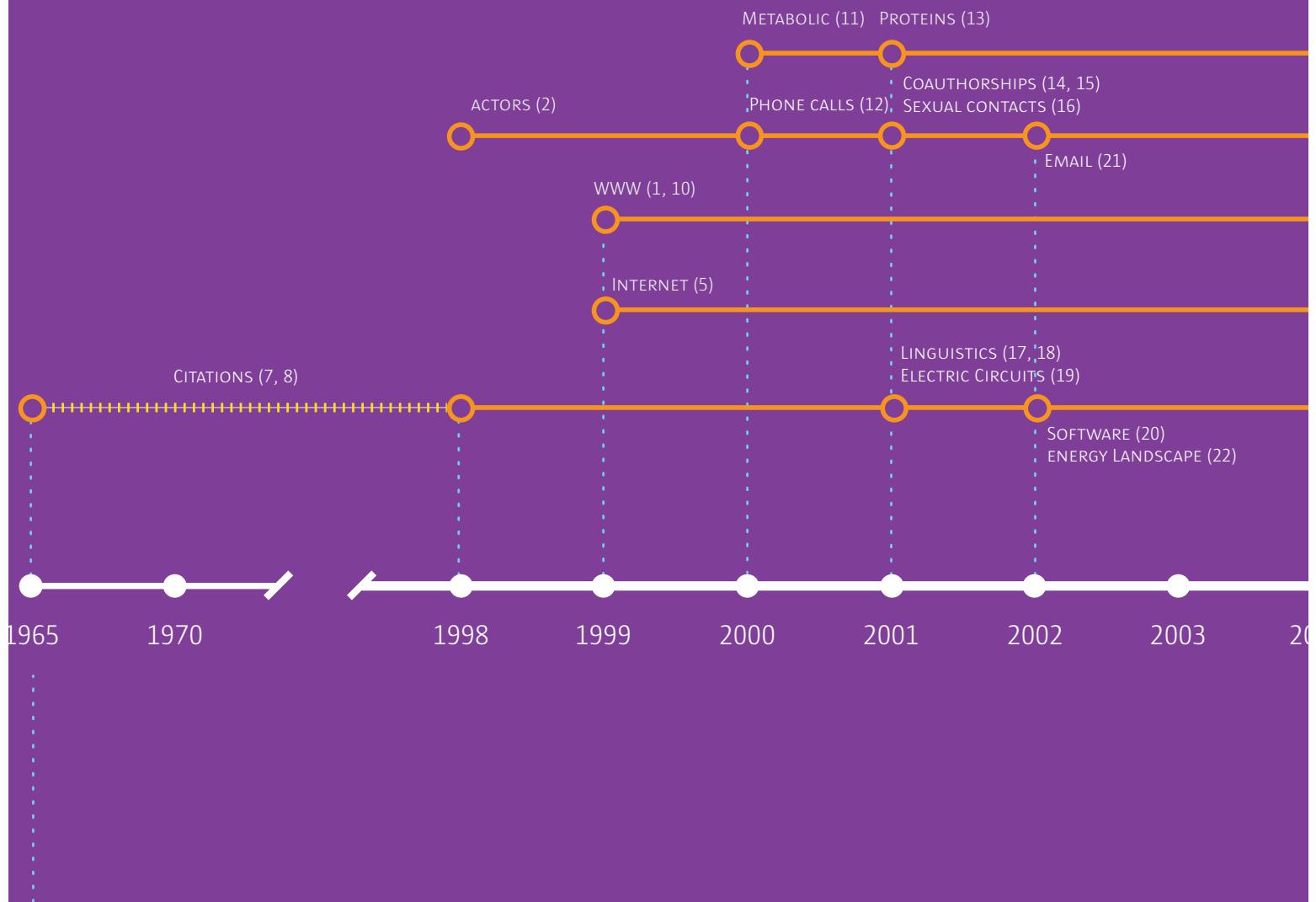
- (a) The degree distribution of the Internet at the router level.
- (b) The degree distribution of the protein-protein interaction network of yeast.
- (c) The degree distribution of the email network of a European university.

In each panel, the dotted line shows the Poisson distribution with the same $\langle k \rangle$ as the real network, indicating that the random network model cannot account for the observed p_k .

SCALE-FREE HISTORY

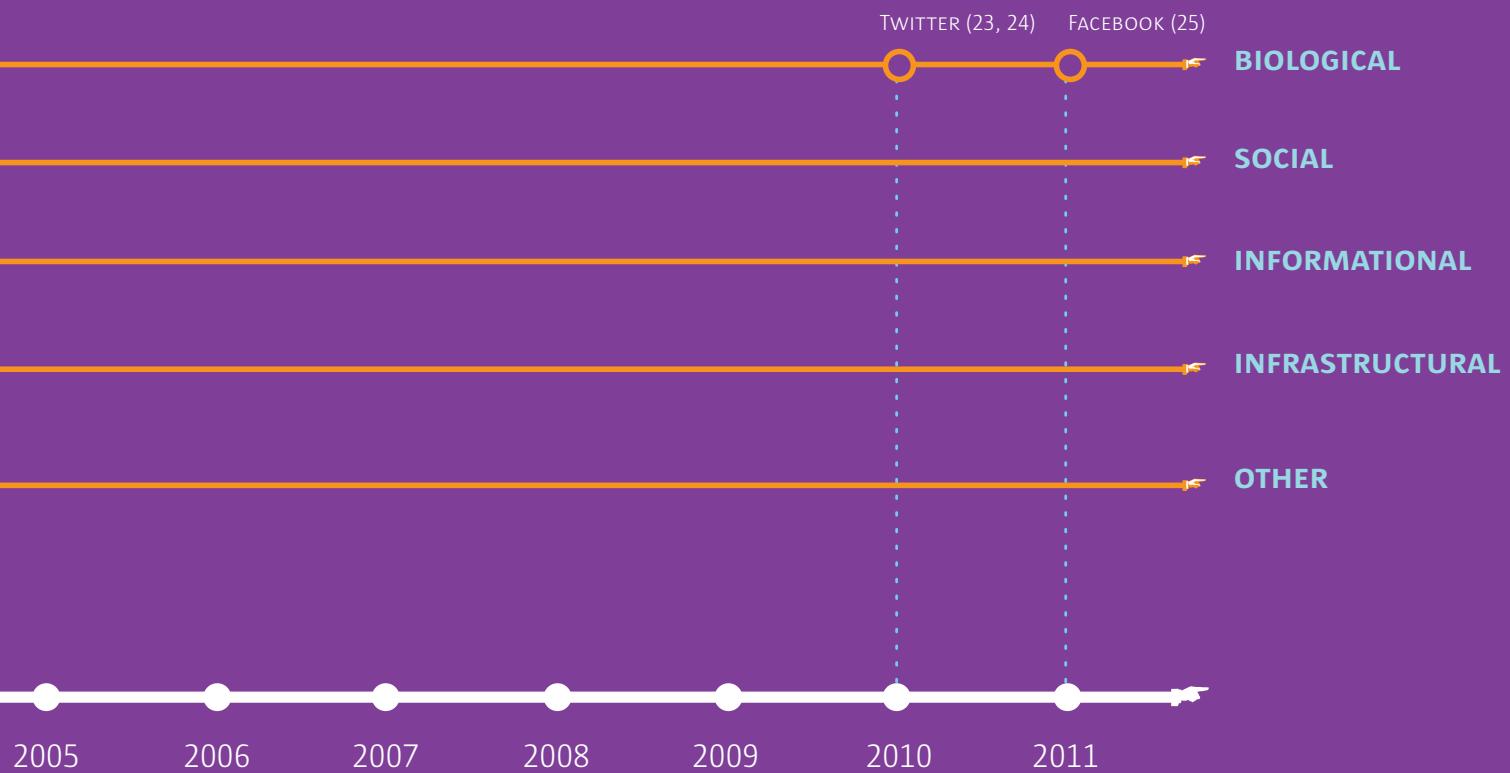
THE TIMELINE OF THE DISCOVERIES REPORTING
THE SCALE-FREE NATURE OF VARIOUS REAL NETWORKS

FIG. 4.10



Many biological, social, and technological networks display the scale-free property. The figure shows the timeline of the discoveries reporting the scale-free nature of various real networks. While there is a clear burst of reports following the 1999 discovery of scale-free networks, in hindsight it is clear that several early papers have reported characteristics that are consistent with what we call today a scale-free topology. For example, Etel de Solla Price reported in 1965 that citations to scientific papers follow a power-law distribution [7], a property independently discovered by Redner in 1998 [8]. This is a consequence of the scale-free nature of citation networks.

A common feature of these early works is that they viewed the observed quantities as scalar events, not as a manifestation of some network phenomena. It wasn't until the 1999 that it was understood that power laws are also a fundamental network property. Indeed, Barabási and Albert, in their 1999 Science paper argued that "we expect that the scale-invariant state observed in all systems for which detailed data has been available to us is a generic property of many complex networks, with applicability reaching far beyond the quoted examples." The 'scale-free network' term was also first used in 1999 [2, 9].



BOX 4.2

Not all network are scale-free

The ubiquity of the scale-free property does not mean that all real networks are scale-free. Indeed, several important networks do not share this property:

- Networks appearing in material science, like the network describing the bonds between the atoms in crystalline or amorphous materials, where each node has exactly the same degree.
- The neural network of the *C. elegans* worm.
- The power grid, consisting of generators and switches connected by transmission lines.

For the scale-free property to emerge the nodes need to have the capacity to link to an arbitrary number of other nodes. These links do not need to be simultaneous: we do not constantly chat with each of our acquaintances and a protein in the cell does not simultaneously bind to each of its potential interaction partners. In general the scale-free property is absent in systems that have a limitation in the number of links a node can have, as such limitations limit the size of the hubs. As illustrated in the image, such limitations are common in materials, explaining why they cannot develop a scale-free topology.

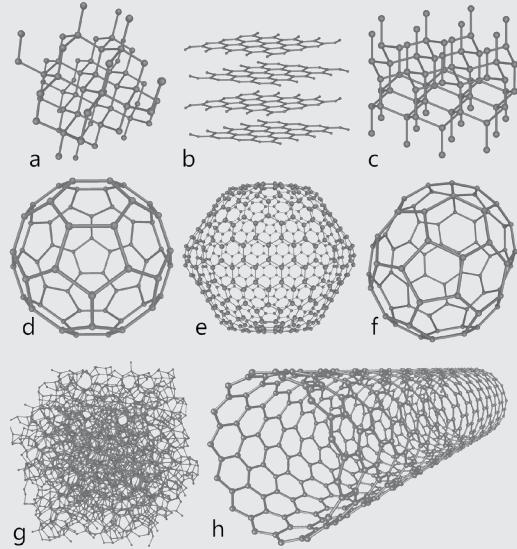


Figure 4.11
The material network

A carbon atom can share only four electrons with other atoms, hence no matter how we arrange these atoms relative to each other, in the resulting network a node can never have more than four links. Hence, hubs are forbidden and the scale-free property cannot emerge. The figure shows several carbon allotropes, each characterized by a different "network", resulting in materials with different physical characteristics, like (a) diamond; (b) graphite; (c) lonsdaleite; (d) C₆₀ (buckminsterfullerene); (e) C₅₄₀ (a fullerene) (f) C₇₀ (another fullerene); (g) amorphous carbon; (h) single-walled carbon nanotube.

Source: http://www.thenanoage.com/Figures/Eight_Allotropes_of_Carbon.png

ULTRA-SMALL PROPERTY

The presence of hubs in scale-free networks raises an interesting question: how do hubs affect the small world property?

[Figure 4.4](#) suggests that they do: airlines build hubs precisely to decrease the number of hops between two airports. The calculations support this expectation, finding that *distances in a scale-free network are either smaller or equal to the distances observed in an equivalent random network*. The precise dependence of the average distance $\langle d \rangle$ on the system size N and the degree exponent γ are captured by the expression [26, 27].

$$d \sim \begin{cases} \text{const.} & \text{if } \gamma = 2, \\ \frac{\ln \ln N}{\ln(\gamma - 1)} & \text{if } 2 < \gamma < 3, \\ \frac{\ln N}{\ln \ln N} & \text{if } \gamma = 3, \\ \ln N & \text{if } \gamma > 3. \end{cases} \quad (4.22)$$

In the following we discuss the behavior of $\langle d \rangle$ in the four regimes predicted by [Eq. 4.22](#), [Fig. 4.12](#):

ANOMALOUS REGIME $\gamma = 2$

According to [Eq. 4.19](#) for $\gamma = 2$ the degree of the biggest hub grows linearly with the system size, i.e. $k_{max} \sim N$. This forces the network into a hub and spoke configuration in which all nodes are at a short distance from each other. In this regime the average path length does not depend on N .

ULTRA-SMALL WORLD $2 < \gamma < 3$

As several real networks have degree exponent between two and three [Table 4.1](#), this regime is of particular practical interest. [Eq. 4.22](#) predicts that the average distance increases as $\ln \ln N$, a significantly slower dependence than the $\ln N$ we derived earlier for random networks. We call networks in this regime ultra-small, as the hubs radically reduce the path length [27]. They do so by linking to a large number of small-de-

gree nodes, creating short distances between them.

To see the implication of the ultra-small property let us consider again the social network with $N \approx 7 \times 10^9$. If the society were to be random, the N -dependent term is $\ln N = 22.66$. In contrast for a scale-free network the N -dependent term is $\ln \ln N = 3.12$ according to Eq. 4.22, supporting our conclusion that hubs radically shrink the distance between the nodes.

CRITICAL POINT $\gamma = 3$

This value is of particular theoretical interest, as the second moment of the degree distribution does not diverge any longer, prompting us to call $\gamma = 3$ the “critical point.” At this critical point the $\ln N$ dependence encountered for random networks returns. Yet the calculations indicate the presence of a double logarithmic correction $\ln \ln N$ [27, 28], which shrink slightly the distances compared to a random network of similar size.

SMALL WORLD $\gamma > 3$

In this regime $\langle k^2 \rangle$ is finite and the average distance follows the small world result derived for random networks. While hubs continue to be present, for $\gamma > 3$ they are not sufficiently large and numerous to have a significant impact on the distance between the nodes.

Taken together, Eq. 4.22 indicates that the more pronounced the hubs are, the more effectively they shrink the distances between the nodes. This conclusion is supported by Fig. 4.11a, which shows the scaling of the average path length for scale-free networks with different γ .

The figure indicates that while for small N the distances in the four regimes are comparable, for large N the differences are remarkable. Further support for this conclusion is provided by the path length distribution for scale-free networks with different γ and N Fig. 4.11b-d. For $N = 10^2$ the path length distributions largely overlap, indicating that at this size differences in γ result in insignificant differences in the path length. For $N = 10^6$, however, p_d observed for different γ are well separated. Fig. 4.11d also shows that the larger the degree exponent, the larger are the distances between the nodes. In summary the scale-free property has two effects on network distances:

- Shrinks the average path lengths.
- Changes the dependence of $\langle d \rangle$ on the system size, as predicted by Eq. 4.21. The smaller γ , the shorter are the distances between the nodes.

Therefore, most scale-free networks of practical interest are not only “small”, but are “ultra-small”. This is a consequence of the hubs, that act as bridges between the many small nodes. Only for $\gamma > 3$ we recover the small-world property encountered in random networks Fig. 4.12.

BOX 4.3

WE ARE ALWAYS CLOSE TO THE HUBS

Frigyes Karinthy in his 1929 short story [30] that introduced the small world concept writes that “it’s always easier to find someone who knows a famous or popular figure than some run-the-mill, insignificant person”.

In other words, we are typically closer to hubs than to less connected nodes. This effect is particularly pronounced in scale-free networks as shown in the figure below. The implications are obvious: there are always short paths linking us to famous individuals like well known scientists or to the president of the United States, as they are hubs with an exceptional numbers of acquaintances. It also means that many of the shortest paths go through these hubs.

In contrast with this expectation, recent measurements designed to replicate the six degrees concept in the online world find that the paths that individuals used to reach their target node involve rather few hubs [31]. That is, individuals involved in successful chains (those that reached their target) were less likely to send a message to a hub than individuals involved in incomplete chains. The reason may be self-imposed, we perceive hubs as busy, hence we contact them only in real need. We therefore avoid them in online experiments of no perceived value to us.

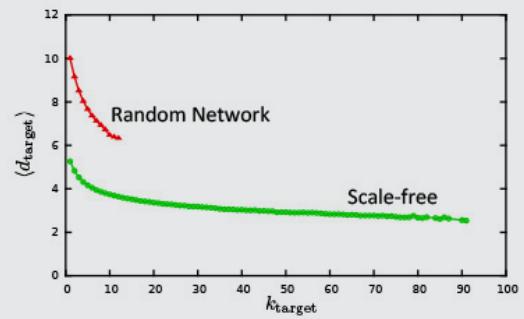


Figure 4.11b
Closing on the hubs

The distance $\langle d_{target} \rangle$ of a node with degree $k = \langle k \rangle$, to a target node with degree k_{target} in a random and a scale-free network. In scale-free networks our distance to the hubs is shorter than in random networks. The figure also documents that in a random network the largest-degree nodes are considerably smaller and hence the path lengths are visibly longer than in a scale-free network. Both networks have $\langle k \rangle = 2$ and $N = 1,000$ and for the scale-free network $\gamma = 2.5$.

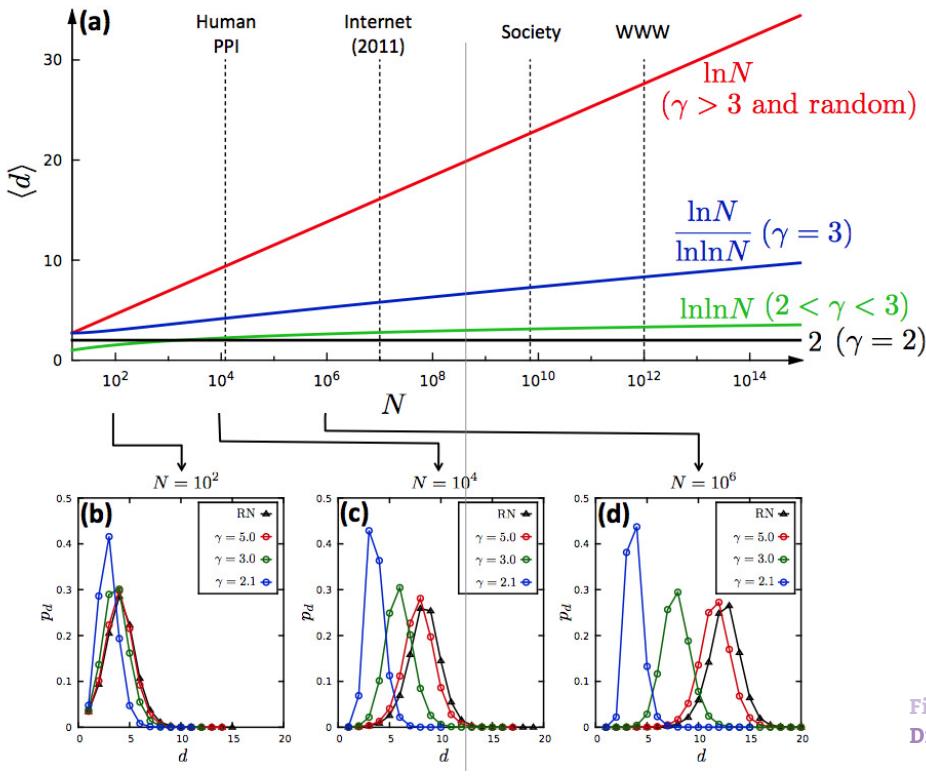


Figure 4.12
Distances in scale-free networks

(a) The scaling of the average path length in the four scaling regimes characterizing a scale-free network: $\ln N$ (scale-free networks with $\gamma > 3$ and random networks), $\ln N / \ln \ln N$ ($\gamma = 3$) and $\ln \ln N$ ($2 < \gamma < 3$). The dotted lines mark the approximate size of several real networks of practical interest. For example, given their modest size, in biological networks the differences in the node to node distances are relatively small in the four regimes. The differences become quite relevant for networks of the size of the social network or the WWW. For these the small-world formula considerably underestimates the real value of $\langle d \rangle$.

(b)(c)(d) Distance distribution for networks of size $N = 10^2, 10^4, 10^6$, illustrating that while for small N ($= 10^2$) the distance distributions is not too sensitive to γ , for large N ($= 10^6$) p_d and $\langle d \rangle$ changes visibly with γ . As (d) shows, the smaller γ , the shorter are the distances between the nodes. The networks were generated using the static model [29] with $\langle k \rangle = 3$.

THE ROLE OF THE DEGREE EXPONENT

Many properties of a scale-free network depend on the value of the degree exponent γ . A close inspection of [Table 4.1](#) indicates that:

- γ varies from system to system, prompting us to explore how the properties of a network change with γ
- For many real systems the degree exponent is between 2 and 3, prompting us to ask: why don't we see systems with $\gamma < 2$ and why are so few systems with $\gamma > 3$? To address these questions next we discuss how the properties of a scale-free network change with γ [Fig. 4.13](#)

ANOMALOUS REGIME ($\gamma \leq 2$)

According to [Eq. 4.18](#), for $\gamma < 2$ the exponent $1/(\gamma - 1)$ is larger than one, hence the fraction of links connected to the largest hub grows faster than the size of the network. This means that for sufficiently large N the degree of the largest hub must exceed the total number of nodes in the network, running out of nodes to connect to. Similarly, for $\gamma < 2$ the average degree $\langle k \rangle$ diverges in the $N \rightarrow \infty$ limit. These odd predictions are only two of the many anomalous features of scale-free networks in this regime. They represent signatures of a deeper problem: large scale-free network with $\gamma < 2$, that lack self-loops or multi-links, cannot exist [BOX 4.4](#). Hence one needs to inspect with caution any research reporting networks with $\gamma < 2$. Such networks can only exist if the hubs have many self-loops or if multiple links can connect the same pair of nodes.

SCALE-FREE REGIME ($2 < \gamma < 3$)

In this regime the first moment $\langle k \rangle$ of the degree distribution is finite but the second and higher moments diverge as $N \rightarrow \infty$. Consequently scale-free networks in this regime are ultra-small (see [SECTION 4.6](#)). [Eq. 4.18](#) predicts that k_{\max} grows with the size of the network with exponent $1/(\gamma - 1)$, which is smaller than one. Hence the market share of the largest hub, k_{\max}/N , representing the fraction of nodes that connect to it, decreases as $k_{\max}/N \sim N^{(2-\gamma)/(\gamma-1)}$.

As we will see in the coming chapters, many interesting features of scale-free networks, from their robustness to failures to anomalous spreading phenomena, are linked to this regime.

RANDOM NETWORK REGIME ($\gamma > 3$)

According to Eq. 4.20 for $\gamma > 3$ both the first and the second moments are finite. For all practical purposes the properties of a scale-free network in this regime are difficult to distinguish from the properties of a random network of similar size. For example Eq. 4.21 indicates that the average distance between the nodes converges to the small-world formula derived for random networks. The reason is that for large γ the degree distribution p_k decays sufficiently fast to make the hubs smaller and less numerous. The larger γ , the smaller are the hubs (see Eq. 4.18), hence the more indistinguishable is the structure and the behavior of a scale-free network from that of a random network.

Table 4.1 and Fig. 4.12 also indicate that there are fewer networks with $\gamma > 3$, prompting us to ask: does this imply that networks with $\gamma > 3$ cannot exist? A quick calculation indicates that they may exist, but it is hard to distinguish them from a random network. To document the presence of a power-law degree distribution we ideally need 2-3 orders of magnitude of scaling, which means that k_{\max} should be at least $10^2 - 10^3$ times larger than k_{\min} . By inverting Eq. 4.18 we can calculate the network size necessary to observe the desired scaling regime between k_{\min} and k_{\max} , obtaining

$$N \gg \frac{k_{\max}}{k_{\min}}^{\gamma-1}. \quad (4.23)$$

For example, in order to document the scale-free nature of a network with $\gamma = 5$ with $k_{\min} \sim 1$ and $k_{\max} \approx 10^2$, according to Eq. 4.23 the size of the network must exceed $N \gg 10^8$. There are very few network maps of this size available for research. Therefore, there may be many real networks with exponent larger than 3, but given their limited size, it is difficult to obtain convincing evidence of their scale-free nature. Hence they are mistakenly classified as networks with an exponential degree distribution.

In summary, we find that the behavior of scale-free networks depends on the value of the degree exponent γ . Theoretically the most interesting regime is $2 < \gamma < 3$, where scale-free networks are ultra-small and $\langle k^2 \rangle$ diverges. Interestingly, many networks of practical interest, from the WWW to protein interaction networks, are in this regime.

BOX 4.4

SCALE-FREE NETWORK WITH $\gamma < 2$ DO NOT EXIST

To see why networks with $\gamma < 2$ are problematic, we need to attempt to build one. A degree sequence that can be turned into simple graph (i.e. a graph lacking multilinks or self-loops) is called graphical [32]. Yet, not all degree sequences are graphical: if for example the number of stubs is odd, then we will always have an unmatched stub, as shown in Fig. 4.13b.

The graphicality of a degree sequence can be tested with an algorithm proposed by Erdős and Gallai [32, 33, 34, 35]. If we apply the algorithm to scale-free networks we find that the number of graphical degree sequences drops to zero for $\gamma < 2$. Hence degree distributions with $\gamma < 2$ cannot be turned into a network. Indeed, for networks in this regime the largest hub grows faster than N . If we do not allow self-loops and multi-links, then the degree of the largest hub cannot exceed $N - 1$.

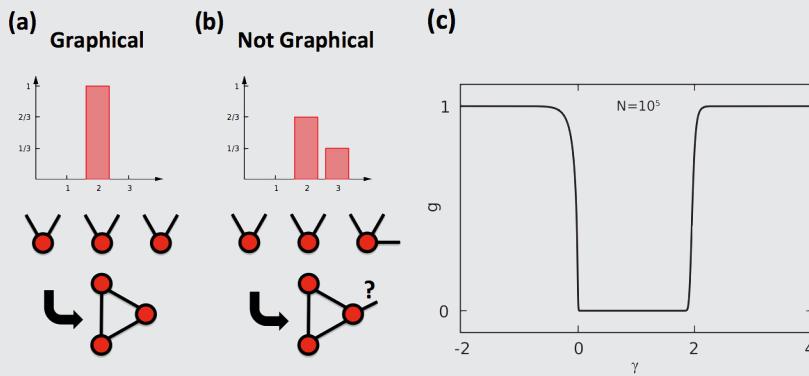


Figure 4.13
Networks with $\gamma < 2$ are not graphical

(a-b) Two degree distributions and the corresponding degree sequences. The difference is limited to the degree of a single node. While we can build a network consistent with the degree distribution (a), it is impossible to build one from (b), as one stub always remains unmatched. Hence (a) is graphical, while (b) is not.

(c) Fraction of networks with a given γ that are graphical. A large number of degree sequences with degree exponent γ and $N = 10^5$ were generated, testing the graphicality of each network.

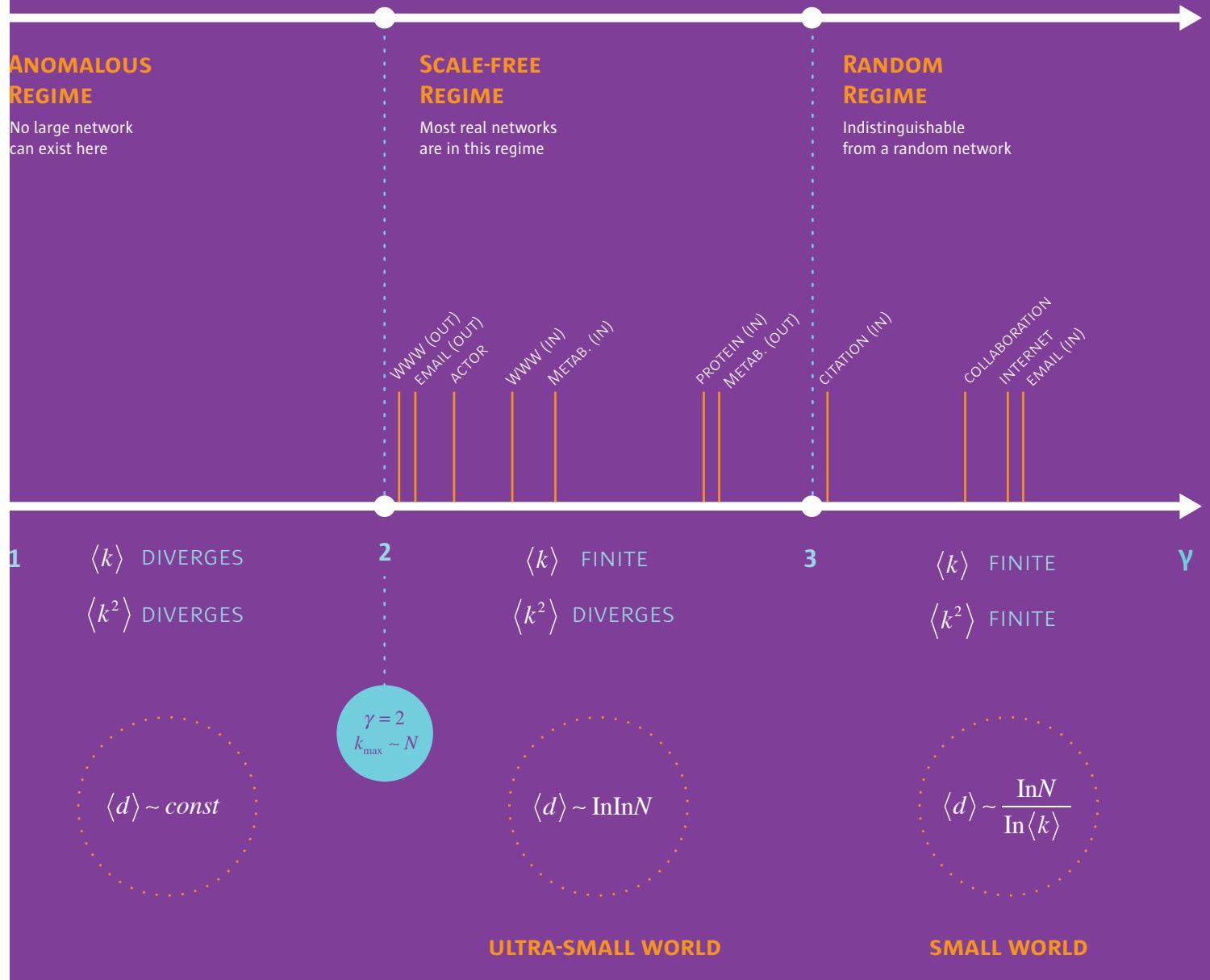
While virtually all networks with $\gamma > 2$ are graphical, it is impossible to find graphical networks with $0 < \gamma < 2$.

DEPENDENT PROPERTIES

A SUMMARY OF THE γ DEPENDENT PROPERTIES
OF SCALE-FREE NETWORKS

FIG. 4.14

The degree exponents shown in the figure were taken from Table 4.1. Note that not all listed γ values show statistical significance, as we lack the proper fitting function. Case in point are the Internet and the email datasets, for which earlier studies reported $\gamma < 3$. To determine the precise value of γ , we need proper models, a topic discussed in Chapter 6.



GENERATING NETWORKS WITH A PRE-DEFINED DEGREE DISTRIBUTION

The Erdős-Rényi model generates networks with a Poisson degree distribution. The empirical results discussed in this chapter indicate, however, that the degree distribution of most real networks significantly deviates from a Poisson form. This raises an important question: how do we generate networks with an arbitrary p_k ? In the following we discuss the three most frequently used algorithms for this purpose.

CONFIGURATION MODEL

The configuration model helps us build a network with a pre-defined degree sequence Fig. 4.15a. In the obtained network each node has a pre-defined degree k_i , but otherwise the network is wired randomly. Consequently the obtained network is often called a random network with a pre-defined degree sequence. By repeatedly applying this procedure to the same degree sequence we can generate different networks with the same p_k Fig. 4.14, panels (2a)-(2c). A couple of a caveats to consider:

- The probability to have a link between nodes of degree k_i and k_j is

$$p_{ij} = \frac{k_i k_j}{2L - 1} \quad (4.24)$$

Indeed, a stub starting from node i can connect to $2L - 1$ other stubs. Of these, k_j are attached to node j . So the probability that a particular stub is connected to a stub of node j is $k_j / (2L - 1)$. As node i has k_i stubs, it will have k_i attempts to link to j , resulting in Eq. 4.24.

- The obtained network will contain self-edges and multi-edges. We can choose to reject stub pairs that lead to these, but if we do so, we may not be able to complete the network. Rejecting self- or multi-edges means that not all possible matchings appear with equal probability. Hence Eq. 4.24 will not be valid any longer, making analytical calculations difficult. The number of self- and multi-edges goes to zero for large networks, so in most cases we do not need to exclude them [39]. The configuration model is frequently used in analytical calculations, as Eq. 4.24 and its inherently random character helps us calculate numerous network measures.

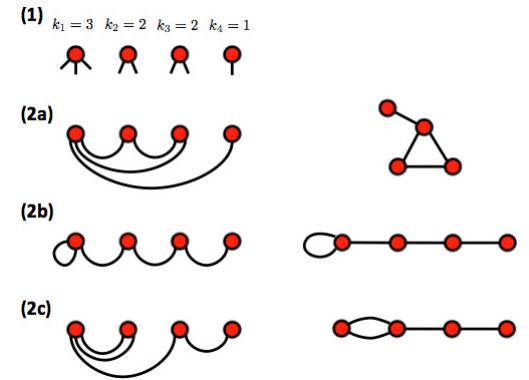


Figure 4.15a
The configuration model

The configuration model allows us to build a network where each node has some pre-defined degree [37, 38]. It consists of the following steps:

(1) Degree sequence: Assign a degree to each node, represented as stubs or half-links. The degree sequence is either generated analytically from a preselected p_k distribution BOX 4.5, or it is extracted from the adjacency matrix of a real network. We must start from an even number of stubs, otherwise we will be left with unpaired stubs.

(2) Network assembly: Randomly select a stub pair and connect them. Then randomly choose another pair from the remaining $2L - 2$ stubs and connect them. This procedure is repeated until all stubs are paired up. Depending on the order in which the stubs were chosen, we obtain different networks. Some networks include cycles (2a), others self-edges (2b) or multi-edges (2c). Yet, the expected number of self- and multi-edges goes to zero in the $N \rightarrow \infty$ limit.

DEGREE PRESERVING RANDOMIZATION

As we explore the properties of a real network, we often need to ask if a certain network property is predicted by its degree distribution alone, or if it represents some additional property not contained in p_k . To answer this question we need to generate networks that are wired randomly, but whose p_k is identical to the original network.

This can be achieved through the degree-preserving randomization [40] described in Fig. 4.14. The idea behind the algorithm is simple: we randomly choose two links in the network and swap them, so that the degree of each of the four involved nodes in the swap remains unchanged. Hence, hubs will stay hubs and small-degree nodes will retain their small degree, but the wiring diagram of the generated network will be randomized. Note that degree-preserving randomization is different from full randomization, where we swap links without preserving the node degrees Fig. 4.14. Complete randomization turns any network into an Erdős-Rényi network, hence independent of the original p_k , the randomized version will have a Poisson degree distribution.

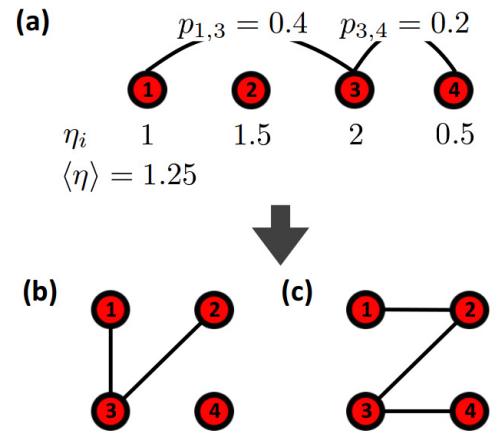


Figure 4.15b
Hidden parameter model

We start with N isolated nodes and assign to each node a “hidden parameter” η_i , which can be randomly selected from a $\rho(\eta)$ distribution or it is provided by a deterministic sequence $\{\eta_i\}$. We next connect each node pair with probability

$$p(\eta_i, \eta_j) = \frac{\eta_i \eta_j}{\langle \eta \rangle N}.$$

For example, the figure shows the probability to connect nodes (1,3) and (3,4). After connecting the nodes, we end up with the networks shown in (b) or (c), representing two independent realizations generated by the same hidden parameter sequence (a). The expected number of links in the obtained network is

$$L = \frac{1}{2} \sum_N^{i,j} \frac{\eta_i \eta_j}{\langle \eta \rangle N} = \frac{1}{2} \langle \eta \rangle N.$$

Just like in the random network model, L will differ from network to network, following a bounded distribution. If we wish to control precisely the average degree $\langle k \rangle$ we can add the L links to the network one by one. The end points i and j of each link are then chosen randomly with a probability proportional to η_i and η_j , following. In this case we connect i and j only if they were not connected previously.

BOX 4.5

GENERATING A DEGREE SEQUENCE WITH POWER-LAW DISTRIBUTION

The degree sequence of an undirected network is a non-increasing sequence of the node degrees. For example, the degree sequence of each of the networks shown in Fig. 4.15a is {3, 2, 2, 1}. As Fig. 4.15a illustrates, the degree sequence in general does not uniquely identify a graph. There can be multiple graphs with the same degree sequence. We often need to generate a degree sequence from a pre-defined degree distribution. Our purpose here is to provide the tools to achieve this. We start from an analytically pre-defined degree distribution, like $p_k \sim k^{-\gamma}$, shown in panel (a). Our goal is to generate a degree sequence $\{k_1, k_2, \dots, k_N\}$ of N degrees that follow the distribution p_k . We start by calculating the complementary cumulative distribution function

$$D(k) = \sum_{k' \geq k} p_{k'}, \quad (4.25)$$

shown in (b). $D(k)$ is between 0 and 1, and the step size at any k equals p_k . Therefore, to generate a sequence of N random numbers following a pre-defined p_k distribution, we generate N random numbers r_i , $i = 1, \dots, N$, chosen from the $(0, 1)$ interval. For each r_i we use the plot in (b) to assign a degree k_i . The obtained $k_i = D^{-1}(r_i)$ set will follow the desired p_k distribution. Note that the degree sequence assigned to a p_k is not unique - we can generate multiple sets of $\{k_1, \dots, k_N\}$ sequences compatible with the same p_k .

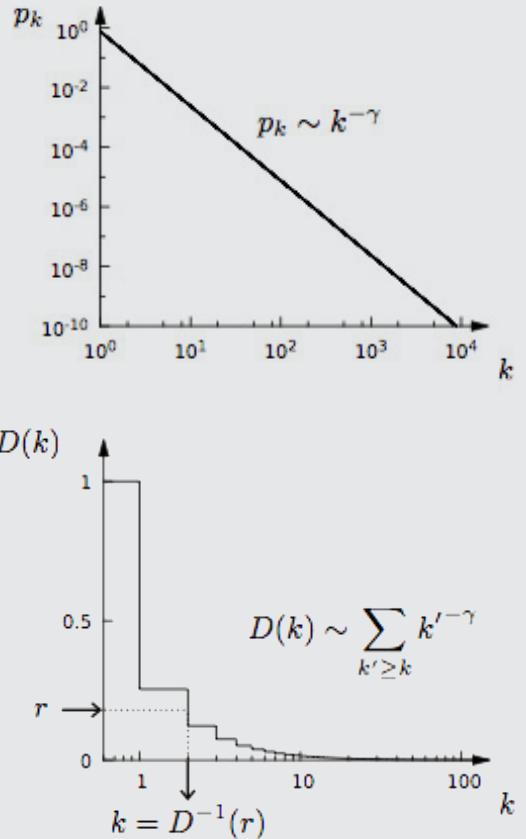


Figure 4.16

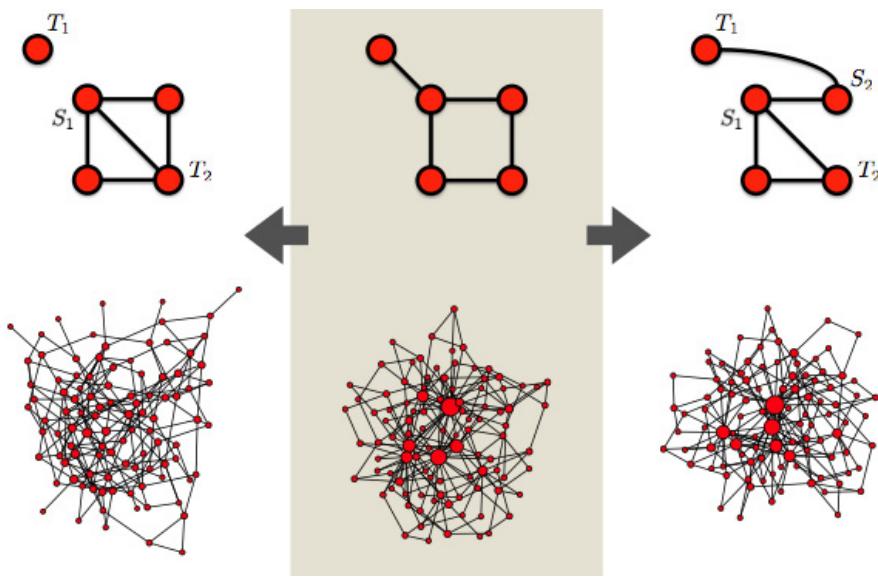


Figure 4.17
Degree preserving randomization

Two randomization methods are used to generate random references to a given network [40]. Full randomization generates a random (Erdős–Rényi) network with the same N and L as the original network. For this we select randomly a source node (S_1) and two target nodes, where the first target is linked directly to the source node (T_1) and the second target is unconnected to it (T_2). We then rewire the S_1-T_1 link, turning it into an S_1-T_2 link. As a result the degree of the target nodes T_1 and T_2 changes. We perform this procedure once for each link in the network.

Degree-preserving randomization generates a network in which each node has exactly the same degree as in the original network, but the network's wiring diagram has been randomized. We select two source (S_1, S_2) and two target nodes (T_1, T_2), such that initially there is a link between S_1 and T_1 , and a link between S_2 and T_2 . We then swap the two links, creating an S_1-T_2 and an S_2-T_1 link. This swap leaves the degree of each node unchanged. We repeat this process until we rewire at least once each link.

Bottom panels: Starting from a scale-free network (middle panel), full randomization eliminates the hubs and turns the network into a random network (left panel). In contrast, degree-preserving randomization leaves the hubs in place and hence the network remains scale-free (right panel).

HIDDEN PARAMETER MODEL

In their most general (and most useful) form the configuration and the rewiring model generate loops and multi-links. Loops and multi-links are absent, however, from many real networks. We can use the hidden parameter model, described in Fig. 4.15b, to generate networks with a pre-defined p_k but without multi-links and self-loops [41, 42, 43]. In the model we start from N isolated nodes and assign each node i a hidden parameter η_i , chosen from a distribution $\rho(\eta)$. The nature of the network generated by the hidden parameter model depends on the selection of a $\{\eta\}$ hidden parameter sequence.

There are two ways to generate the appropriate hidden parameters:

- (i) η_i can be a sequence of N random number chosen from a pre-defined $\rho(\eta)$ distribution. In this case the degree distribution of the obtained network is

$$p_k = \int \frac{e^{-\eta} \eta^k}{k!} \rho(\eta) d\eta. \quad (4.26)$$

- (ii) η can come from a deterministic sequence $\{\eta_1, \eta_2, \dots, \eta_N\}$. In this case the degree distribution of the obtained network is

$$p_k = \frac{1}{N} \sum_j \frac{e^{-\eta_j} \eta_j^k}{k!}. \quad (4.27)$$

The hidden parameter model offers a particularly simple method to generate a scale-free network. Indeed, using

$$\eta_i = c / i^\alpha, i = 1, \dots, N. \quad (4.28)$$

as the sequence of hidden parameters, according to Eq. 4.27 the obtained network will have the degree distribution

$$p_k \sim k^{-(1+1/\alpha)} \quad (4.29)$$

for large k . We can use $\langle \eta \rangle$ to tune $\langle k \rangle$ as Eq. 4.26 and Eq. 4.27 imply $\langle k \rangle = \langle \eta \rangle$. The three methods discussed above for creating networks with a pre-defined p_k raise the following question: how do we decide which one to use? Our choice depends on whether we start from a degree sequence $\{k_i\}$ or a degree distribution p_k and whether we can tolerate self-loops and multiple links between two nodes. The decision tree involved in this choice is provided in Fig. 4.18.

In summary, the configuration model, degree-preserving randomization and the hidden parameter model are attractive because they generate networks with a pre-defined degree distribution and allow us to analytically calculate several network properties.

We will turn to these each time we explore if a certain network property is a consequence of the network's degree distribution, or it represents some emerging property of the modeling network **BOX 4.6**. Yet, these mod-

els also have a number of limitations:

- These algorithms do not tell us why a network has a certain degree distribution
- Several important network characteristics, present in real networks, from clustering to degree correlations, are lost during randomization

Hence the networks generated by these algorithms are a bit like a photograph of a painting: at first look they appear to be the same as the original. But upon closer inspection we realize that many details, from the texture of the canvas to the brush strokes, are lost.

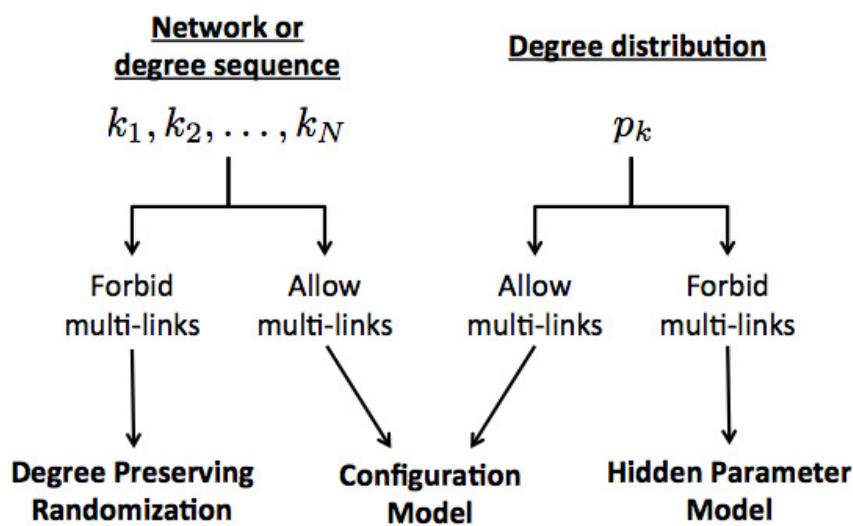


Figure 4.18
Choosing the proper generative model

The choice of the appropriate generative model depends on our starting point as well as our tolerance towards self-loops and multi-links. If we start from the analytical form of the degree distribution, p_k , then the goal is to generate networks whose degree distribution is consistent with p_k . In this case if we allow self-loops and multi-links, the configuration model is an appropriate choice; if we wish to forbid them, then the hidden parameter model is a better choice.

If we start from a real network or known degree sequence our goal is often to generate networks with the degree sequence identical to the original network. Again if we allow self-loops and multi-links, the configuration model is an appropriate choice; if we wish to forbid them, we can use degree-preserving randomization.

BOX 4.6

TESTING THE SMALL-WORLD PROPERTY

A common practice in the network literature is to compare the distances observed in a real network to the small-world formula Eq. 4.19 from CHAPTER 3. Yet, Eq. 4.19 was derived for random networks, while most real networks do not have a Poisson degree distribution. If the network is scale-free, then Eq. 4.22 offers the appropriate formula. That, however, provides only the scaling of the distance with N , and not its absolute value. Hence instead of trying to fit the average distance, we often ask the following question: are the distances observed in the real network comparable with the distances observed in a randomized network with the same degree distribution? We can use degree preserving randomization to answer this. We illustrate the procedure on the protein interaction network (PIN) of yeast.

- (i) Original p_d : we start by measuring the distance distribution p_d of the original network, obtaining $\langle d \rangle = 5.61$ (red curve).
- (ii) Full randomization: next we generate a random network with the same N and L as the original network. The obtained p_d (blue curve) is visibly shifted to the right, providing $\langle d \rangle = 7.13$, much larger than the original $\langle d \rangle = 5.61$. It is tempting to conclude that the protein interaction network is affected by some unknown organizing principle that keeps the distances shorter than expected in a random configuration. The result (iii) shows that this would be a flawed conclusion, as the difference is explained by the degree distribution.
- (iii) Degree preserving randomization: as the original network is scale-free, the proper random reference is a network with the same degree distribution as the original. Hence we determine p_d after degree-preserving randomization, finding that it is comparable to the original p_d (green curve).

This indicates that a random network overestimates the distances between the nodes, as it is missing the hubs presented in the original network. The network obtained by degree preserving randomization preserves these hubs, and its distances are comparable to the original network. This example illustrates the importance of choosing the proper random reference frame.

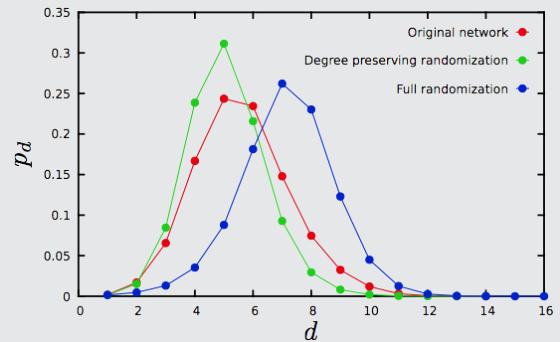


Figure 4.19
Randomizing real networks

The distance distribution p_d (red symbols) between each node pair in *S. Cerevisiae* protein-protein interaction network Table 4.1. The purple symbols provide the path-length distribution obtained under full randomization, which turns the original network into an Erdős-Rényi network with the same N and L as the original network Fig. 4.17.

The green symbols correspond to p_d of the network obtained after degree-preserving randomization, which keeps the degree of each node unchanged.

We have: $\langle d \rangle = 5.61 \pm 1.64$ (original), $\langle d \rangle = 7.13 \pm 1.62$ (full randomization), $\langle d \rangle = 5.08 \pm 1.34$ degree-preserving randomization.

SUMMARY

There are two main reasons why the scale-free property played a key role in the emergence of network science.

First, many networks of scientific and practical interest, from the WWW to the cell, are scale-free.

Second, once the hubs, that accompany the scale-free property, are present, they have an enormous impact on the system's behavior. The ultra-small property offers a first hint of the hubs's impact on a network's properties; we will encounter many more in the coming chapters.

As we continue exploring the consequences of the scale-free property, we must keep in mind that the power-law form [Eq. 4.1](#) is rarely seen in this pure form in real systems. The reason is simple: a host of processes affect the topology of real networks, which also influence the shape of the degree distribution. We will discuss these processes in the coming chapters. The diversity of these processes and the complexity of the resulting p_k confuses those who approach these networks through the narrow perspective of the quality of fit to a pure power law. Instead the scale-free property tells us that we must distinguish between two rather different classes of networks:

- Bounded networks are networks whose degree distribution decrease exponentially or faster for high k . Examples of p_k in this class include the Poisson, Gaussian, or the simple exponential distribution. The Erdős-Rényi network is the best known example of the networks belonging to this class. Bounded networks lack outliers, consequently most nodes have comparable degrees. Real networks in this class include highway networks, the power grid or the atomic networks observed in crystalline or amorphous materials.
- Unbounded networks are networks whose degree distribution has a fat tail in the high- k region. Networks with a power-law degree distribution [Eq. 4.1](#) offer a representative example of this class. A common property of these networks is that the node degrees span several orders

of magnitude, differences that are difficult to explain using a bounded distribution. Outliers, or exceptionally high-degree nodes, are not only allowed but expected in these networks. Networks in this class include the WWW, the Internet, the protein interaction networks, and many social and online networks. While it would be desirable to fit and statistically validate the precise form of the degree distribution, often it is sufficient to decide the class to which a given network belongs: bounded or unbounded (see ADVANCED TOPICS 4A). If the degree distribution is bounded, the random network model offers a reasonable starting point to understand its topology. If the degree distribution is unbounded, a scale-free network offers a better approximation.

In summary, to understand the properties of real networks, it is often sufficient to remember that in scale-free networks a few highly connected hubs coexist with a large number of small nodes. In contrast in random networks most nodes have comparable degrees and hubs are absent. The presence or absence of the hubs plays an important role in the system's behavior. The purpose of this chapter was to explore the basic characteristics of scale-free networks. We are left, therefore, with an important question: why are networks scale-free? The next chapter will provide the answer. Keeping up with the framework established in the previous chapter, the results discussed in this chapter allow us to formulate our next network law:

The Second Law: scale-free property

Many real networks are characterized by a fat-tailed degree distribution. This means that many small-degree nodes are held together by a few hubs.

Let us recap the validity of this law in the context of the three criteria established in CHAPTER 3:

A. Quantitative formulation: Eq. 4.1 offers the quantitative formulation of the Second Law, indicating that the degree distribution of such networks can be approximated by a power law.

B. Universality: as discussed in SECTION 4.5, the scale-free property is a common feature of many real networks, from the WWW to the protein interaction network in the cell.

C. Non-random origins: the scale-free property represents a dramatic deviation from the Poisson degree distribution characterizing random networks, hence it can not be explained in the context of the random network model.

BOX 4.7

At a glance Scale-free networks

DEGREE DISTRIBUTION

Discrete form:

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}.$$

Continuous form:

$$p(k) = (\gamma - 1) k_{\min}^{\gamma-1} k^{-\gamma}.$$

SIZE OF THE LARGEST HUB

$$k_{\max} \sim k_{\min} \frac{1}{N^{\gamma-1}}.$$

MOMENTS OF p_k

$2 < \gamma < 3$: $\langle k \rangle$ finite, $\langle k^2 \rangle$ diverges when $N \rightarrow \infty$.

$\gamma > 3$: $\langle k \rangle$ and $\langle k^2 \rangle$ finite.

DISTANCES IN A SCALE-FREE NETWORK

$$d \sim \begin{cases} \text{const.} & \text{if } \gamma = 2, \\ \frac{\ln \ln N}{\ln(\gamma - 1)} & \text{if } 2 < \gamma < 3, \\ \frac{\ln N}{\ln \ln N} & \text{if } \gamma = 3, \\ \ln N & \text{if } \gamma > 3. \end{cases}$$

ADVANCED TOPICS 4.A

POWER LAWS

Power laws have a convoluted history in natural and social sciences, being interchangeably called fat-tailed, heavy-tailed, long-tailed, Pareto, or Bradford distributions. They also have a series of close relatives, like log-normal, Weibull, or Lévy distributions. The purpose of this section is to discuss the properties of some of the most frequently encountered distributions in network science and their relationship to the power law function discussed in this chapter.

Many quantities in nature, from the height of individuals to the probability of being in a car accident, follow bounded distributions. A common property of these is that p_k decays either exponentially (e^{-x}), or faster than exponentially (e^{-x^2/σ^2}) for high x . Consequently events with high x are extremely rare, the largest expected x being unable to exceed some upper value x_{max} that is not too different from $\langle x \rangle$ (it is “bounded”). The high- x regime is often called the tail of the distribution, and given the absence of numerous events in the tail, these distributions are also called thin tailed. Well known examples of such bounded distributions are the Poisson, Gaussian (normal), or the exponential distribution [Table 4.2](#).

In contrast the terms “fat tailed”, “heavy tailed”, “long tailed”, or “unbounded” refer to p_k whose decay at large x is slower than exponential. In these distributions one often encounters events characterized by very large x values, unusually called outliers or rare events. The power-law distribution of [Eq. 4.1](#) represents the best known example of such unbounded distributions. In the following we will discuss the basic properties of the most commonly encountered bounded and unbounded distributions in network science [Table 4.2](#).

BOUNDED DISTRIBUTIONS (EXPONENTIALS)

Analytically the simplest bounded distribution is the exponential distribution $e^{-\lambda x}$. Within network science the most prominent bounded distribution is the Poisson distribution, capturing the degree distribution of a random network. Outside of network science the most frequently encountered member of this class is the normal (Gaussian) distribution.

A common property of bounded distributions comes in their tail: for high x they decay exponentially or faster. Consequently, the expected largest x obtained after we draw N numbers from a bounded p_x grows as $x_{max} \sim \log(N)$ or slower. This means that outliers, representing unusually high x -values, are rare. They are so rare that they are effectively forbidden, meaning that they do not occur with any meaningful probability. Instead, most events drawn from a bounded distribution are not too far from $\langle x \rangle$.

UNBOUNDED DISTRIBUTIONS (POWER LAWS)

An instantly recognizable feature of an unbounded distribution is that the magnitude of the events x drawn from it vary widely, spanning several orders of magnitude. The most prominent member of this class is the power-law distribution discussed in [SECTION 4.2](#). Its relevance to networks is provided by several factors:

- Many quantities occurring in networks science, like degrees, link weights and betweenness centrality, follow a power-law distribution in many real and model networks.
- The power-law form is analytically predicted by some of the most fundamental network models [CHAPTER 5](#).

In contrast with bounded distributions, in unbounded distributions the size of the largest event after N trials scales as $x_{max} \sim N^{\zeta}$ where ζ is some integer related to the exponent γ characterizing the p_x distribution. As N^{ζ} grows fast, rare events or outliers occur with a noticeable frequency, often dominating the properties of the system.

CROSSOVER DISTRIBUTION (LOG-NORMAL, STRETCHED EXPONENTIAL)

Several functions interpolate between bounded and unbounded distributions. This means that depending on their parameters, they can be used to fit unbounded distributions, but technically speaking they are bounded, as their tail for large x decays exponentially or faster. In the following we discuss the properties of the most frequently encountered crossover distributions.

A power law with exponential cut-off is often used in network theory to fit the degree distribution of real networks. Its density function has the form:

$$p_k = Cx^{-\gamma} e^{-\lambda x} \quad (4.30)$$

$$C = \frac{\lambda^{1-\gamma}}{\Gamma(1-\gamma, \lambda x_{min})}, \quad (4.31)$$

where $x > 0$ and $\gamma > 0$. The analytical form of [Eq. 4.30](#) directly captures its crossover nature: it combines a power-law term, a key component of unbounded distributions, with an exponential term, responsible for its bounded tail. We can explore its crossover characteristics by taking the logarithm of [Eq. 4.30](#),

$$\ln p_x = \ln C - \gamma \ln x - \lambda x. \quad (4.32)$$

For $x \ll 1/\lambda$ the second term on the r.h.s dominates, suggesting that the distribution follows a power law with exponent γ . Once $x \geq 1/\lambda$, the λx term overcomes the $\ln x$ term, resulting in an exponential cutoff for high x .

Stretched exponential (Weibull distribution) is similar to Eq. 4.30 except that we have a fractional power law in the exponential. Its density function has the form

$$p_x = Cx^{\beta-1}e^{-\lambda x^\beta} \quad (4.32)$$

$$C = \beta x^{-\beta} \exp \left(x_{\min} / \lambda \right)^\beta . \quad (4.33)$$

In most applications x varies between 0 and $+\infty$. In Eq. 4.32 β is the *stretching exponent*, determining the properties of p_x :

- For $\beta = 1$ we recover a simple exponential function
- If β is between 0 and 1, the graph of $\log p_x$ versus x is “stretched”, meaning that it spans several orders of magnitude in x . This is the regime where a stretched exponential is difficult to distinguish from a pure power law. The closer β is to 0, the more similar is p_x to the power law x^1
- By taking a logarithm of Eq. 4.32,

$$\ln p_x \sim (\beta - 1) \ln x - \lambda x^\beta , \quad (4.34)$$

we can see why the stretched exponential is often used to approximate a power law distribution. Indeed, for small β and not too large x the function will be indistinguishable from a power law with slope $(\beta-1)$. For large x the term λx^β becomes dominant, generating an exponential cutoff in p_x .

- If $\beta > 1$ we observe a “compressed” exponential function, meaning that x varies in a very narrow range.
- For $\beta = 2$ Eq. 4.32 reduces to the normal distribution.

As we will see in CHAPTERS 5 and 6, several important network models predict a stretched exponential degree distribution.

A *log-normal distribution (Galton or Gibrat distribution)* emerges if $\ln x$ follows a normal distribution. Typically a variable follows a log-normal distribution if it is the product of many independent positive random numbers. We encounter log-normal distributions in finance, representing the compound return from a sequence of trades, where the compound return is the product of the individual trades. The probability density function of a log-normal distribution is

$$p_x = \frac{1}{\sqrt{2\pi}\sigma x} \exp \left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right] \quad (4.35)$$

Hence a log-normal is like a normal distribution except that its variable in the exponential term is not x , but $\ln x$. To understand why a log-normal is occasionally used to fit a power law distribution, let us take the logarithm of Eq. 4.35,

$$\ln p_x = \ln \frac{1}{\sqrt{2\pi}\sigma} - \ln x - \frac{(\ln x - \mu)^2}{2\sigma^2} \quad (4.36)$$

If $\ln x \ll \mu$ then the last term is negligible and the distribution follows a power law with slope -1 due to the second term $\ln x^{-1}$.

Therefore, for distributions that appear to follow a power law with slope -1, a log-normal function will likely offer a reasonable fit. For large σ the log-normal distribution may resemble power laws with other exponents too (see dashed line in Fig. 4.20 with slope 2.5). Note that for reasons that are discussed in Box 4.4, a degree distribution with $\gamma=1$ is forbidden in most real networks, hence log-normal distributions are rarely used to approximate a network's degree distribution. In summary, in most areas where we encounter fat-tailed distributions, there is an ongoing debate about the form of the distribution that offers the best fit to the data. Common candidates include a simple power law, a stretched exponential, or a log-normal function. In many systems it is impossible to distinguish these distribution based on empirical data only. Hence as long as there is empirical data to be fitted, the debate surrounding the best fit will never die out.

The debate can be best resolved by developing accurate mechanistic models, which analytically predict the expected degree distribution. We will see in the coming chapters that the distributions that are analytically predicted by network theory are the Poisson, simple exponential, stretched exponential, and power law. The remaining distributions in Table 4.2 are occasionally used to fit the degrees of some networks, despite the fact that we lack theoretical backing to support their relevance for network science.

NAME	p_x	C_i	$\langle x \rangle$	$\langle x^2 \rangle$
Exponential (continuous)	$e^{-\lambda x}$	$\lambda e^{\lambda x_{\min}}$	$\lambda^{-1} + x_{\min}$	$\frac{(\lambda x_{\min} + 1)^2 + 1}{\lambda^2}$
Exponential (discrete)	$e^{-\lambda x}$	$(1 - e^{-\lambda})e^{-\lambda x_{\min}}$	$(e^\lambda - 1) + x_{\min}$	$\frac{e^\lambda + 1}{(e^\lambda - 1)^2} + \frac{2x_{\min}}{e^\lambda - 1} + x_{\min}^2$
Poisson	$\mu^x / x!$	$\left[e^\mu - \sum_{k=0}^{x_{\min}-1} \frac{\mu^k}{k!} \right]^{-1}$	$\mu - e^{-\mu} \sum_{x=0}^{x_{\min}-1} \frac{\mu^x}{x!} x$	$\mu^2 + \mu - e^{-\mu} \sum_{x=0}^{x_{\min}-1} \frac{\mu^x}{x!} x^2$
Power law (continuous)	$x^{-\alpha}$	$(\alpha - 1)x_{\min}^{\alpha-1}$	$\begin{cases} x_{\min}^{\frac{\alpha-1}{\alpha-2}} & \text{if } \alpha > 2 \\ \infty & \text{if } \alpha \leq 2 \end{cases}$	$\begin{cases} x_{\min}^2 \frac{\alpha-1}{\alpha-3} & \text{if } \alpha > 3 \\ \infty & \text{if } \alpha \leq 3 \end{cases}$
Power law (discrete)	$x^{-\alpha}$	$1/\zeta(\alpha, x_{\min})$	$\begin{cases} \frac{\zeta(\alpha-1, x_{\min})}{\zeta(\alpha, x_{\min})} & \text{if } \alpha > 2 \\ \infty & \text{if } \alpha \leq 2 \end{cases}$	$\begin{cases} \frac{\zeta(\alpha-2, x_{\min})}{\zeta(\alpha, x_{\min})} & \text{if } \alpha > 3 \\ \infty & \text{if } \alpha \leq 3 \end{cases}$
Power law with cutoff (exponential)	$x^{-\alpha} e^{-\lambda x}$	$\frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha, \lambda x_{\min})}$	$\lambda^{-1} \frac{\Gamma(2-\alpha, \lambda x_{\min})}{\Gamma(1-\alpha, \lambda x_{\min})}$	$\lambda^{-2} \frac{\Gamma(3-\alpha, \lambda x_{\min})}{\Gamma(1-\alpha, \lambda x_{\min})}$
Stretched exponential	$x_X^{\beta-1} e^{-\lambda x^\beta}$	$\beta \lambda e^{\lambda x_{\min}^\beta}$	$\lambda^{-1/\beta} e^{\lambda x_{\min}^\beta} \Gamma(1/\beta + 1, \lambda x_{\min}^\beta)$	$\lambda^{-2/\beta} e^{\lambda x_{\min}^\beta} \Gamma(2/\beta + 1, \lambda x_{\min}^\beta)$
Log-normal	$\frac{1}{x} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$	$\sqrt{\frac{2}{\pi\sigma^2}} \left[\operatorname{erfc}\left(\frac{\ln x - \mu}{\sqrt{2\sigma}}\right) \right]^{-1}$	$\frac{e^{\mu + \sigma^2/2}}{1 - \operatorname{erf}\left[\frac{-\mu + \ln x_{\min}}{\sqrt{2\sigma}}\right]}$	$\frac{e^{2(\mu + \sigma^2)}}{1 - \operatorname{erf}\left[\frac{\mu + 2\sigma^2 - \ln x_{\min}}{\sqrt{2\sigma}}\right]}$
Gaussian	$\exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]$	$\sqrt{\frac{1}{2\pi\sigma^2}}$	μ	$\sigma^2 + \mu^2$

Table 4.2
Distributions in network science

The table lists several frequently encountered distributions in network science. For each distribution we show the density function p_x , the appropriate normalization constant C such that

$$\int_{x=x_{\min}}^{\infty} Cf(x) dx = 1$$

for the continuous case or

$$\sum_{x=x_{\min}}^{\infty} Cf(x) = 1$$

for the discrete case. Given that $\langle x \rangle$ and $\langle x^2 \rangle$ play an important role in network theory, we list the analytical form of these two quantities for each distribution. As many of these distributions diverge at $x = 0$, $\langle x \rangle$ and $\langle x^2 \rangle$ are calculated assuming that there is a small cutoff x_{\min} in the system. In networks x_{\min} often corresponds to the smallest positive degree, $k_{\min-1}$, or could reflect the smallest degree k_{\min} for which the appropriate distribution offers a good fit.

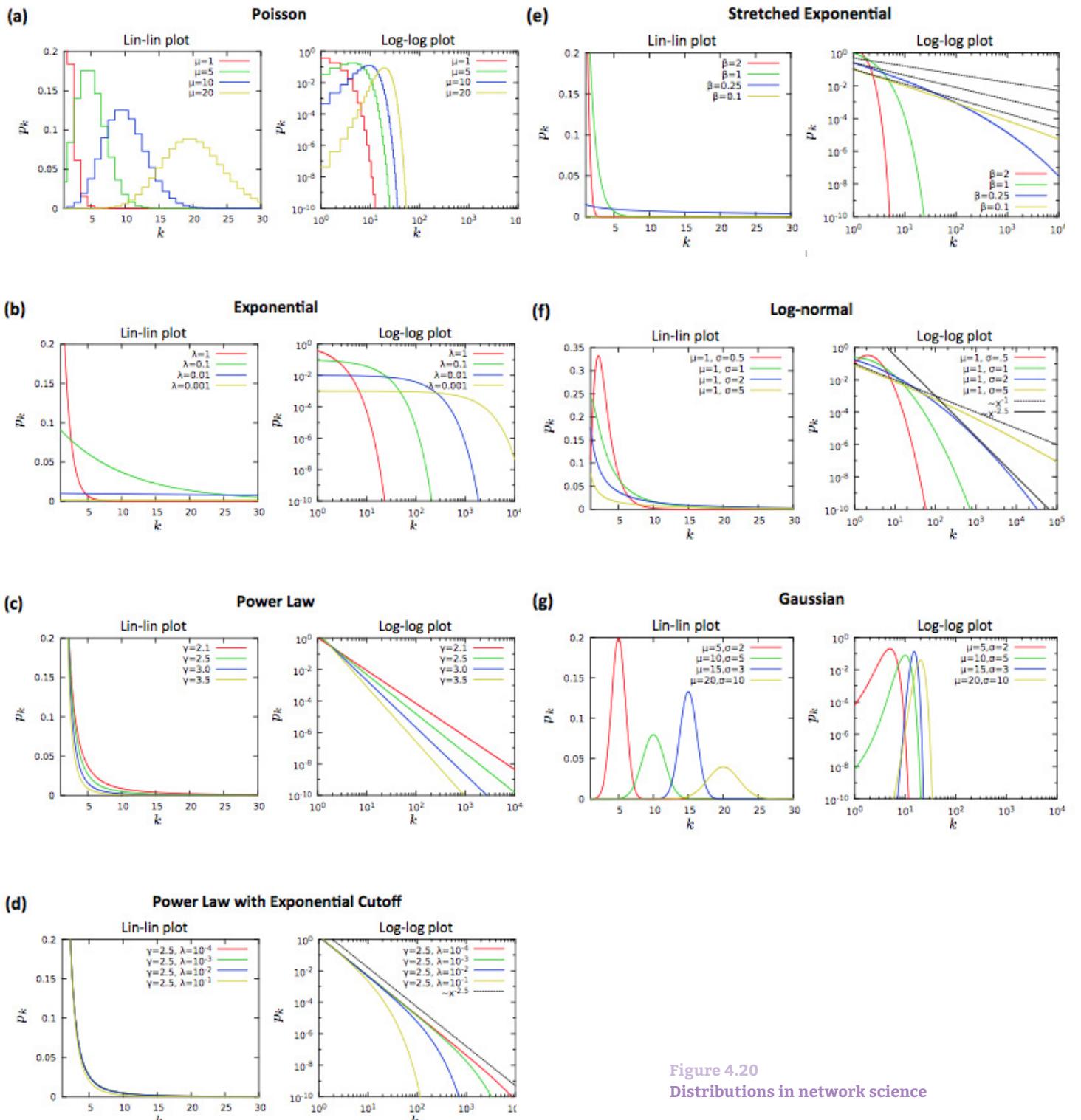


Figure 4.20
Distributions in network science

Each figure shows the linear and the log-log plot for the most frequently encountered distributions in network science. For definitions see Table 4.2.

ADVANCED TOPICS 4.B

PLOTTING A POWER-LAW DEGREE DISTRIBUTION

Plotting the degree distribution is an integral part of analyzing the properties of a network. The process starts with obtaining N_k , the number of nodes with degree k . This can be provided by direct measurement or by a model. From N_k we determine $p_k = N_k / N$. The question is, how to plot p_k to best extract its properties.

USE A LOG-LOG PLOT

In a scale-free network numerous nodes with one or two links coexist with a few hubs, representing nodes with thousands or even millions of links. Using a linear k -axis will compress the numerous small degree nodes in the small- k region, rendering them invisible. Similarly, as there are orders of magnitude differences in p_k for $k=1$ and for some large k , if we plot p_k on a linear vertical axis, its value for large k will appear to be zero (see Fig. 4.21). The use of a log-log plot avoids these problems. We can either use logarithmic axes, with powers of 10 (used throughout this book) or we can plot $\log p_k$ in function of $\log k$ (equally correct, but slightly harder to read). Note that points with $N_k=0$ or ($p_k=0$) are not shown on a log-log plot as $\log 0=-\infty$.

AVOID LINEAR BINNING

The most flawed method (yet frequently seen in the literature) is to simply plot $p_k = N_k/N$ on a log-log plot Fig. 4.21b. This is called linear binning, as each bin has the same size $\Delta k = 1$. For a scale-free network linear binning results in an instantly recognizable plateau at large k , consisting of numerous data points that form a horizontal line Fig. 4.21b. This plateau has a simple explanation: as typically we have only one copy of each high degree node, for high k we either have $N_k=0$ (no node with degree k) or $N_k=1$ (a single node with degree k). Consequently linear binning will either give $p_k=0$, not visible on a log-log plot, or $p_k = 1/N$, which effectively applies to all hubs, generating a plateau at $p_k = 1/N$. This plateau affects our ability to estimate the degree exponent γ . For example, if we attempt to fit a power law to the data shown in Fig. 4.21b using linear binning, the fit provides γ that is quite different from real value $\gamma=2.5$. The reason is that under linear binning we have a large number of nodes in small k bins, hence in this regime we can confidently fit p_k . We have too few nodes in the large k bins for

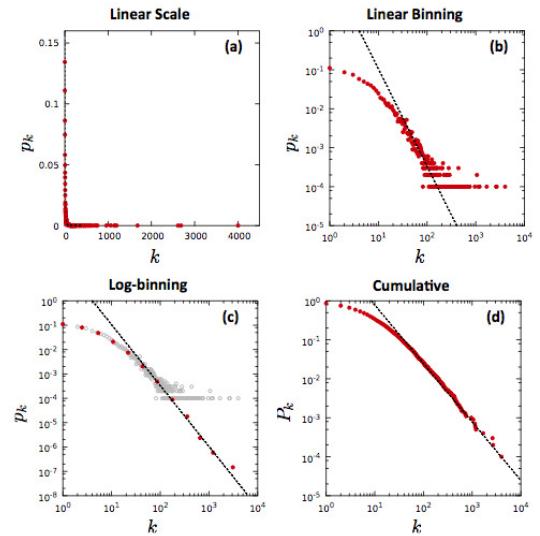


Figure 4.21
Plotting degree distributions

(a-d) The degree distribution of the form $p_k \sim (k + k_0)^{-\gamma}$, with $k_0=10$ and $\gamma=2.5$, plotted using the three procedures described in the text:

(a) linear binning. It is impossible to see the distribution on a lin-lin scale. This is the reason why we always use log-log plots for scale-free networks.

(b-d): The degree distribution shown on a log-log plots using (b) linear binning, (c) logarithmic binning, and (d) plotting the cumulative distribution.

a proper statistical estimate of p_k , hence the plateau biases our fit. Yet, it is precisely this high- k regime that plays a key role in determining γ . Increasing the bin size will not solve this problem. It is therefore recommended to avoid linear binning for fat tailed distributions.

USE LOGARITHMIC BINNING

Logarithmic binning aims to correct for the non-uniform sampling observed for linear binning. For log-binning we let the bin sizes increase with the degree, making sure that each bin has a comparable number of nodes. For example, we can choose the bin sizes to be multiples of 2, so that the first bin has size $b_0=1$, containing all nodes with $k=1$; the second has size $b_1=2$, containing nodes with degrees $k=2, 3$; the third bin has size $b_2=4$ containing nodes with degrees $k=4, 5, 6, 7$. In general, the n^{th} bin has size 2^{n-1} and contains all nodes with degrees $k=2^{n-1}, 2^{n-1}+1, \dots, 2^{n-1}-1$. Note that the bin size can increase with arbitrary increments, $b_n = c_n$, where $c > 1$. The degree distribution is given by $p_{\langle k_n \rangle} = N_n / b_n$, where N_n is the number of nodes found in the bin n of size b_n , and $\langle k_n \rangle$ is the average degree of the nodes in bin b_n . The logarithmically binned p_k is shown in [Fig. 4.21c](#). Note that now the scaling extends into the high- k plateau, previously invisible under linear binning. This indicates that logarithmic binning extracts useful information from the high degree nodes as well [BOX 4.8](#).

USE CUMULATIVE DISTRIBUTION

Another way to extract information from the tail of p_k is to plot the cumulative distribution

$$P_x = \sum_{q=k}^{\infty} P_q, \quad (4.37)$$

which again enhances the statistical significance the high-degree region. If p_k follows the power law, then the cumulative distribution will scale as

$$P_x \sim k^{-\gamma+1}. \quad (4.38)$$

The cumulative distribution will again eliminate the plateau observed for linear binning and leads to an extended scaling region [Figure 4.21d](#), allowing for a more accurate estimate of the degree exponent.

In summary, plotting the degree distribution to fully extract its features requires special attention. Mastering the tools of the process can help us better explore the properties of real networks [BOX 4.9](#).

BOX 4.8

The impact of log-binning

To illustrate the rationale for log-binning, we compare three binning strategies: linear binning, log-binning, and variable bins, when the bin lengths were chosen such that each bin contains exactly the same number of events. As the figure shows, for logarithmic binning the bin sizes decrease exponentially with the bin number.

Indeed, choosing the bin sizes to vary between 2^{n-1} and 2^n , we obtain that the number of events in each bin decreases as $2^{-(\gamma-1)n}$. Yet, the bin size in case of linear binning decreases even faster, effectively running out of events.

The impact of log-binning is most visible in (b) where we show the obtained degree distribution. As one can see, both the variable binning and the linear binning considerably limits the scaling regime compared to the log-binning strategy.

Note that to compare the three methods we set the total number of bins to 10 in all cases.

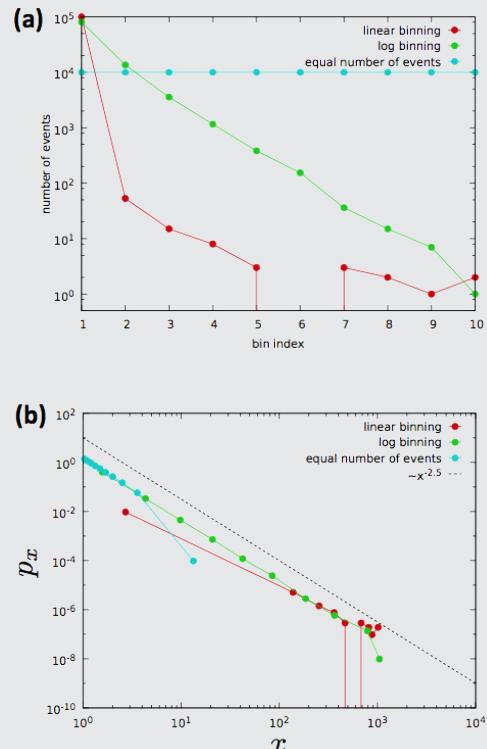


Figure 4.22

BOX 4.9

The degree distribution of real networks

In real systems we rarely observe a degree distribution that follows a pure power law. Instead, for most real systems p_k has the shape shown schematically in (a), with some recognizable features:

- Low-degree saturation is a common deviation from the power law behavior. Its signature is a flattened p_k for $k < k_{sat}$. This indicates that we have fewer small degree nodes than expected for a pure power law. The origin of the saturation will be explained in **CHAPTER 6**.
- High-degree cutoff appears as a rapid drop in p_k for $k > k_{cut}$, indicating that we have fewer high-degree nodes than expected in a pure power law. This also limits the size of the largest hub, making it smaller than predicted by [Eq. 4.23](#). High-degree cutoffs emerge if there are inherent limitations in the number of links a node can have. For example, in social networks individuals have difficulty maintaining a meaningful relationship with an exceptionally large number of acquaintances.

Given the widespread presence of such cutoffs we often fit the degree distribution to

$$p_x = a(k + k_{sat})^{-\gamma} \exp -\frac{k}{k_{cut}} . \quad (4.39)$$

where k_{sat} accounts for the degree saturation, and the exponential term accounts for the high- k cutoff. To extract the full extent of the scaling we plot

$$p_x = p_x \exp \frac{k}{k_{cut}} \quad (4.40)$$

in function of $\tilde{k} = k + k_{min}$. According to [Eq. 4.40](#) $\tilde{p} \sim \tilde{k}^{-\gamma}$, correcting for the two cutoffs, as shown in (b). One occasionally encounters the claim that the presence of low-degree or high-degree cutoffs implies that the network is not scale-free. This is a misunderstanding of the scale-free property: most properties of scale-free networks are insensitive to the low-degree saturation. Only the high-degree cutoff affects the system's properties by limiting the divergence of the second moment $\langle k^2 \rangle$. The presence of such cutoffs means that additional phenomena take place in the system, that need to be understood.

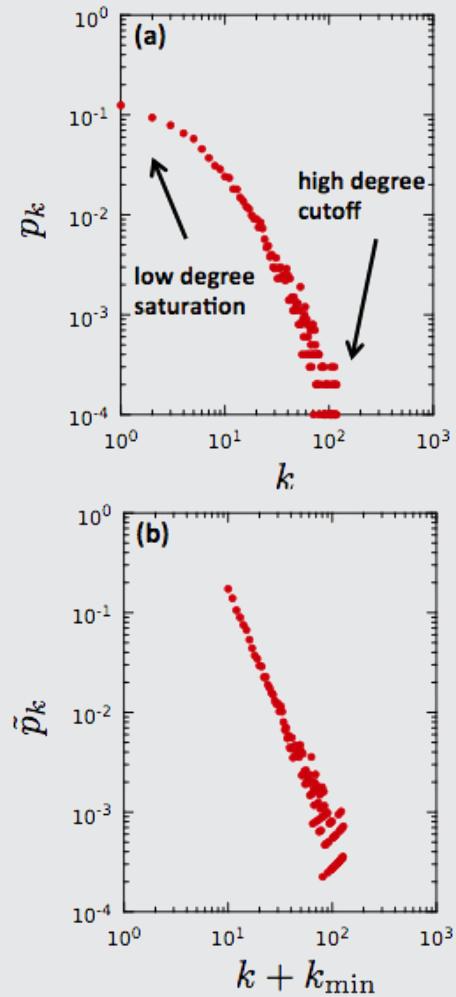


Figure 4.23
Rescaling the degree distribution

(a) The frequently observed form of a degree distribution in real data, characterized by low and high degree cutoffs.

(b) By plotting the rescaled \tilde{p}_k in function of $(k + k_{min})$, as suggested by [Eq. 4.39](#), the degree distribution follows a power law for all degrees.

ADVANCED TOPICS 4.C

ESTIMATING THE DEGREE EXPONENT

As discussed in SECTION 4.7, the properties of scale-free networks depend on γ , raising the need to accurately estimate the degree exponent γ . We face several difficulties, however, when we try to fit a power law to real data. The most important one is the fact that the scaling is rarely valid for the full range of the degree distribution.

Rather we observe so called small- and high-degree cutoffs BOX 4.9, denoted by k_{min} and k_{max} , within which we can min max observe a clear scaling region. Note that k_{min} and k_{max} are different from K_{min} and K_{max} , which correspond to the smallest and largest degrees in a network. Here we focus on estimating the small degree cutoff K_{min} , as the high degree cutoff can be approximated in a similar fashion. Before implementing this procedure, the reader is advised to consult the discussion on systematic problems provided at the end of this section.

FITTING PROCEDURE

As the degree distribution typically comes as a list of positive integers $k=0, 1, 2, \dots, k_{max}$, we aim to estimate γ from a discrete set of data points. We follow [44] and the algorithmic tools to perform the fits are available at <http://tuvalu.santafe.edu/~aarong/powerlaws/>. We use the degree distribution of citation networks to illustrate the procedure. The network consists of $N=384,362$ nodes, each representing a research paper published between 1890 and 2009 in the family of journal published by the American Physical Society. The network has $L=2,353,984$ links, each representing a citation from a published research paper to some other publication in the dataset (outside citations are ignored). See [45] for an overall characterization of the full dataset Figure 4.24a. The steps of the fitting process are:

1. Pick a value of k_{min} between k_{min} and k_{max} . Estimate the value of the degree exponent corresponding to this k_{min} using

$$\gamma = 1 + N^{-1} \sum_{i=1}^N \ln \frac{k_i}{K_{min} - \frac{1}{2}} \quad (4.41)$$

2. With the obtained (γ, k_{min}) parameter pair assume that the degree distribution has the form

$$p(k) = \frac{1}{\zeta(\gamma, K_{min})} k^{-\gamma}, \quad (4.42)$$

hence the associated cumulative distribution function (CDF) is

$$P(k) = 1 - \frac{\zeta(\gamma, k)}{\zeta(\gamma, K_{min})}. \quad (4.43)$$

3. Use the Kormogorov-Smirnov test to determine the maximum distance D between the CDF of the data $S(k)$ and the fitted model provided by Eq. 4.43 with the selected (γ, k_{min}) parameter pair,

$$D = \max_{k \geq K_{min}} |S(k) - P(k)|. \quad (4.44)$$

Eq. 4.44 identifies the degree for which the difference D between the empirical distribution $S(k)$ and the fitted distribution Eq. 4.43 is the largest.

4. Repeat steps (1-3) by scanning the whole k_{min} range from k_{min} to k_{max} . We aim to identify the k_{min} value for which D provided by Eq. 4.44 is minimal. To illustrate the procedure, we plot D in function of k_{max} for the citation network Fig. 4.24b. The plot indicates that D is minimal for $k_{min}=49$, and the corresponding γ estimated by Eq. 4.41, representing the optimal fit, is $\gamma=2.79$. The standard error for the obtained degree exponent is

$$\sigma_\gamma = \sqrt{\frac{1}{N} \left(\frac{\zeta(\gamma, K_{min})}{\zeta(\gamma, K_{min})} - \frac{\zeta(\gamma, K_{min})}{\zeta(\gamma, K_{min})}^2 \right)} \quad (4.45)$$

which implies that the best fit is for exponent $\gamma \pm \sigma_\gamma$. For the citation network we obtain $\sigma_\gamma=0.003$, hence $\gamma=2.79$ (3).

Note that Eq. 4.45 represents an approximation, but typically the results provided by it is within 1% of the real value as long as $k_{min}>6$. Furthermore, in order to obtain a reasonable estimate for γ , we need $N>50$. Smaller datasets should be treated with caution.

GOODNESS-OF-FIT

Just because we obtained a (γ, k_{min}) pair that represents an optimal fit to our dataset, does not mean that the power law itself is a good model for the studied distribution. We therefore need to use a goodness-of-fit test, which generates a p -value that quantifies the plausibility of the power law hypothesis. The most often used procedure [12] consists of the following steps:

- (i) Use the cumulative distribution Eq. 4.43 to estimate the KS distance between the real data and the best fit, that we denote by D^{real} . This is step 3 above, taking the value of D for k_{min} that offered the best fit

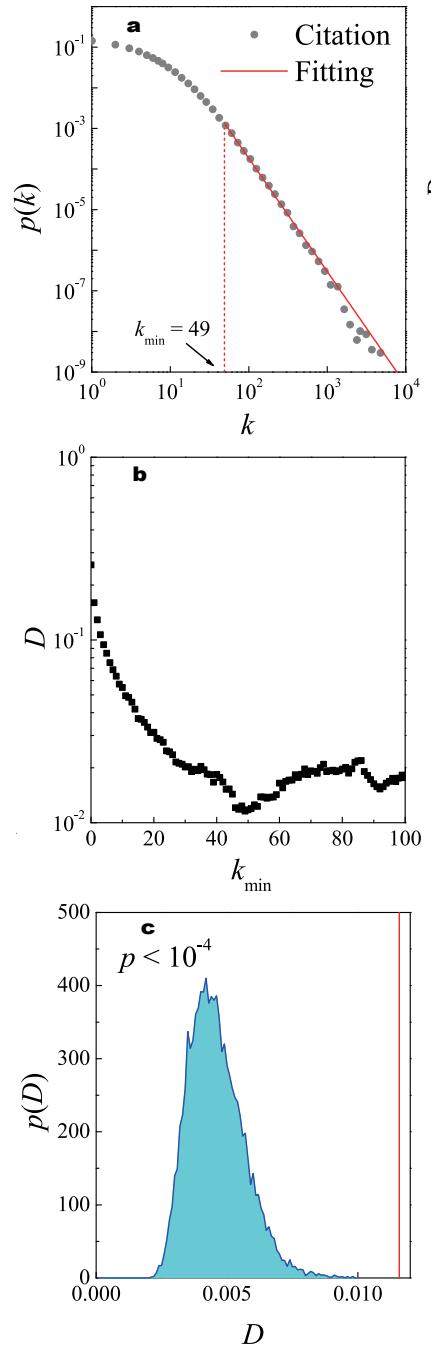


Figure 4.24
Maximum likelihood estimation

(a) The degree distribution p_k of the citation network, where the straight line represents the best based on model Eq. 4.39.

(b) The values of Kormogorov-Smirnov test vs. k_{min} , where the red lines indicate the minimum value of D and the corresponding k_{min} .

(c) $p(D^{synthetic})$ for $M=10,000$ synthetic data, where the red line corresponds to the D value from the citation network (a-b).

to the data. For the citation data we obtain $D^{\text{real}} = 0.01158$ for $k_{\min} = 49$ Fig. 4.24.

(ii) Use Eq. 4.42 to generate a degree sequence of N degrees (i.e. the same number of random numbers as the number of nodes in the original dataset) and substitute the obtained degree sequence for the empirical data, determining $D^{\text{synthetic}}$ for this hypothetical degree sequence. Hence $D^{\text{synthetic}}$ represents the distance between a synthetically generated degree sequence, consistent with our degree distribution, and the real data.

iii. The goal is to see if the obtained $D^{\text{synthetic}}$ is comparable to D^{real} . For this we repeat step (ii) M times ($M \gg 1$), and each time we generate a new degree sequence and determine the corresponding $D^{\text{synthetic}}$, eventually obtaining the $p_{D^{\text{synthetic}}}$ distribution. Plot $p_{D^{\text{synthetic}}}$ and show as a vertical bar D^{real} Fig. 4.24c. If D^{real} is within the $p_{D^{\text{synthetic}}}$ distribution, it means that the distance between the model providing the best fit and the empirical data is comparable with the distance expected from random degree samples chosen from the best fit distribution. Hence the power law is a reasonable model to the data. If, however, D^{real} falls outside the $(p_{D^{\text{synthetic}}})$ distribution, then the power law is not a good model - some other function is expected to describe the original p_k .

While the distribution shown in Figure 4.20 may be in some cases useful to offer a visual illustration, in general is better to assign a p -number to the fit, given by

$$p = \int_D^\infty P_{D^{\text{synthetic}}} dD^{\text{synthetic}}. \quad (4.46)$$

The closer p is to 1, the more likely that the difference between the empirical data and the model can be attributed to statistical fluctuations alone; if p is small, the model is not a plausible fit to the data.

Typically, the model is accepted if $p > 1\%$. For the citation network we obtain $p < 10^{-4}$, indicating that a pure power law is not a suitable model for the original degree distribution. This outcome is somewhat surprising, as the power-law nature of citation data has been documented repeatedly since 1960s [7, 8]. This failure offers a lesson on the limitation of the blind application of the fitting procedures.

FITTING REAL DISTRIBUTIONS

To correct the problem, we note that the fitting model Eq. 4.44 eliminates all the data points with $k < k_{\min}$. As the citation network is fat tailed, choosing $k_{\min} = 49$ forces us to discard over 96% data points. Yet, there is statistically useful information in the $k < k_{\min}$ regime, that is ignored by the previous fit. We therefore introduce an alternate model that resolves this problem.

As we discussed in **BOX 4.9**, the degree distribution of many real networks, like the citation network, can not be described by a pure power law, but has the form

$$p_k = \frac{1}{\sum_{k=1}^{\infty} (k + k_{sat})^{-\gamma} e^{-k/k_{cut}}} (k + k_{sat})^{-\gamma} e^{-k/k_{cut}} \quad (4.47)$$

and the associated CDF is

$$P_k = \sum_{k=1}^{\infty} (k + k_{sat})^{-\gamma} e^{-k/k_{cut}}, \quad (4.48)$$

where k_{sat} and k_{cut} correspond to low- k saturation and the large- k cutoff, respectively. The difference between our earlier procedure and [Eq. 4.47](#) is that we now do not discard the points that deviate from a pure power law, but we use a function that may offer a better fit to the whole degree distribution, from k_{min} to k_{max} .

Our goal is to find the fitting parameters k_{sat} , k_{cut} , and γ of the model [Eq. 4.47](#), which we achieve through the following steps:

- A. Pick a value for k_{sat} and k_{cut} between k_{min} and k_{max} . Estimate the value of the degree exponent γ using the steepest descend method that maximizes the log-likelihood function

$$\log \mathcal{L}(\gamma | k_{min}, k_{cut}) = \sum_{i=1}^N \log p(k_i | \gamma, k_{min}, k_{cut}). \quad (4.49)$$

That is, for fixed (k_{sat}, k_{cut}) we vary γ until we find the maximum of L . The steepest descent method provides γ (k_{min} , k_{cut}) for which [Eq. 4.48](#) is maximal.

- B. With the obtained $\gamma(k_{sat}, k_{cut})$ assume that the degree distribution has the form. Calculate the Kormogorov Smirnov parameter D [Eq. 4.47](#) between the cumulative degree distribution (CDF) of the original data and the fitted model provided by [Eq. 4.47](#).

- C. Change k_{sat} and k_{cut} , and repeat steps (1-3), scanning with k_o from $k_{min}=0$ to k_{max} and with k_{cut} from $k_{min}=k_o$ to k_{max} . The goal is to identify k and k values for which D is minimal. We illustrate this by plotting D in function of k_{sat} for serval k_{cut} values in [Fig. 4.25a](#) for our citation sample. The (k_{sat}, k_{cut}) for which D is minimal, and the corresponding γ is provided by [Eq. 4.41](#), will represent the optimal parameters of the fit. For our dataset the optimal fit is obtained for $k_{sat}=12$ and $k_{cut}=5691$, providing the degree exponent $\gamma = 3.028$. We find that now D for the real data is within the generated $p(D)$ distribution [Fig.4.25c](#), and the associated p -value is 69%.

SYSTEMATIC FITTING ISSUES

The procedure described above may offer the impression that determining the degree exponent is a cumbersome but straight forward process. In reality the existing fitting methods have some well known limitations:

1. A pure power law is really an idealized distribution that emerges in its form (1) only in simple models [CHAPTER 5](#). In reality, a whole range of processes contribute to the topology of real networks, affecting the precise shape of the degree distribution. These processes will be described in [CHAPTER 6](#). If p_k does not follow a pure power law, the methods described above, designed to fit a power law to the data, will inevitably fail to detect statistical significance. That does not necessarily mean that the network is not scale-free (but it could also mean that). Most often it means that we have not yet gained a proper understanding of the precise form of the degree distribution, hence we are fitting the wrong functional form of p_k to the dataset.

2. The statistical tools used above to test the goodness of the fit rely on the Kolmogorov-Smirnov criteria, which measures the maximum distance between the fitted model and the dataset. If all data points follow a perfect power law, but a single point for some reason deviates from the curve, we will lose the fit's statistical significance. In real systems there are numerous reasons for such local deviations, that have little impact on the system's overall behavior. Yet, removing these "outliers" could be seen as data manipulation; if kept, however, one cannot detect the statistical significance of the power law fit. A good example is provided by the actor network, whose degree distribution follows a power law for most degrees. There is a single outlier, at $k = 1,287$, thanks to the 1956 movie, *Around the World in Eighty Days*.

This is the only movie, where IMDB lists all the uncredited extras in the cast. Hence the movie appears to have 1,288 actors. The second largest movie in the dataset has only 340 actors. Since the extras are only listed for this movie, each of them have links only to the 1,287 extras that played in the same movie, leading to a local peak in p_k at $k=1,287$. Thanks to this peak, the degree distribution, fitted to a power law fails to pass the Kolmogorov-Smirnov criteria. Indeed, as indicated in [Table 4.3](#), neither the pure power law fit, nor a power law with high-degree cutoff offers a statistically significant fit.

3. Thanks to the issues discussed above, the methodology described above often predicts a small scaling regime, forcing us to remove a huge fraction of the nodes (often as many as 99%, see [Table 4.4](#)), to obtain a statistically significant fit. Once plotted next to the original dataset, the obtained fit can be at times ridiculous, even if the method indicates statistical significance. The bottom line, estimating the degree exponents is still not an exact science. We continue to lack methods that would estimate the statistical significance of a proper fit in a manner that would be acceptable to a practitioner. The blind application of the tools describe above often leads to either fits that obviously do not capture the trends in the data, or to a false rejection of the power-law hypothesis. An important improvement will be provided by our ability to derive the expected form of the degree distribution, discussed in [CHAPTER 6](#).

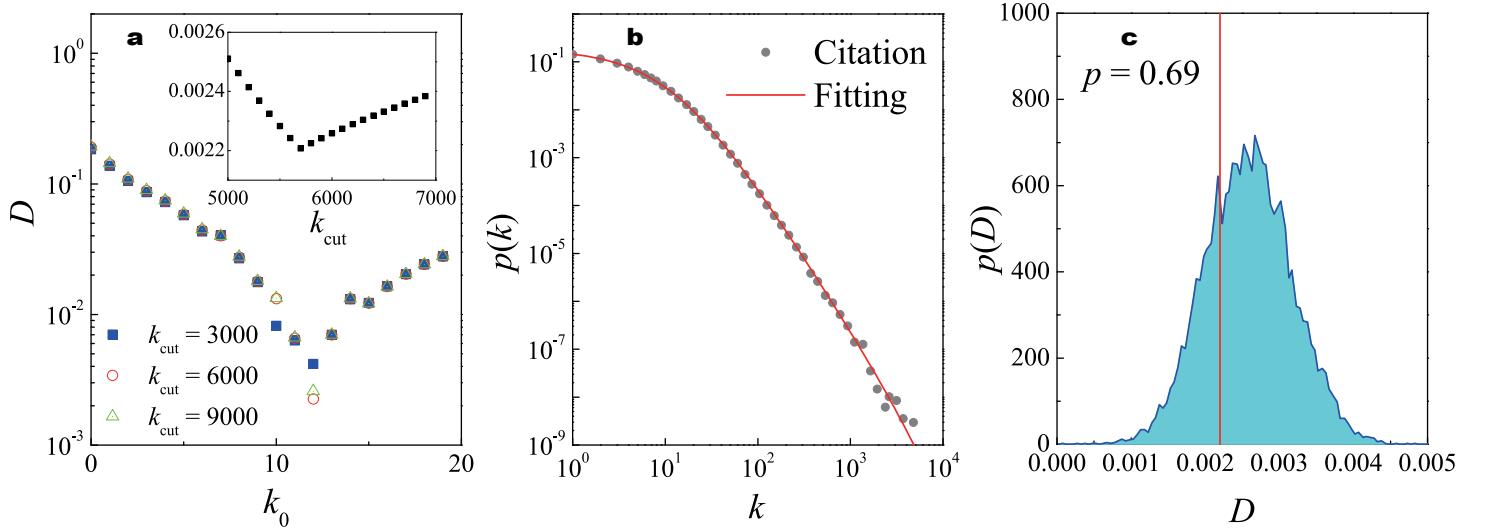


Figure 4.25

Estimating the scaling parameters for citation networks

(a) The Kormogorov-Smirnov parameter D vs. k_0 for $k_{cut} = 3.000, 6.000, 9.000$, respectively, showing that $k_{sat} = 12$ corresponds to the minimal D . Inset: D vs. k_{cut} for $k_{sat} = 12$, indicating that $k_{cut} = 5.691$ minimizes D .

(b) Degree distribution p_k where the straight line represents the best fitting estimated from (a).

(c) $pD^{\text{synthetic}}$ for $M = 10.000$ synthetic data, where the red line corresponds to the D value from the citation network (a-b).

NETWORK NAME	λ	k_{\min}	P-VALUE	PERCENTAGE
Power Grid	0.5174		0.91	12%

Table 4.3
Exponential Fitting

For the power grid a power law does not offer a statistically significant fit as the underlying network is not scale-free. We used the fitting procedure described in this section to now fit the exponential function $e^{-\lambda k}$ to the degree distribution of the power grid, obtaining a statistically significant fit in this case. The table shows the obtained λ parameters, the k_{\min} over which the fit is valid, the obtained p -value, and the percentage of data points included in the fit.

NETWORK NAME $k^{-\gamma}; [k_{\min}, \infty]$					$(k + k_{\text{sat}})^{-\gamma} e^{-k/k_{\text{cut}}}$			
	γ	k_{\min}	P-VALUE	PERCENT	γ	k_{sat}	k_{cut}	P-VALUE
Internet	3.42	72	0.13	0.6%	3.55	88	500	0.00
WWW-ND (in)2	.001		0.00	100%	1.970		660	0.00
WWW-ND (out)2	.317		0.00	15%2	.828		8500	0.00
Power Grid	4.00	50	.001	2%	8.561	91	40	.00
Mobile Phone Calls (in)	4.69	90	.342	.6%6	.951	51	00	.00
Mobile Phone Calls (out)	5.01	11	0.77	1.7%	7.23	15	10	0.00
Email-PRE (in)3	.438	80	.11	0.2%	2.27	08	500	0.00
Email-PRE (out)2	.033		0.00	1.2%2	.550		8500	0.00
Science Collaboration3	.352	50	.0001	5.4%	1.501	71	20	.00
Actor Network2	.12	54	0.00	33%	--		-0	.00
Citation Network (in)2	.79	51	0.00	3.0%	3.03	12	5691	0.69
Citation Network (out)	4.00	19	0.00	14%-	0.16	51	00	.00
E.coli Metabolism (in)	2.43	30	.00	57%	3.851	91	20	.00
E.coli Metabolism (out)	2.90	50	.00	34%	2.56	15	10	0.00
Yeast Protein Interactions	2.897		0.67	8.3%2	.952		90	0.52
WWW-stanford (in)2	.15	30	.00	44.9%	2.86492	4	222	0.00
WWW-stanford (out)3	.976	20	.000	.6%3	.96102	17	128	0.00
Email-PNAS (in)2	.811	90	.0005	22.2%	0.54	02	50	.00
Email-PNAS (out)2	.272	60	.929	.3%0	.920		36	0.00

Table 4.4
Fitting parameters for real networks

The estimated degree exponents and the appropriate fit parameters for several networks studied in this book. We implemented two fitting strategies, the first aiming to fit a pure power law in the region (k_{\min}, ∞) and the second fits a power law with saturation and exponential cutoff to the whole dataset. In the table we show the obtained γ exponent and k_{\min} for the fit with the best statistical significance, the p -value for the best fit and the percentage of the data included in the fit. In the second case we again show the exponent γ , the two fit parameters, k_{sat} and k_{cut} , and the p -value of the obtained fit. Note that $p > 0.01$ is considered to be statistically significant.

BIBLIOGRAPHY

- [1] H. Jeong, R.Albert, and A.-L. Barabási. Internet: Diameter of the world-wide web. *Nature*, 401:130-131, 1999.
- [2] A.-L. Barabási and R.Albert. Emergence of scaling in random networks. *Science*, 286:509-512, 1999.
- [3] V. Pareto. *Cours d'Économie Politique: Nouvelle édition par G.- H. Bousquet et G. Busino*, Librairie Droz, Geneva, 299–345, 1964.
- [4] A.-L. Barabási. *Linked: The New Science of Networks*. (Plume, New York). ISBN 0-452-28439-2, 2002.
- [5] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *Proceedings of SIGCOMM. Comput. Commun. Rev.* 29: 251-262, 1999.
- [6] R. Pastor-Satorras and A.Vespignani. *Evolution and Structure of the Internet: A Statistical Physics Approach*. (Cambridge University Press, Cambridge), 2004.
- [7] D. J. De Solla Price. Networks of Scientific Papers, *Science* 149: 510-515, 1965.
- [8] S. Redner. How Popular is Your Paper? An Empirical Study of the Citation Distribution, *Eur. Phys. J. B* 4, 131, 1998.
- [9] A. L. Barabási, R.Albert, and H. Jeong. Mean-field theory of scale-free random networks. *Physica A* 272:173-187, 1999.
- [10] R. Kumar, P. Raghavan, S. Rajalopagan, and A.Tomkins. Extracting Large-Scale Knowledge Bases from the Web. *Proceedings of the 25th VLDB-Conference*, Edinburgh, Scotland, pp.639-650, 1999.
- [11] H. Jeong, B.Tombor, R.Albert, Z. N. Oltvai, and A.-L. Barabási. The

large-scale organization of metabolic networks, *Nature* 407, 651-654, 2000.

[12] W. Aiello, F. Chung, and L.A. Lu. Random graph model for massive graphs, *Proc. 32nd ACM Symp.Theor. Comp*, 2000.

[13] H. Jeong, B.Tombor, S. P. Mason,A.-L. Barabási, and Z.N. Oltvai. Lethality and centrality in protein networks, *Nature* 411, 41-42, 2001.

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

[15] M. E. J. Newman. The structure of scientific collaboration networks, *Proc. Natl.Acad. Sci.* 98(2), 404-409, 2001.

[16] F.Liljeros,C.R.Edling,L.A.N.Amaral,H.E.Stanley, and Y.Aberg. The Web of Human Sexual Contacts, *Nature* 411, 907-908, 2001.

[17] R. Ferrer i Cancho and R.V. Solé. The small world of human language, *Proc. R. Soc. Lond. B* 268: 2261-2265, 2001.

[18] S.Yook, H. Jeong, and A.-L. Barabási. (unpublished), (2001).

[19] R. Ferrer i Cancho, C. Janssen, and R.V. Solé. Topology of technology graphs: Small world patterns in electronic circuits, *Phys. Rev. E* 64, 046119, 2001.

[20] S.Valverde and R.V. Sole. Hierarchical Small Worlds in Software Architecture, *arXiv:cond-mat/0307278*, 2003.

[21] H. Ebel, L.-I. Mielsch, and S. Bornholdt. Scale-free topology of e-mail networks, *Phys. Rev. E* 66, 035103(R), 2002.

[22] J. P. K. Doye. Network Topology of a Potential Energy Landscape: A Static Scale-Free Network, *Phys. Rev. Lett.* 88, 238701, 2002.

[23] H. Kwak, C. Lee, H. Park, S. Moon. What is Twitter, a social network or a news media?, Proceedings of the 19th international conference on World wide web, 591-600, 2010.

[24] M. Cha, H. Haddadi, F. Benevenuto and K. P. Gummadi. Measuring user influence in Twitter: The million follower fallacy, Proceedings of international AAAI Conference on Weblogs and Social, 2010.

[25] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow. The Anatomy of the Facebook Social Graph, *arXiv:1111.4503*, 2011.

[26] B. Bollobás and O. Riordan. The Diameter of a Scale-Free Random Graph. *Combinatorica*, 24: 5-34, 2004.

[27] R. Cohen and S. Havlin. Scale free networks are ultrasmall, *Phys.*

[28] R. Cohen and S. Havlin. Complex Networks - Structure, Robustness and Function. (Cambridge University Press, Cambridge), 2010.

[29] K.-I. Goh, B. Kahng, and D. Kim. Universal behavior of load distribution in scale-free networks. Phys. Rev. Lett. 87, 278701, 2001.

[30] F. Karinthy. Láncszemek, in Minden másképpen van (Budapest: Atheneum Irodai es Nyomdai R-T. Kiadása, 1929), 85–90. English translation is available in: M.E. J. Newman, A.-L. Barabási, and D. J. Watts, The Structure and Dynamics of Networks (Princeton University Press, Princeton), 2006.

[31] P. S. Dodds, R. Muhamad and D.J.Watts. An experimental study to search in global social networks, Science 301, 827-829, 2003.

[32] P. Erdős and T. Gallai. Graphs with given degrees of vertices. Matematikai Lapok, 11:264-274, 1960.

[33] C.I. Del Genio, H. Kim, Z.Toroczkai, and K.E. Bassler. Efficient and exact sampling of simple graphs with given arbitrary degree sequence. PLoS ONE, 5(4):e10012, 04 2010.

[34] V. Havel. A remark on the existence of finite graphs. Casopis Pest. Mat., 80:477-480, 1955.

[35] S. Hakimi. On the realizability of a set of integers as degrees of the vertices of a graph. SIAM J. Appl. Math., 10:496-506, 1962.

[36] I. Charo Del Genio, G.Thilo, and Kevin E. Bassler. All scale-free networks are sparse. Phys. Rev. Lett., 107:178701, 10 2011.

[37] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, European J. Combin. 1: 311– 316, 1980.

[38] M. Molloy and B. A. Reed. Critical Point for Random Graphs with a Given Degree Sequence. Random Structures and Algorithms, 6: 161-180, 1995.

[39] M. Newman. Networks: An Introduction. (Oxford University, Oxford), 2010.

[40] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. Science, 296:910-913, 2002.

[41] G. Caldarelli, I. A. Capocci, P. De Los Rios, and M.A. Muñoz. Scale-Free Networks from Varying Vertex Intrinsic Fitness, Phys. Rev. Lett. 89, 258702, 2002.

[42] B. Söderberg. General formalism for inhomogeneous random

graphs, Phys. Rev. E 66, 066121, 2002.

[43] M. Boguñá and R. Pastor-Satorras. Class of correlated random networks with hidden variables, Phys. Rev. E 68, 036112, 2003.

[44] A.Clauset,C.R.Shalizi, and M.E.J.Newman.Power-law distributions in empirical data. SIAM Review S1: 661-703. ArXiv e- prints, 2009.

[45] S. Redner. Citation statistics from 110 years of physical review. Physics Today, 58:49, 2005.

CHAPTER 5

THE BARABÁSI-ALBERT MODEL

Introduction

Growth and preferential attachment 1

The Barabási-Albert Model 2

Degree dynamics 3

Degree distribution 4

The absence of growth or preferential attachment 5

Non-linear preferential attachment 6

The origins of preferential attachment 7

Diameter and clustering coefficient 8

Homework

Summary

ADVANCED TOPICS 5.A

Deriving the degree distribution

ADVANCED TOPICS 5.B

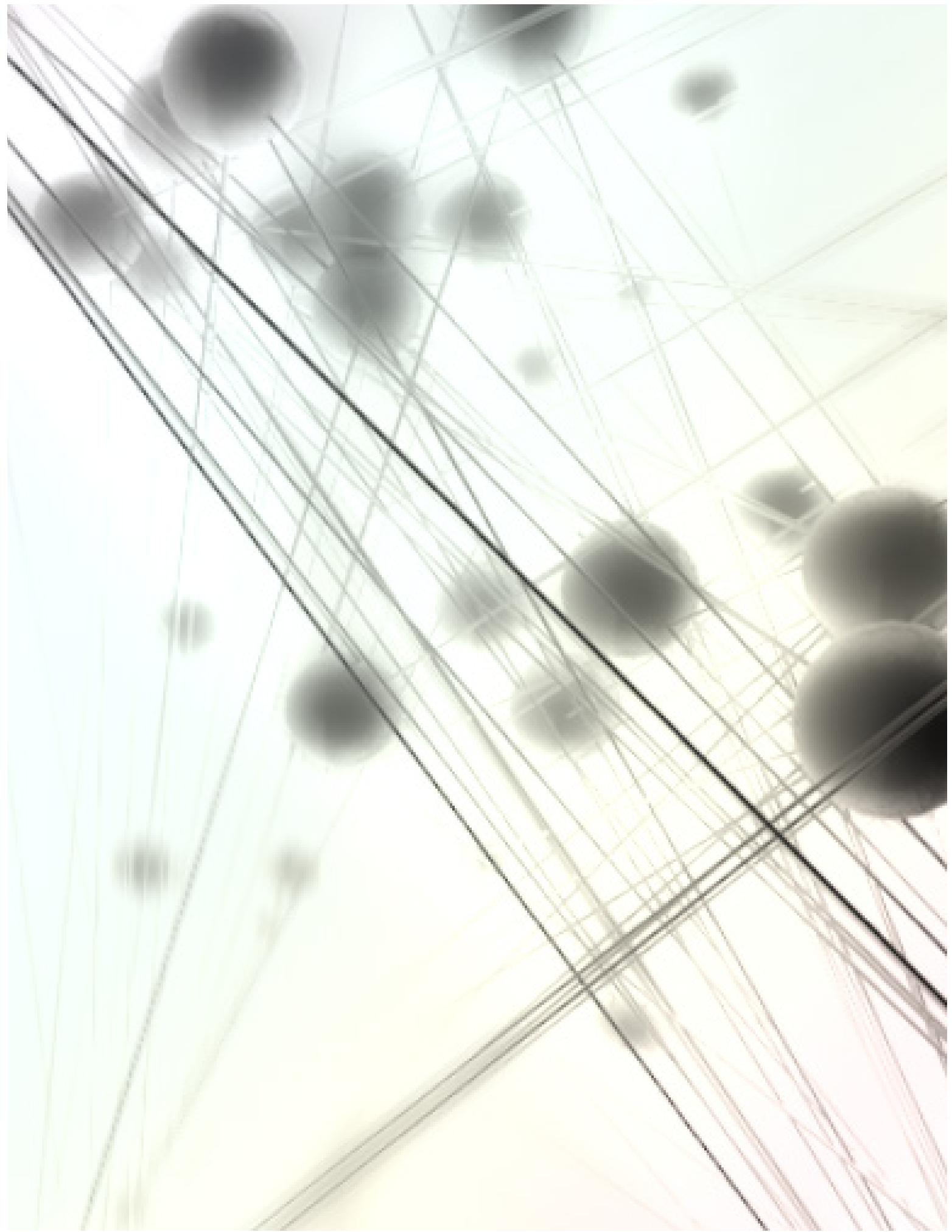
Nonlinear preferential attachment

ADVANCED TOPICS 5.C

The clustering coefficient

Bibliography

Figure 5.0 (front cover)
Universality: G. Musella



INTRODUCTION

Hubs represent the most striking difference between a random and a scale-free network. Their very existence raises several fundamental questions:

- Why does the random network model of Erdős and Rényi fail to reproduce the hubs and the power laws observed in many real networks?
- Why do so different systems as the WWW or the cell converge to a similar scale-free architecture?

The last question is particularly puzzling given the fundamental differences in the nature, origin, and scope of the systems that display the scale-free property:

- The nodes of the cellular network are proteins or metabolites, while the nodes of the WWW are documents, representing information without a physical manifestation.
- The links within the cell are binding interactions and chemical reactions, while the links of the WWW are URLs, or small segments of computer code.
- The history of these two systems could not be more different: the cellular network is shaped by 4 billion years of evolution, while the WWW is a few decades old.
- The purpose of the metabolic network is to chemically build the basic chemical components the cells need for life, while the purpose of the WWW is information access and delivery.

To understand why so different systems converge to a similar architecture we need to first understand the mechanism responsible for the emergence of the scale-free property. This is the main topic of this chapter. Giv-

en the major differences between the systems that display the scale-free property, the explanation must be simple and fundamental. The answers will change the way we view and model networks, forcing us to move from describing a network's topology to modeling the evolution of complex systems.

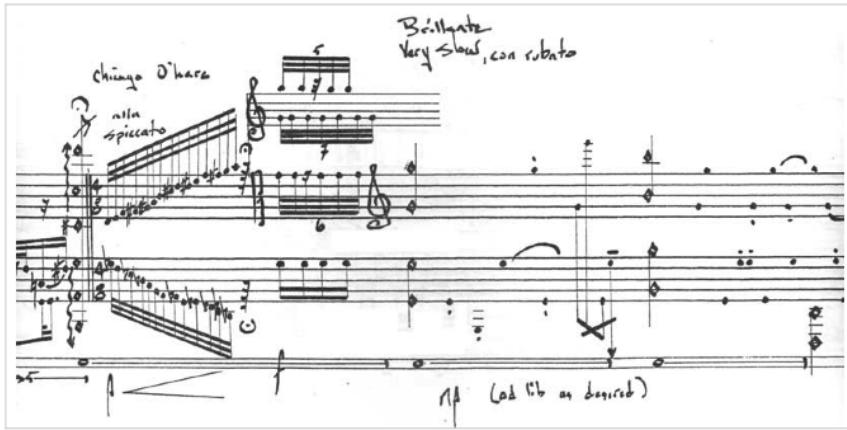


Figure 5.1
Scale-free sonata

Composed by Michael Edward Edgerton in 2003, *1 sonata for piano* was inspired by scale-free networks. The music obeys the principles of a growing network, incorporating growth and preferential attachment. The interplay between the music and networks is explained by the composer:

“6 hubs of different length and procedure were distributed over the 2nd and 3rd movements. Musically, the notion of an airport was utilized by diverting all traffic into a limited landing space, while the density of procedure and duration were varied considerably between the 6 differing occurrences.”

The Image shows the beginning of what Edgerton calls Hub #5.

GROWTH AND PREFERENTIAL ATTACHMENT

Our journey towards understanding the origin of the scale-free property by asking: why are hubs and power laws absent from the model? The answer emerged in 1999, highlighting two hidden assumptions of the Erdős-Rényi model, each of which are violated in real networks [1]. Next we discuss these two assumptions separately.

NETWORKS EXPAND THROUGH THE ADDITION OF NEW NODES

The random network model assumes that we have a *fixed* number of nodes, N . The role of the modeler is to connect these nodes, while keeping N unchanged. Yet, in most real networks the number of nodes is not fixed, but continually *grows* thanks to the addition of new nodes. Let us consider a few examples:

- In 2001 the WWW had a single node, the first webpage build by Tim Berners-Lee, the creator of the Web. Today the Web has over a trillion (10^{12}) documents, an extraordinary number that was reached through the continuous addition of new documents by millions of individuals and institutions [Fig. 5.2a](#).
- The collaboration and citation networks continually expand through the publication of new research papers [Fig. 5.2b](#).
- The Hollywood actor network continues to expand through the release of new movies [Fig. 5.2c](#).
- At first the protein interaction network within our cells may appear to be static, as we inherit our genes (and hence our proteins) from our parents. Yet, it is not: the number of genes grew from a few to the over 20,000 genes present today in a human cell over four billion years.

Consequently, if we wish to model these networks, we cannot resort to a static model. Rather our approach must acknowledge that networks are the product of a steady growth process.

NODES PREFER TO LINK TO THE MORE CONNECTED NODES

The random network model assumes that we randomly choose the interaction partners of a node. In most real networks, however, new nodes prefer to link to the more connected nodes, a process called *preferential attachment*. Consider a few examples:

- We are familiar with only a tiny fraction of the trillion or more documents available on the WWW. The nodes we know are not entirely random, but we all heard about Google and Facebook, but we rarely encounter the billions of less-prominent nodes that populate the Web. As our knowledge is biased towards the more connected nodes, we are more likely to link to a high-degree node than to a node with only few links.
- With more than a million scientific papers published each year, no scientist can attempt to read them all. The more cited is a paper, the more likely that we will notice it. Therefore, our citations are biased towards the more cited publications, representing the high-degree nodes of the citation network.
- The more movies an actor has played in, the more familiar is a casting director with her skills. Hence, the higher the degree of an actor in the actor network, the higher are the chances that she will be considered for a new role.

In summary, the random network model differs from real networks in two important characteristics:

GROWTH

While the random network model assumes that the number of nodes, N , is fixed (time invariant), real networks are the result of a growth process that continuously increases N .

PREFERENTIAL ATTACHMENT

While nodes in random networks randomly choose their interaction partner, in real networks new nodes prefer to link to the more connected nodes.

There are many other differences between real and random networks, some of which will be discussed in the coming chapters. Yet, as we show next, growth and preferential attachment have a particularly important role shaping a network's degree distribution.

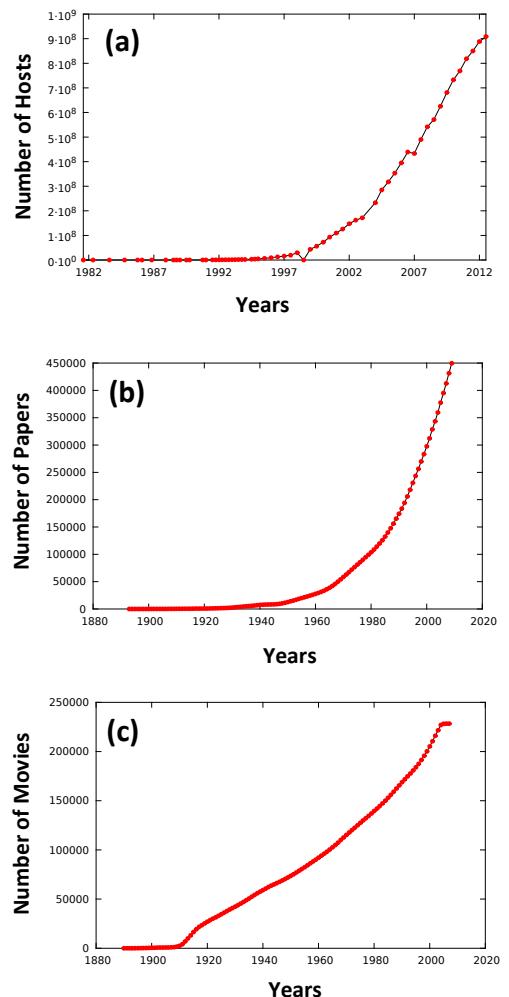


Figure 5.2
The growth of networks

(a) The evolution of the number of WWW hosts, documenting the Web's rapid growth. After <http://www.isc.org/solutions/survey/history>.

(b) The number of scientific papers published in *Physical Review* journals since the journal's funding in 1893. The observed growth drives the growth of both the science collaboration network as well as the citation network. Over the century the *Physical Review* portfolio has split several times, responding to the exponential growth of the number of research papers and to specialization. Today the corpus features *Physical Review Letters*, *Physical Review A, B, C, D, E, X* and *Reviews of Modern Physics*.

(c) Number of movies listed in IMDB.com, reflecting the growth of the Hollywood movie enterprise, and with that the growth of the actor network.

BOX 5.1

PREFERENTIAL ATTACHMENT: A BRIEF HISTORY

Preferential attachment has emerged repeatedly in mathematics and social sciences. Consequently today we can encounter it under different names in the scientific literature:

- It made its first appearance in 1923 in the celebrated urn model of the Hungarian mathematician György Pólya (1887-1985) [2], proposed to explain the nature of certain distributions. Hence, in mathematics preferential attachment is often called a *Pólya process*.
- George Udmy Yule (1871-1951) in 1925 used preferential attachment to explain the power-law distribution of the number of species per genus of flowering plants [3]. Hence, in statistics preferential attachment is often called a *Yule process*.
- Rober Gibrat (1904-1980) in 1931 proposed that the size and the growth rate of a firm are independent. Hence, larger firms grow faster [4]. Called *proportional growth*, this is a form of preferential attachment.
- George Kinsley Zipf (1902-1950) in 1941 used preferential attachment to explain the fat tailed distribution of wealth in the society [5].
- Modern analytical treatments of preferential attachment use of the master equation approach pioneered by the economist Herbert Alexander Simon (1916-2001). Simon used preferential attachment in 1955 to explain the fat-tailed nature of the distributions describing city sizes, word frequencies in a text, or the number of papers published by scientists [6].
- Building on Simon's work, Derek de Solla Price (1922-1983) used preferential attachment to explain the citation statistics of scientific publications, referring to it as *cumulative advantage* [7].
- In sociology preferential attachment is often called the *Matthew effect*, named by Robert Merton (1910-2003) [8] after a passage in the Gospel of Matthew: "For everyone who has will be given more, and he will have an abundance. Whoever does not have, even what he has will be taken from him."
- The term *preferential attachment* was introduced in the 1999 paper by Barabási and Albert [1] to explain the ubiquity of power laws in networks.

Note that the distributions characterized from Pólya to Merton describe scalar quantities, like the number of individuals with the same income or the size of cities. In contrast the Barabási-Albert model aims to describe networks. Networks have a wide array of topological characteristics that are absent from scalar distributions, but which are deeply affected by the power-law nature of the degree distribution.

THE BARABÁSI-ALBERT MODEL

The recognition that growth and preferential attachment coexist in real networks has lead to the introduction of a minimal model capable of generating networks with power-law degree distribution [1]. The model is defined as follows:

We start with m_0 nodes, the links between which are chosen arbitrarily, as long as each node has at least one link. The network develops following two steps Fig. 5.3:

(A) GROWTH

At each timestep we add a new node with m ($\leq m_0$) links that connect the new node to m nodes already in the network.

(B) PREFERENTIAL ATTACHMENT

The probability $\Pi(k_i)$ that one of the links of the new node connects to node i depends on the degree k_i of node i as

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \quad (5.1)$$

Preferential attachment is a probabilistic rule: a new node is free to connect to any node in the network, whether it is a hub or has a single link. Eq. 5.1 implies, however, that if a new node has a choice between a degree-two and a degree-four node, it is twice as likely that it connects to the degree-four node. The model defined by steps (A) and (B) is called the *Barabási-Albert* model after the authors of the paper that introduced it in 1999 [1]. One may also encounter it in the literature as the BA model or the *scale-free* model. After t timesteps the Barabási-Albert model generates a network with $N = t + m_0$ nodes and $m_0 + mt$ links. As Fig. 5.4 shows, the network generated by the model has a power-law degree distribution, with a degree exponent $\gamma=3$.

As Fig. 5.3 indicates, while most nodes in the network have only a few links, a few gradually turn into hubs. The hubs are the result of a *rich-gets-*

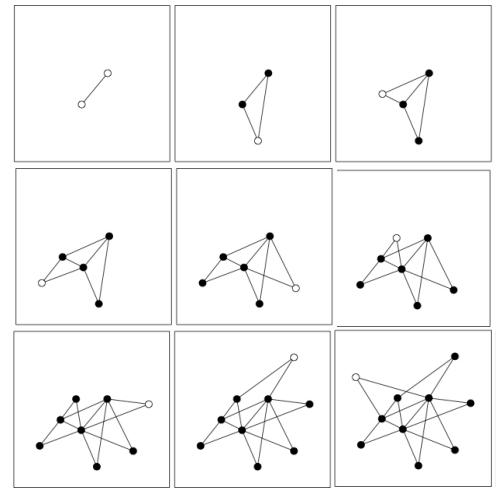


Figure 5.3
Time evolution of the Barabási-Albert model

The sequence of images shows the gradual emergence of a few highly connected nodes, or hubs, through growth and preferential attachment. White circles mark the newly added node to the network, which decides where to connect its two links ($m=2$) through preferential attachment Eq. 5.1. After [9].

richer phenomenon: due to preferential attachment new nodes are more likely to connect to the more connected nodes than to the smaller degree nodes. Hence, the more connected nodes will acquire links at the expense of the less connected nodes, eventually turning into hubs.

In summary, the Barabási-Albert model indicates that two simple mechanisms, growth and preferential attachment, are responsible for the emergence of networks with a power-law degree distribution. The origin of the power law and the associated hubs is a *rich-gets-richer phenomena* induced by the coexistence of these two ingredients. Yet, to understand the model's behavior and to quantify the emergence of the scale-free property, we need to describe the model's mathematical properties, which is the subject of the next section.

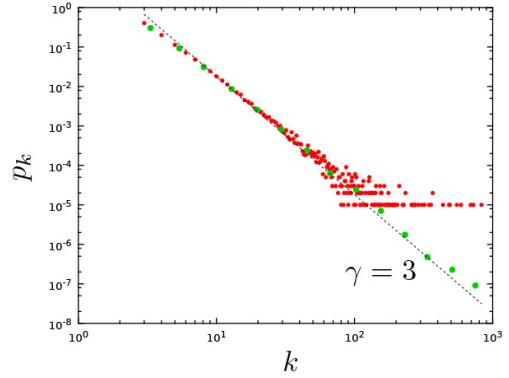


Figure 5.4
The degree distribution

The degree distribution of a network generated by the Barabási-Albert model. The plot shows p_k for a single network of size $N=100,000$ and $m=3$. It shows both the linearly-binned (red symbols) as well as the log-binned version (green symbols) of p_k . The straight line is added to guide the eye and has slope $\gamma=3$, corresponding to the resulting network's degree distribution.

To: ~~REKA~~ REKA ALBERTI

from A.L. Barabási

Rika: Probabb meg leprogramozni a Lévétkezőt
Söt működik; ami esetleg megoldja a váró hatalig függ-
végét.

MODEL 1: $t=0$: m vertices, m edges.

at time t : always add a new vertex, which
has m edges (non-directed) coming out of it.
The ends of the edges will be randomly
connected to the existing vertices in the system.

then: $t=1$: we have $(m+1)$ vertices, and m edges,
coming out of the new vertex.

$t=2$: $(m+2)$ vertices; $2m$ edges

$t=3$ $(m+3)$ vertices; $3m$ edges

thus a new vertex ~~is~~ always has m edges ~~X~~
whose ends are connected randomly to the
already present vertices in the system.

The average connectivity after time t : $\langle \delta \rangle = \frac{tm}{m+tm} \rightarrow \underline{m}$

Determine $P(\ell)$ at different times t_1, t_2, \dots , the

(where t_1, \dots are large),

and see if it is exponential or ~~or~~ power law!

Model 2: like Model 1, with
the following difference:

When you add a new vertex with the M
edges, you do not connect it with equal
probability to the existing vertices, but:

P_i : probability that the edge will be
connected to vertex i (which has edges b_i) is

$$P_i = \frac{b_i}{Z_M} \quad ; \text{ where: } (Z_M) \text{ is the total number of vertices in the system at time } t.$$

$$\left(\sum_i P_i = \frac{\sum b_i}{Z_M} = \frac{Z_M}{Z_M} = 1 \right)$$

This model will create flocking: the
vertices will grow faster than those
hopefully giving a power law scale

~~Again: determine $P(k)$~~

Kirke Farold doabrikat a ~~lukkes~~
Kirke Farold doabrikat a ~~lukkes~~
distr.: 351 26082622, and
mej meddik (all phiteren nær nu)

Q.S. Not verry, men nops va+omvendt. Model
as $(P(k) \sim k^{-\gamma})$, men a model

Donet le-hd tann venstre os en ~~lukkes~~:
op-dott vertex-ede o b*i*-je. Model 1:
Model 2:

(Vejris høgge et en dott
vertex konnektiviteten idéen)

Figure 5.5
The scale-free fax

The discovery of the Barabási-Albert model is recounted in *Linked* [9] describing a workshop in Porto, Portugal, that the author attended:

"During the summer of 1999 very few people were thinking about networks, and there were no talks on the subject during this workshop. But networks were very much on my mind. I could not help carrying with me on the trip our unresolved questions: Why hubs? Why power-laws? [...] Before I left for Europe, Réka Albert and I agreed that she would analyze these networks. On June 14, a week after my departure, I received a long email from her detailing some ongoing activities. At the end of the message there was a sentence added like an afterthought: "I looked at the connectivity distribution too, and in almost all systems (IBM, actors, power grid), the tail of the distribution follows a power law."

Réka's email suddenly made it clear that the Web was by no means special. I found myself sitting in the conference hall paying no attention to the talks, thinking about the implications of this finding. If two networks as different as the Web and the Hollywood acting community both display power-law degree distribution, then some universal law or mechanism must be responsible. If such a law existed, it could potentially apply to all networks. During the first break between talks I decided to withdraw to the quiet of the seminary where we were housed during the workshop. I did not get far, however. During the fifteen-minute walk back to my room a potential explanation occurred to me, one so simple and straightforward that I doubted it could be right. I immediately returned to the university to fax Réka, asking her to verify the idea using the computer. A few hours later she emailed me the answer. To my great astonishment, the idea worked."

The Figure is a reproduction of the two-page fax sent on June 14, 1999 from Porto to Réka Albert, describing the model that we call today the Barabási-Albert model.

BOX 5.2

THE MATHEMATICAL DEFINITION OF THE BARABÁSI-ALBERT MODEL

The definition of the Barabási-Albert model provided in this chapter leaves many mathematical details of the model unspecified:

- It does not specify the precise initial configuration of the first m_0 nodes.
- It does not specify whether the m links assigned to a new node are added one by one, independent of each other, or simultaneously. These problems were recognized by Riordan and Bollobás [10], who offered a definition that addresses these shortcomings. In contrast with the original model, Riordan and Bollobás allows for multiple edges and loops, showing later that their number will be negligible. The resulting model, called the Linearized Chord Diagram (LCD), is defined as follows:

Consider a fixed sequence of nodes v_1, v_2, \dots , where the degree of the node v_i is k_i . We build a graph $(G_1^{(t)})_{t \geq 0}$ so that $G^{(t)}$ is a graph on $v_i, 1 \leq i \leq t$ as follows: start with $G_1^{(0)}$ corresponding to an empty graph with no nodes, or with $G_1^{(1)}$ graph with one node and one loop. Given $G_1^{(t-1)}$ generate $G_1^{(t)}$ by adding the node v_t together with a single link between v_t and v_s , where i is chosen with probability

$$p_r(i=s) = \begin{cases} k_{G_1^{(t-1)}} \frac{k(v_s)}{(2t-1)}, & \text{if } 1 \leq s \leq t-1 \\ \frac{1}{(2t-1)}, & \text{if } s=t \end{cases} \quad (5.2)$$

That is, we place a link from node v_t to node v_i , where the probability that node i is chosen as the target of this new link is proportional to its degree k_i at the time, the new link already contributing to the degree of v_t . Hence, the new node v_t can also link to itself with probability $1/(2t-1)$. For $m > 1$, we add m links from v_t one by one, counting the previous links together with the outward half of the newly added link as contributing to the degrees.

DEGREE DYNAMICS

To understand the time evolution of the Barabási-Albert model, we first focus on the time-dependent degree of a node [11]. In the model a node has a chance to increase its degree each time a new node enters the network. When a new node joins the network, it will link to m of the $N(t)$ nodes present in the system. The probability that it chooses node i is given by Eq. 5.1. Assuming that k_i is a time-dependent continuous real variable, the rate at which node i acquires links follows the equation

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m \frac{k_i}{\sum_{j=1}^{N-1} k_j}. \quad (5.3)$$

The coefficient m describes that each new node arrives with m links. Hence, node i has m chances to be chosen. The sum in the denominator of Eq. 5.3 goes over all nodes in the network except the newly added node, thus

$$\sum_{j=1}^{N-1} k_j = 2mt - m. \quad (5.4)$$

Therefore Eq. 5.4 becomes

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t - 1}. \quad (5.5)$$

For large t the term (-1) can be neglected in the denominator, obtaining

$$\frac{\partial k_i}{k_i} = \frac{1}{2} \frac{\partial t}{t}. \quad (5.6)$$

By integrating Eq. 5.6 and using the fact that $k_i(t_i)=m$, meaning that node i joins the network at time t_i with m links, we obtain

$$k_i(t) = m \left(\frac{t}{t_i} \right)^\beta. \quad (5.7)$$

The exponent β is the network's dynamical exponent and has the value

$$\beta = \frac{1}{2}. \text{ Eq. 5.7} \text{ offers a number of predictions:}$$

- The degree of each node increases following a power-law with the same dynamical exponent $\beta = 1/2$ Fig. 5.6, implying that all nodes follow the same growth law.
- The growth in the degrees is sublinear (i.e. $\beta < 1$). In contrast in the Erdős-Rényi model $\langle k \rangle$ increases as $k_i \sim t$ if we add links one by one to the network. The sublinear nature of Eq. 5.7 is a consequence of the growing nature of the Barabási-Albert model: each new node has more nodes to link to than the previous nodes. Hence, with time each node competes for links with an increasing pool of nodes.
- The earlier node i was added, the higher is its degree $k_i(t)$. Hence, hubs are large not because they grow faster, but because they arrived earlier, a phenomenon called *first-mover advantage* in marketing and business.
- The growth rate of a node (i.e. the rate at which the node i acquires new links) is given by the derivative of Eq. 5.7

$$\frac{dk_i(t)}{dt} = \frac{m}{2} \frac{1}{\sqrt{t_i t}}, \quad (5.8)$$

indicating that older nodes acquire more links in a unit time (as they have smaller t_i), as well as that the rate at which a node acquires links decreases with time as $t^{-1/2}$. Hence, less and less links go to a node.

Taken together, the Barabási-Albert model offers a dynamical description of a network's evolution, capturing the fact that in real networks nodes arrive one after the other, connecting to the earlier nodes. This sets up a competition for links during which the older nodes have an advantage over the younger nodes, eventually turning into hubs.

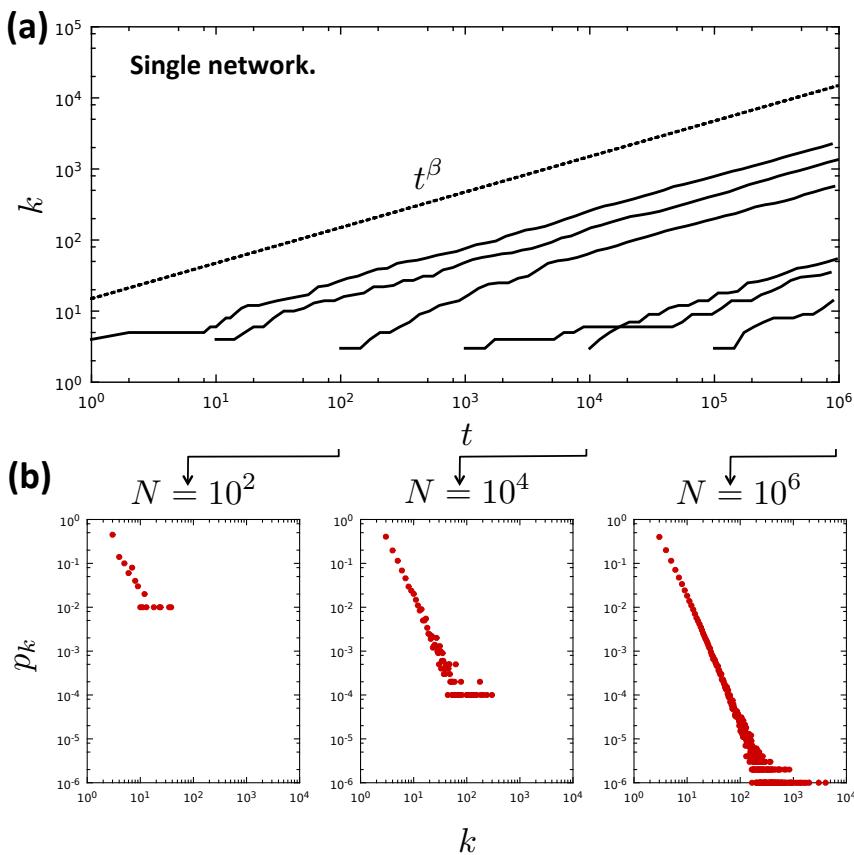


Figure 5.6
Degree dynamics

(a) Time dependence of the degrees of nodes added at time $t = 1, 10, 10^2, 10^3, 10^4, 10^5$ (continuous lines from left to right). One can see that each node increases its degree following the same law Eq. 5.7. Also, at any moment the older nodes have higher degrees. The dotted line corresponds to the analytical prediction Eq. 5.7 with $\beta = 1/2$.

(b) Degree distribution of the network after the addition of $N = 10^2, 10^4$, and 10^6 nodes, i.e. at time $t = 10^2, 10^4$, and 10^6 (illustrated by arrows in (a)). The larger the network, the more obvious is the power-law nature of the degree distribution. Note that p_k is plotted using linear binning, to better show the gradual emergence of the scale-free state.

BOX 5.3

THE MATHEMATICAL DEFINITION OF THE BARABÁSI-ALBERT MODEL

As we compare the predictions of the various network models with real data, we often have to decide how to measure time in networks. Real networks have evolved over rather different time scales: the first webpage was created in 1991, giving the WWW a history of a few decades at most. Given its trillion documents, this means that on average the WWW added more than a thousand nodes each second. In contrast the human cell is the result of 4 billion years of evolution; hence with roughly 20,000 genes, the cellular network added a node every 200,000 years. Given these enormous time-scale differences it seems impossible to use real time to compare the dynamics of these networks. Therefore, in network theory we use event time, that is, we advance time each time there is a change in the network topology. For example, in the Barabási-Albert model the addition of each new node corresponds to a new time step. Consequently in the model $t=N$. In more complicated models a distinct time step is assigned to each event—like the addition of a new node, the arrival of a new link, or the deletion of a node, any attempt to change the network topology. Obviously, if needed, we can establish a direct mapping between event time and the physical time.

DEGREE DISTRIBUTION

The distinguishing feature of networks generated by the Barabási-Albert model is their power-law degree distribution Fig. 5.4. In this section we calculate the functional form of p_k , helping us understand its origin. A number of analytical tools are available to calculate the model's degree distribution. The simplest is the continuum theory that we started developing in the previous section [1, 11]. It predicts that the degree distribution follows **BOX 5.4**,

$$p(k) \sim 2m^{1/\beta} k^{-\gamma} \quad (5.9)$$

with

$$\gamma = \frac{1}{\beta} + 1 = 3. \quad (5.10)$$

Eq. 5.9 tells us that the degree distribution follows a power law with exponent $\gamma=3$, in agreement with the numerical results shown in Fig. 5.4 and Fig. 5.7. In turn **Eq. 5.10** links the degree exponent, γ , a quantity characterizing the network topology, to the dynamical exponent, β , that characterizes a node's temporal evolution. While the continuum theory predicts the correct degree exponent ($\gamma=3$), it fails to accurately predict the pre-factors of **Eq. 5.9**. This is why we use a proportional sign in **Eq. 5.9**, rather than equality. The exact degree distribution, with the correct pre-factors, can be obtained using a master [12] or rate equation [13] approach or calculated exactly using the LCD model [10] **BOX 5.2**. As we show in **ADVANCED TOPICS 5.A**, the exact form of the degree distribution of the Barabási-Albert model is

$$p_k = \frac{2m(m+1)}{k(k+1)(k+2)}. \quad (5.11)$$

Eq. 5.11 has several notable implications:

- For large k , **Eq. 5.11** reduces to $p_k \sim k^{-3}$, or $\gamma = 3$, in line with **Eq. 5.9** and **Eq. 5.10**.
- The degree exponent γ is independent of m , a prediction that agrees with the numerical results **Fig. 5.7a**.

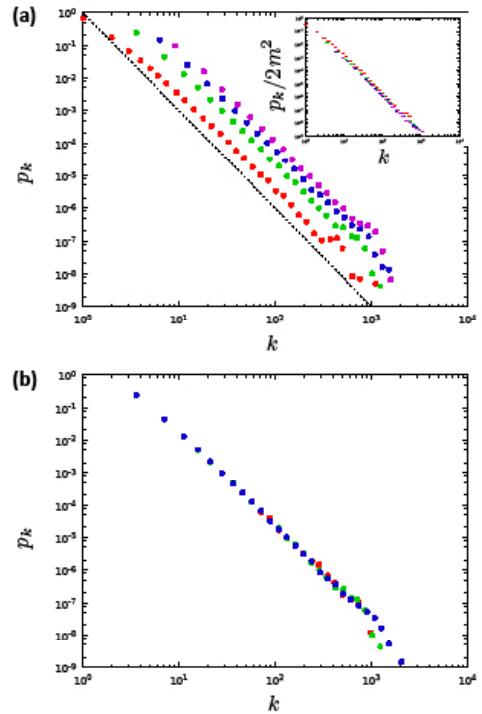


Figure 5.7
Probing the analytical predictions

(a) To show that p_k is independent of the parameters m and m_0 , we generated networks with $N=100,000$ and $m_0=m=1$ (red), 3 (green), 5 (blue), and 7 (purple). The fact that the curves are parallel to each other indicates that γ is independent of m and m_0 . The slope of the dashed line is -3. Inset: **Eq. 5.11** predicts $p_k \sim 2m^2$, hence $p_k/2m^2$ should be independent of m . Indeed, by plotting $p_k/2m^2$ vs. k all curves in the main plot collapse into a single curve.

(b) The Barabási-Albert model predicts that p_k is independent of N . To test this we plot p_k for $N = 50,000$ (red), 100,000 (green), and 200,000 (blue), with $m_0=m=3$. The obtained p_k are practically indistinguishable, indicating that the degree distribution is time invariant.

- The power-law degree distribution observed in real networks describes systems of rather different age and size. Hence, a proper model should lead to a time-independent degree distribution. Indeed, Eq. 5.11 predicts that the degree distribution of the Barabási-Albert model is time independent, resulting in the emergence of a stationary scale-free state. Numerical simulations support this prediction, indicating that p_k observed for different t (or N) fully overlap Fig. 5.7b.
- Eq. 5.11 predicts that the coefficient of the power-law distribution is proportional to $m(m + 1)$ (or m^2 for large m), again confirmed by numerical simulations Fig. 5.7, inset).

In summary, the analytical calculations confirm that the Barabási-Albert model generates a power-law degree distribution, predicting the value of the degree exponent as $\gamma=3$. The exponent is independent of the parameters m and m_0 . The calculations predict that the degree distribution is stationary (i.e. time invariant), explaining why networks with different history, size and age develop a similar degree distribution.

BOX 5.4

CONTINUUM THEORY

To calculate the degree distribution of the Barabási-Albert model we first determine the probability that the degree $k_i(t)$ of node i is smaller than a value k , i.e. $P(k_i(t) < k)$. Using Eq. 5.7, we can write

$$P(k_i(t) < k) = P\left(t_i > \frac{m^{1/\beta} t}{k^{1/\beta}}\right). \quad (5.12)$$

In the model we add the nodes at equal time intervals BOX 5.3. To capture this temporal uniformity we write the probability that a node arrives at time t_i as a random variable with a constant probability density

$$P(t_i) = \frac{1}{m_0 + t}. \quad (5.13)$$

Substituting Eq. 5.13 into Eq. 5.12 we obtain the cumulative distribution

$$P(k) = P\left(t_i \leq \frac{m^{1/\beta} t}{k^{1/\beta}}\right) = 1 - \frac{m^{1/\beta} t}{k^{1/\beta} (t + m_0)} \quad (5.14)$$

We obtain the degree distribution $p(k)$ by taking the derivative of the cumulative function, i.e.

$$p(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^{1/\beta} t}{m_0 + t} \frac{1}{k^{1/\beta+1}}, \quad (5.15)$$

which for $t \gg m_0$ reduces to Eq. 5.9.

THE ABSENCE OF GROWTH OR PREFERENTIAL ATTACHMENT

The coexistence of growth and preferential attachment in the Barabási-Albert model raises an important question: are they both necessary for the emergence of the scale-free property? In other words, could we generate a scale-free network with only one of the two ingredients? To address these questions, next we discuss two limiting cases of the model, each containing only one of the two ingredients [1, 11].

MODEL A

To test the role of preferential attachment we keep the growing character of the network (ingredient A) and eliminate preferential attachment (ingredient B). Hence, Model A starts with m_0 nodes and evolves following these steps:

(1) Growth

At each time step we add a new node with $m(\leq m_0)$ links that links to m previous nodes.

(2) Random attachment

The probability that a new node links to a node with degree k_i is

$$\Pi(k_i) = \frac{1}{(m_0 + t - 1)}. \quad (5.16)$$

That is, $\pi(k_i)$ is independent of k_i , indicating that the new nodes choose randomly the nodes they link to. The continuum theory predicts that for Model A $k_i(t)$ increases logarithmically with time, i.e.

$$k(t) = m \ln \left(e \frac{m_0 + t + 1}{m_0 + t_i + 1} \right) \quad (5.17)$$

a much slower increase than the power law Eq. 5.7 derived earlier. Consequently the degree distribution becomes exponential Fig. 5.8a

$$p_k = \frac{e}{m} \exp \left(-\frac{k}{m} \right). \quad (5.18)$$

As an exponential function decays much faster than a power law, it does not support hubs. Therefore the lack of preferential attachment eliminates the network's scale-free character and the hubs.

MODEL B

To test the role of growth we next keep preferential attachment (ingredient B) and eliminate growth (ingredient A). Hence, Model B starts with N nodes and evolves following this step:

Preferential Attachment:

At each time step a node is selected randomly and connects to a node i with degree k_i already present in the network, where i is chosen with probability Eq. 5.1.

In Model B, the number of nodes remains constant during the network's evolution, but the number of links increases linearly with time. As a result the degree of each node also increases linearly with time Fig. 5.8b, inset

$$k_i(t) \approx \frac{2}{N}t \quad (5.19)$$

Indeed, in each time step we add a new link, without changing the number of nodes. At early times, when there are only a few links in the network (i.e. $L \ll N$), each new link connects previously unconnected nodes. In this stage the model's evolution is indistinguishable from the Barabási-Albert model with $m=1$. Numerical simulations show that in this regime the model develops a degree distribution with a power-law tail Fig. 5.8b. Yet, p_k is not stationary, as after roughly $T \approx N^2$ time steps the network converges to a complete graph. Consequently, after a transient period ($t \ll N$) the node degrees start to converge to the average degree Eq. 5.19 and the degree distribution becomes peaked Fig. 5.8b. For $t \rightarrow N(N-1)/2$ the degree distribution becomes $p_k = \delta(N-1)$, i.e. the network turns into a complete graph in which all nodes have degree $k_{\max} = N-1$. Therefore, in the absence of growth the network is not stationary, becoming a complete graph with time.

In summary, the failure of Models A and B to reproduce the empirically observed scale-free distribution indicates that growth and preferential attachment are simultaneously needed for the emergence of the scale-free property.

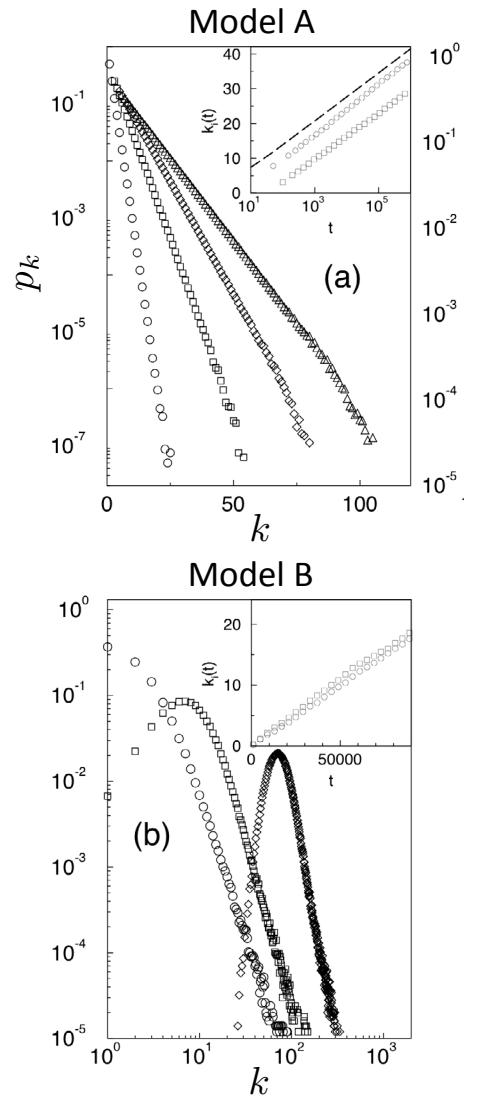


Figure 5.8
Model A and Model B

Numerical simulations of Model A and B, probing the role of growth and preferential attachment.

(a) Degree distribution for Model A, that incorporates growth but lacks preferential attachment. The symbols correspond to $m_0=m=1$ (circles), 3 (squares), 5 diamonds), 7 (triangles) and $N=800,000$.

Inset: Time evolution of the degree of two vertices added at $t=7$ and $t=97$ for $m_0=m=3$. The dashed line follows $k_i(t)=m \ln(m_0+t^1)$ as predicted by Eq. 5.16 for large t .

(b) Degree distribution for Model B, that lacks growth but incorporates preferential attachment, shown for $N=10,000$ and $t=N$ (circles), $t=5N$ (squares), and $t=40N$ (diamonds).

Inset: Time dependence of the degrees of two vertices for system size $N=10,000$, indicating that $k_i(t)$ grows linearly, as predicted by Eq. 5.19. After [14].

MEASURING PREFERENTIAL ATTACHMENT IN REAL NETWORKS

In the previous sections we showed that growth and preferential attachment are responsible for the scale-free property. The presence of growth in real systems is obvious: all large networks arrived to their current size by continuously adding new nodes. But to convince ourselves that preferential attachment is also present in real networks, we need to detect it experimentally. In this section we show how to detect preferential attachment by measuring the $\pi(k)$ function in real networks. We start by noting that preferential attachment incorporates two hypotheses:

HYPOTHESIS 1

The likelihood to connect to a node depends on the node's degree k . This is in contrast with the random network model, for which $\pi(k)$ is independent of k .

HYPOTHESIS 2

The functional form of $\pi(k)$ is linear in k .

Both hypotheses can be tested by measuring $\pi(k)$. To be specific, we can determine $\pi(k)$ for networks for which we [14, 15] know the time at which each node joined the network, or we have at least two network maps collected at not too distant moments in time.

Consider a network for which we have two different maps, the first taken at time t and the other at time $t + \Delta t$ Fig. 5.9. During the Δt time frame some nodes did not change their degree, so for these $k(t+\Delta t) = k(t)$. For nodes that did alter their degree we measure the change $\Delta k_i = k_i(t+\Delta t) - k_i(t)$. According to Eq. 5.1, the relative change $\Delta k_i/\Delta t$ should follow

$$\frac{\Delta k_i}{\Delta t} \propto \Pi(k_i) \quad (5.20)$$

providing the functional form of preferential attachment. For Eq. 5.20 to be valid we must keep Δt small, so that the changes in Δk are relatively small. But it must not be too small so that there are still detectable differences between the two networks.

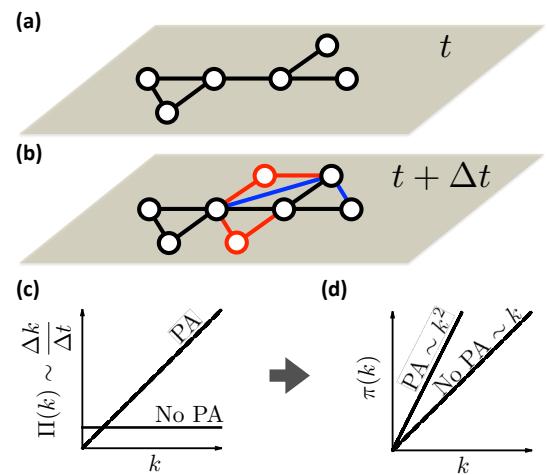


Figure 5.9
Detecting preferential attachment

If we have access to two maps of the same network, taken at time t and $t + \Delta t$, comparing them allows us to measure the $\pi(k)$ function that governs preferential attachment. Specifically we look at nodes that have gained new links thanks to the arrival of new nodes, like the two new red nodes at $t + \Delta t$. The blue lines correspond to links that connect previously disconnected nodes, called internal links. Their role is discussed in CHAPTER 6.

(c) In the presence of preferential attachment $\Delta k/\Delta t$ will depend linearly on a node's degree at time t .

(d) The scaling of the cumulative preferential attachment function helps us detect the presence or absence of preferential attachment.

In practice the obtained $\Delta k_i/\Delta t$ curve is typically noisy, particularly for small networks. To reduce this noise we often measure the cumulative function

$$\pi(k) = \sum_{k_i=0}^k \Pi(k_i). \quad (5.21)$$

In the absence of preferential attachment we expect $\pi(k)=\text{constant}$, hence, $\pi(k) \sim k$ according to Eq. 5.21. If preferential attachment is present, i.e. $\pi(k_i)=k_i$, we expect $\pi(k) \sim k^2$. Fig. 5.10 shows the measured $\pi(k)$ for four real networks. For each system we observe a faster than linear increase in $\pi(k)$, indicating the presence of preferential attachment. Fig. 5.10 also suggests that $\pi(k)$ can be approximated with

$$\Pi(k) \sim k^\alpha. \quad (5.22)$$

For the Internet and citation networks we have $\alpha \approx 1$, indicating that $\pi(k)$ depends linearly on k , as assumed in Eq. 5.2. This is in line with Hypotheses 1 and 2. For the co-authorship and the actor network the best fit provides $\alpha=0.9 \pm 0.1$ indicating the potential presence of a sublinear preferential attachment.

In summary, Eq. 5.20 helps us detect the presence (or absence) of preferential attachment in real networks. The measurements show that the attachment probability depends on the node degree, in line with Hypothesis 1. Yet, we also find that while in some systems preferential attachment is linear, in others it can be sublinear, hence, Hypothesis 2 is occasionally violated. The implications of this non-linearity is discussed in the next section.

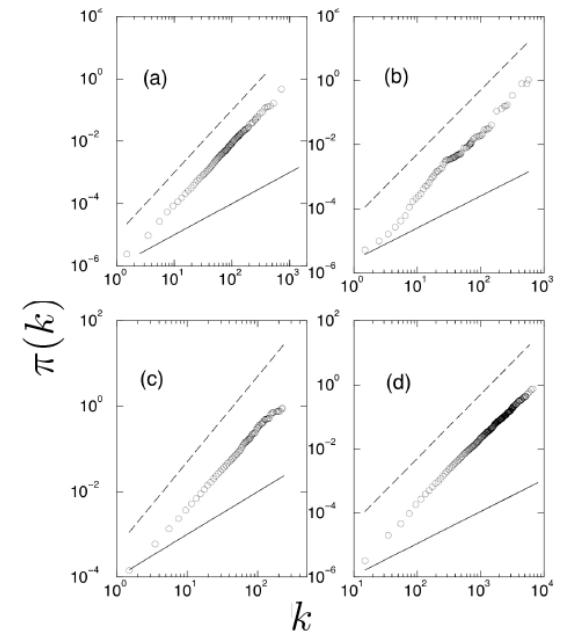


Figure 5.10
Evidence of preferential attachment

The figure shows the cumulative preferential attachment function $\pi(k)$, defined in Eq. 5.21, for several real systems:

- (a) A citation network
- (b) The Internet
- (c) Neuroscience scientific collaboration network
- (d) Actor network

In each panel we have two lines to guide the eye: the dashed line corresponds to linear preferential attachment ($\pi(k) \sim k^2$) and the continuous line indicates the absence of preferential attachment ($\pi(k) \sim k$). In line with Hypothesis 1 we detect a k -dependence in each dataset. Yet, in (c) and (d) $\pi(k)$ grows slower than k^2 , indicating that for these two systems preferential attachment is sublinear, violating Hypothesis 2. Note that these measurements only consider links added through the arrival of new nodes, ignoring the addition of internal links. After [14].

NON-LINEAR PREFERENTIAL ATTACHMENT

The observation of sublinear preferential attachment in Fig. 5.9 raises an important question: what is the impact of this nonlinearity on the network topology? To answer this we replace the linear preferential attachment Eq. 5.1 with Eq. 5.21 and re-calculate the degree distribution of the Barabási-Albert model. The behavior for $\alpha=0$ is clear: in the absence of preferential attachment, the model reduces to Model A discussed in SECTION 5.4. Consequently the degree distribution will follow the simple exponential function Eq. 5.17. For $\alpha = 1$ we recover the Barabási-Albert model, obtaining a scale-free network with degree distribution Eq. 5.14. We next focus on the case when $\alpha \neq 0$ and $\alpha \neq 1$. The calculation, providing p_k for an arbitrary α , is presented in ADVANCED TOPICS 5.B, predicting several scaling regimes [13]:

SUBLINEAR PREFERENTIAL ATTACHMENT ($0 < \alpha < 1$)

For any $\alpha > 0$ new nodes favor the more connected nodes over the less connected nodes. Yet, for $\alpha < 1$ the bias is not sufficient to generate a scale-free degree distribution. Instead, in this regime the degrees follow the stretched exponential distribution SECT. 4.10

$$p_k \sim k^{-\alpha} \exp\left(\frac{2\mu(\alpha)}{\langle k \rangle(1-\alpha)} k^{1-\alpha}\right) \quad (5.23)$$

where $\mu(\alpha)$ is a function that depends only weakly on α . For $\alpha \rightarrow 1$ Eq. 5.22 reduces to the degree distribution of the BA model. Indeed for $\alpha=1$ we have $\mu=2$, and $\lim_{\alpha \rightarrow 1} \frac{k^{1-\alpha}}{1-\alpha} = \ln k$. Therefore $p_k \sim k^{-1} \exp(-2\ln k) = k^{-3}$. The exponential cutoff in Eq. 5.22 implies that sublinear preferential attachment limits the size and the number of the hubs.

Sublinear preferential attachment also affects the size of the largest degree, k_{\max} . In CHAPTER 4 we showed that for a scale-free network the degree of the largest node scales polynomially with time Eq. 4.14. For sub-linear preferential attachment we have

$$k_{\max} \sim (\ln t)^{1/(1-\alpha)} \quad (5.24)$$

a logarithmic dependence that predicts a much slower growth of the maximum degree than the polynomial. This slower growth is the reason

why the hubs are smaller for $\alpha < 1$ Fig. 5.10.

SUPERLINEAR PREFERENTIAL ATTACHMENT ($\alpha > 1$)

For $\alpha > 1$ the tendency to link to highly connected nodes is enhanced, accelerating the *rich-gets-richer process*. The consequence of this is most obvious for $\alpha > 2$, when the model predicts a *winner-takes-all* phenomenon: almost all nodes connect to a single or a few super-hubs. Hence, we observe the emergence of a hub-and-spoke network, in which most nodes link directly to a few central nodes. The situation for $1 < \alpha < 2$ is less extreme, but similar. This winner-takes-all process impacts the time dependence of the largest hub as well, finding that Fig. 5.12.

Hence for $\alpha > 1$ the largest hub links to a finite fraction of nodes in the system.

$$k_{\max} \sim t \quad (5.25)$$

In summary, nonlinear preferential attachment introduces deviations from the power law degree distribution, either limiting the size of the hubs ($\alpha < 1$), or leading to super-hubs ($\alpha > 1$, Fig. 5.12). Hence, $\pi(k)$ needs to depend strictly linearly on the degrees for the resulting network to have a pure power law p_k . While in many systems we do observe such a linear behavior, in others, like the scientific collaboration network and the actor network, preferential attachment is sublinear, limiting the size of the hubs. This sublinear form of $\pi(k)$ could be responsible for the systematic deviations from a pure power-law degree distribution observed in the previous chapter. Hence for these systems a stretched exponential Eq. 5.22 should offer a better fit to the degree distribution.

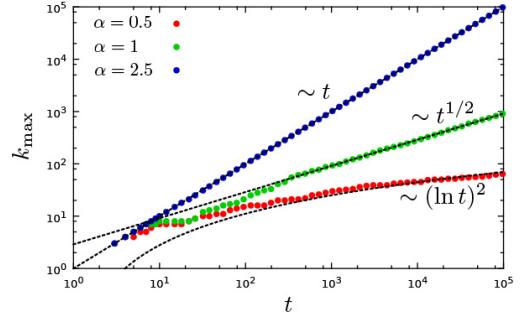


Figure 5.11
The growth of the hubs

The nature of preferential attachment affects the degree of the largest node. While in a scale-free network ($\alpha=1$) the biggest hub grows as $t^{1/2}$ (green curve) Eq. 4.14, for sublinear preferential ($\alpha<1$) attachment this dependence becomes logarithmic (red curve, Eq. 5.24). For superlinear preferential attachment the biggest hub grows linearly with time, always grabbing a finite fraction of all links (blue curve), Eq. 5.25. The symbols are provided by a numerical simulation; the dotted lines represent the analytical predictions.

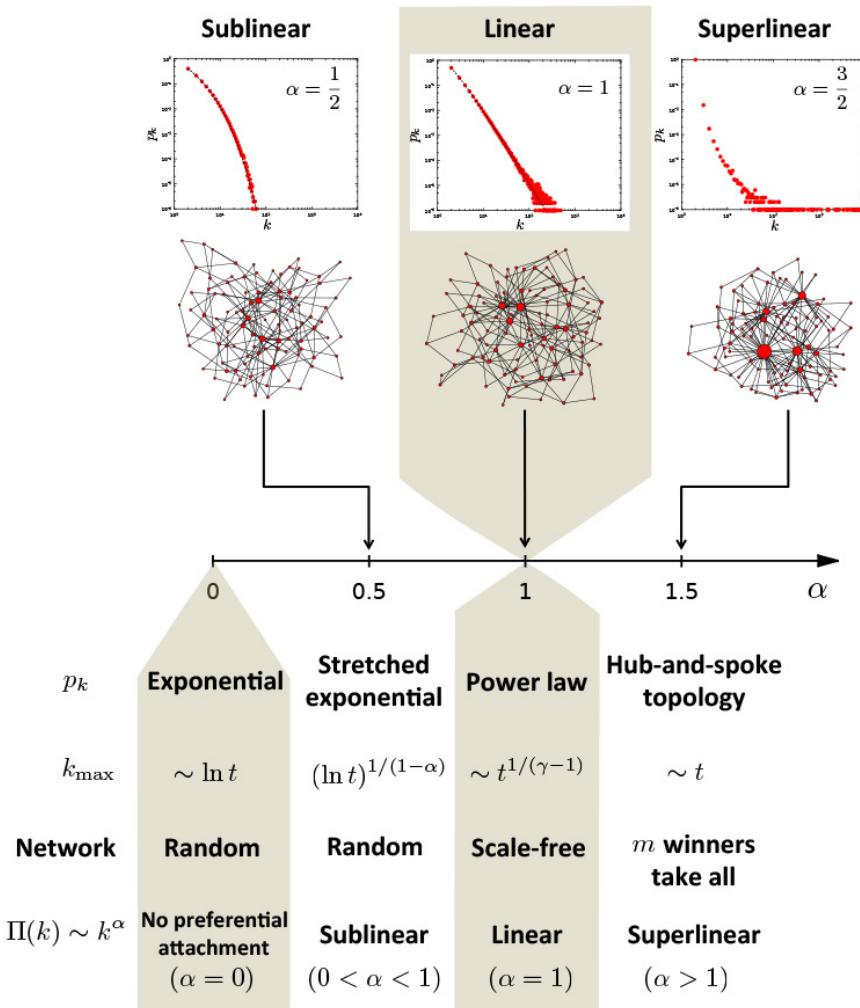


Figure 5.12
Nonlinear preferential attachment

The scaling regimes characterizing networks driven by non-linear preferential attachment. The three top panels show p_k for different α values ($N=10^4$). The network maps under them show the obtained topologies for $N=100$ nodes. The theoretical results indicate the existence of four scaling regimes:

No preferential attachment ($\alpha=0$)

The network has a simple exponential degree distribution following Eq. 5.18. Hubs are absent and the resulting network behaves like a random network.

Sublinear preferential attachment ($0 < \alpha < 1$)

The degree distribution follows the stretched exponential Eq. 5.23, resulting in fewer and smaller hubs than in a scale-free network. As $\alpha \rightarrow 1$ the cutoff length increases, hence p_k follows a power law over an increasing range of degrees.

Linear preferential attachment ($\alpha=1$)

This corresponds to the Barabási-Albert model, hence the degree distribution follows a power law.

Superlinear preferential attachment ($\alpha > 1$)

The high-degree nodes are disproportionately attractive and the network follows a winner-takes-all dynamics, developing a hub-and-spoke topology. In this configuration the earliest nodes become super hubs and all subsequent nodes link to them. Hence p_k shown for $\alpha=1.5$ (top panel), indicates the coexistence of many small nodes with a few super hubs in the vicinity of $k=10^4$.

THE ORIGINS OF PREFERENTIAL ATTACHMENT

Given the key role preferential attachment plays in the evolution of real networks, we must ask, Where does preferential attachment come from? The question can be broken down to two narrower issues:

Why does $\pi(k)$ depend on k ?

Why is the dependence of $\pi(k)$ linear in k ?

In the past decade we witnessed the emergence of two philosophically different approaches to these questions. In the first class belong models that view preferential attachment as a result of an interplay between random events and some structural property of a network. These mechanisms do not require global knowledge of the network and rely on random actions, hence we will call them *local* or *random* mechanisms. A second class of models assume that each new node or link is preceded by a cost-benefit analysis, balancing various needs with the available resources. These models assume familiarity with the whole network and rely on optimization principles, prompting us to call them *global* or *optimized* mechanisms. The purpose of this section is to discuss these two approaches.

LOCAL MECHANISMS

The link selection model offers perhaps the simplest example of a local or random mechanism capable of generating preferential attachment [16]. It is defined as follows:

- **Growth**

At each time step we add a new node to the network.

- **Link selection**

We select a link at random and connect the new node to one of the two nodes at the two ends of the selected link. This procedure is inherently local and random, as one does not need to know anything about the overall network topology to connect the new node. To show that this simple mechanism generates linear preferential attachment, we write the probability q_k that the node at the end of a randomly chosen

link has degree k as

$$q_k = Ckp_k. \quad (5.26)$$

Eq. 5.26 captures two effects:

(i) The higher the degree of a node, the higher the chance that it will be located at the end of the chosen link.

(ii) The more degree- k nodes are in the network (i.e., the higher is p_k), the more likely that a degree k node will be at the end of the link.

In **Eq. 5.26** the value of C can be calculated using the normalization condition $\sum q_k = 1$, obtaining $C=1/\langle k \rangle$. Hence the probability that we find a degree- k node at the end of a randomly chosen link is

$$q_k = \frac{kp_k}{\langle k \rangle}. \quad (5.27)$$

a quantity called excess degree. **Eq. 5.27** also represents the probability that a new node connects to a node with degree k in the link selection model, hence it plays the role of the preferential attachment $\pi(k)$. Therefore **Eq. 5.26** indicates that random link selection generates preferential attachment that is linear in then degree. While link selection is perhaps the simplest mechanism for preferential attachment, it is neither the first nor the most popular in the class of models relying on local mechanisms. That distinction goes to is the *copying model*, described in **Fig. 5.13**.

OPTIMIZATION

A longstanding assumption of economics is that humans make rational decisions, balancing cost against benefits. Innother words, each individual aims to maximize its personal advantage. This is the starting point of rational choice theory in economics [21] and it is a hypothesis central to modern political science, sociology and philosophy. As we discuss below, such rational decisions can lead to preferential attachment as well [22, 23, 24].

Consider the Internet, whose nodes are routers or autonomous systems (AS), connected to each other via cables. Establishing a new Internet connection between two routers requires laying down a cable between them. As this can be costly, each new link is preceded by careful cost-benefit analysis. Each new router must agree with the nodes it links to that they will transmit the data packets leaving from or arriving to the new node (peering relationship). Therefore each new node will choose its link to balance access to good network performance (like proper bandwith) with the cost of laying down a new cable i.e. physical distance). This can be a conflicting desire, as the closest node does not always offer the best network performance. For simplicity let us assume that the nodes are all located on a unit square. At each time step we add a new node by randomly choosing a point within the square. When deciding where to connect the new node i to the existing nodes, we calculate the cost function [22]

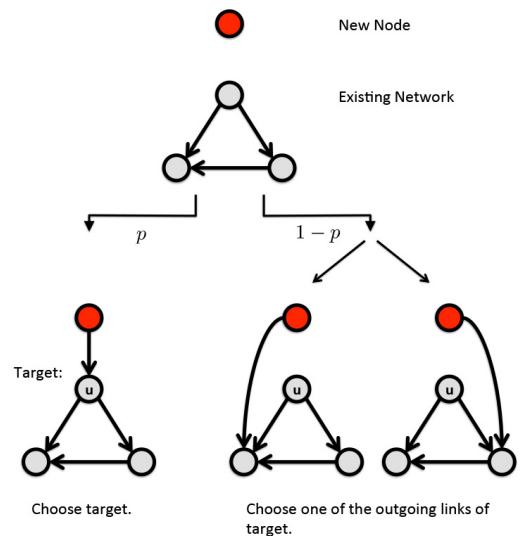


Figure 5.13
Copying model

When building a new webpage, authors tend to borrow links from webpages covering similar topics, a process captured by the copying model [17, 18]. In the model, in each time step a new node with a single link is added to the network. To choose the target node we randomly select a node u and follow a two-step procedure:

(a) **Random Connection:** With probability p the new node links to u .

(b) **Copying:** With probability $1-p$ we randomly choose an outgoing link of node u and link the new node to the link's target. Hence, the new node *copies* one of the links of an earlier node. For step (a) the probability of selecting a particular node is $1/N$. Step (b) is equivalent with selecting a node linked to a randomly selected link. The probability of selecting a degree- k node through the copying process of step (b) is $k/2L$ for undirected networks. That is, the likelihood that the new node will connect to a degree- k node follows preferential attachment $\pi(k) = p / N + (1-p)k/(2L)$, which is linear in k . The popularity of the copying model lies in its adaptability to real systems:

- **Social networks:** The more acquaintances an individual has, the higher is the chance that she will be introduced to new individuals by her existing acquaintances. Without friends, it is difficult to make new friends.

- **Citation Networks:** No scientist can be familiar with all papers published on a certain topic. If we assume that authors decide what to cite by randomly selecting references from the papers they have already read, then papers with more citations are more likely to be cited again.

- **Protein interaction networks:** Gene duplication, a common mechanism leading to next genes in a cell, can be mapped into the copying model, explaining the scale-free nature of protein interactions networks [19, 20].

$$C_i = \min_j [\delta d_{ij} + h_j] \quad (5.28)$$

for each node j already in the network, where d_{ij} is the Euclidean distance between i and j , and h_j is the network-based distance of node j to the first node in the network, designated as the “center” of the network Fig. 5.14. Hence h_j captures the “resources” offered by node j , in the form of its distance to the network’s center. The calculations indicate the emergence of three distinct network topologies, depending on the value of the parameter δ in Eq. 5.28 and Fig. 5.15:

STAR NETWORK $\delta < (1/2)^{1/2}$

For $\delta = 0$ the Euclidean distances are irrelevant, hence each node will simply link to the central node, turning the network into a star. This star configuration persists for any $\delta < (1/2)^{1/2}$, guaranteeing that the h_{ij} term dominates over the δd_{ij} term in Eq. 5.28.

RANDOM NETWORK $\delta \geq N^{1/2}$

For very large δ the contribution provided by the distance term δd_{ij} overwhelms h_j in Eq. 5.27. In this case each new node will connect to the node closest to it. The resulting graph is a dynamic version of the Euclidean minimum spanning tree. The resulting network will have a bounded degree distribution, like a random network Fig. 5.15.

SCALE-FREE NETWORK $4 \leq \delta \leq N^{1/2}$

Numerical simulations and analytical calculations [22] indicate that for intermediate δ values the network develops a scale-free topology.

The origin of the power law is rooted in two competing mechanisms:

- (i) Optimization: Each node has a basin of attraction, so that nodes landing in this basin will always link to it Fig. 5.14. The size of each basin correlate with h_j of node j at its center, which in turn correlates with the node’s degree k_j Fig 5.14f.
- (ii) Randomness: We choose randomly the position of the new node, ending in one of the N basins of attraction. The node with the largest degree largest basin of attraction, will gain the most new nodes and links. This leads to preferential attachment, documented in Fig. 5.15d.

In summary, the microscopic mechanisms responsible for preferential attachment can have two fundamentally different origins **BOX 5.5**: it can be rooted in random processes, like link selection or copying, or in optimization process, when new nodes balance conflicting criteria as they decide where to connect. These results help us understand why preferential attachment is present in so different systems as the cell or the Internet. The diversity of the mechanisms discussed in this section suggest that preferential attachment is so widespread precisely because it can come from both rational choice and random actions [25]. Most complex systems are driven by processes that have a bit of both. Hence luck or reason, preferential attachment wins either way.

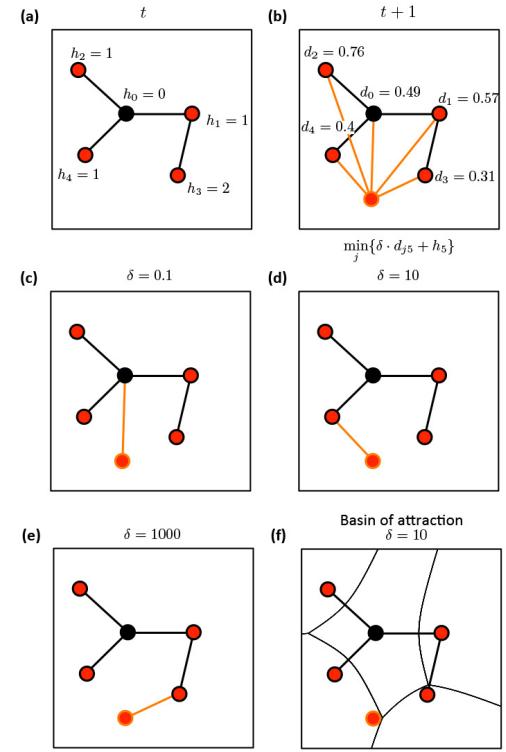


Figure 5.14
Optimization Model

(a) A small network configuration, where the h_j term in the cost function of Eq. 5.28 is shown for each node. Here h_j represents the network-based distance of node j from node $i=0$. Hence $h_0=0$ and $h_3=2$.

(b) A new node, shown in orange, will choose the node j to connect to by minimizing c_j of Eq. 5.28. If $\delta = 0$ or δ is small the new node will connect to the central node with $h_j=0$.

(c)-(e) As we increase δ , the balance in Eq. 5.28 changes, forcing the new node to connect to different nodes. The panels (c)-(e) show the choice of the new node (orange) for a different values of δ for the given network configuration.

(f) The basin of attraction for each node for $\delta=10$. A new node arriving inside a particular basin will always link to the node at the center of the basin. The size of each basin depends on the degree of the node at its center. Indeed, the smaller is h_j , the larger can be the distance to the new node while still minimizing the cost Eq. 5.28. Yet, the higher the degree of node j , the smaller is its expected distance to the central node h_j .

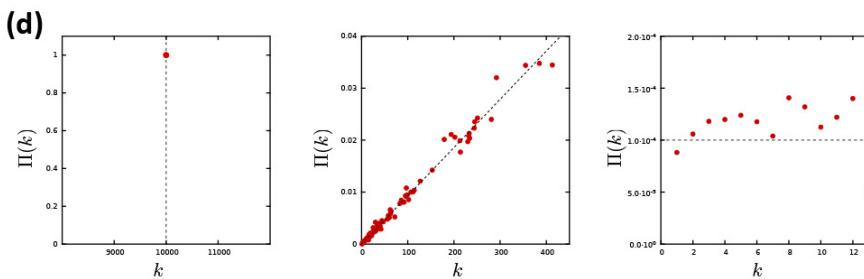
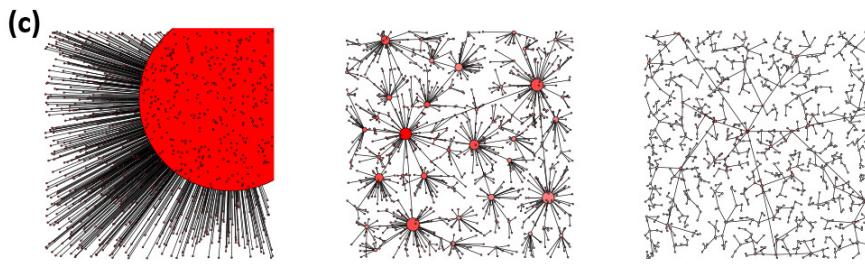
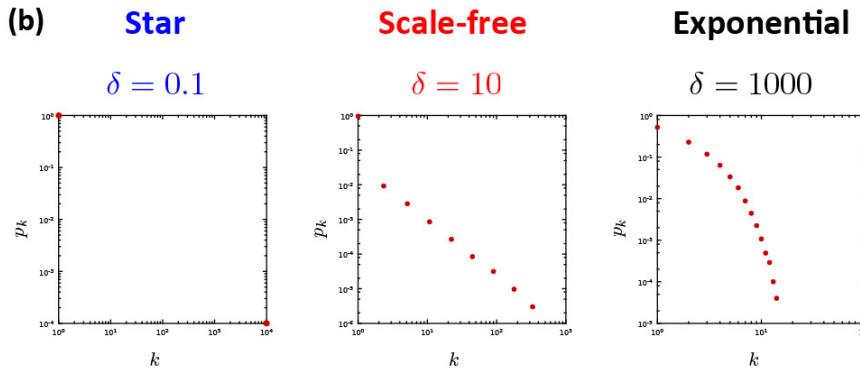
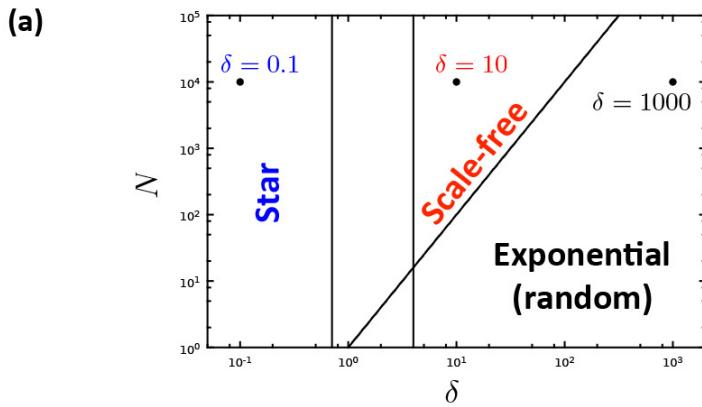


Figure 5.15
Scaling in the optimization model

(a) A schematic diagram, describing the three main classes of networks generated by the optimization model: star topology, scale-free topology and exponential networks. The structure of the network in the unmarked area is unknown. The boundary of the star configuration is given by $s=(1/2)^{1/2}$. Indeed, the maximum distance between two nodes on a square lattice with unit length, over which the model is defined, is the diagonal $2^{1/2}$. Therefore if $\delta < 1/2^{1/2}$, for any new node $\delta d_i < 1$. In this case the cost of connecting to the central node is $c_i = \delta d_{ij} + 0$, which is always lower than connecting to any other node at the cost of $f(i, j) = \delta d_{ij} + 1$. Therefore for $\delta < (1/2)^{1/2}$ all nodes connect to node 0, resulting in a network dominated by a single hub (star network, see (c)). The oblique line making the boundary of the scale-free regime is $\delta = N^{1/2}$. Indeed, if nodes are placed randomly on the unit square, then the typical distance between neighbors decreases as $N^{-1/2}$. Hence, if $d_{ij} \sim N^{-1/2}$ then $\delta d_{ij} \geq h_{ij}$ for most node pairs. Typically the path length to the central node h_j grows slower than N (in small-world networks $h_j \sim \log N$, in scale-free networks $h_j \sim \ln \ln N$). Therefore C_i is dominated by the δd_{ij} term and the smallest C_i is achieved by minimizing the distance-dependent term. Note that strictly speaking the transition only occurs in the $N \rightarrow \infty$.

(b) Degree distribution for networks emerging in the three phases discussed above for $N=10^4$.

(c) Typical topologies generated by the optimization model for the selected δ values. The node size is chosen to be proportional to its degree.

(d) We used the method described in SECT. 5.7 to measure the preferential attachment function $\Pi(k)$. Starting from a network with $N=10,000$ nodes we added a new node and measured the degree of the node that it connected to. We repeated this procedure 10,000 times, obtaining $\Pi(k)$. The plots indicate the presence of a linear preferential attachment in the scale-free phase, but its absence in the star and the exponential phases.

BOX 5.5

LUCK OR REASON: AN ANCIENT DEBATE

The tension between randomness and optimization, two apparently antagonistic explanations for power laws, is by no means new: in the 1960s Herbert Simon and Benoit Mandelbrot have engaged in a fierce public dispute over this very topic. Simon proposed that randomness is responsible for the power-law nature of word frequencies. Mandelbrot, however, fiercely defended an optimization-based framework.

The debate spanned seven papers and several years and is one of the most vicious scientific disagreement on record. It is hard to know what set it off—it may have been Simon's brief note in his 1955 paper [26], dismissing Mandelbrot's explanation that power laws observed in linguistics are rooted in an optimization process [27]. Mandelbrot responded with a comment [28] stating that 'Simon's model is analytically circular.' The essence of Simon's lengthy reply a year later is well summarized in its abstract: 'Dr. Mandelbrot's principal and mathematical objections to the model are shown to be unfounded' [29]. This prompted a 19 page response by Mandelbrot entitled 'Final Note [...]', stating that 'most of Simon's (1960) reply was irrelevant' [30] and, ensuring that this will not be the final note. Sure enough, Simon's subsequent reply states that 'this present "Reply" refutes the almost entirely new arguments introduced by Dr. Mandelbrot' [31].

That inspired a paper creatively entitled a "Post Scriptum to "Final Note,"" by Mandlebrot [32], stating that 'My criticism has not changed since I first had the privilege of commenting upon a draft of Simon,' Simon's final note ends but does not resolve the debate: "Dr. Mandelbrot has proposed a new set of objections to my 1955 models of Yule distributions. Like earlier objections, these are invalid." [33].

In the context of networks the argument titled in Simon's favor the power laws observed in complex networks appear to be driven by randomness and preferential attachment. Yet, as we seek to explain the origins of preferential attachment, the optimization-based ideas proposed by Mandelbrot play an important role.

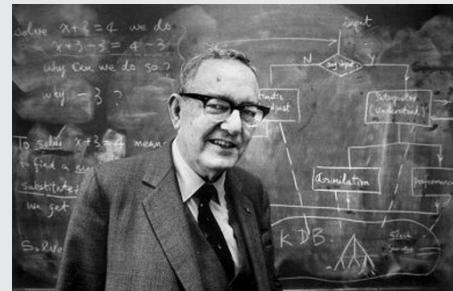


Figure 5.15a
Herbert Simon



Figure 5.15b
Benoit Mandelbrot

DIAMETER AND CLUSTERING COEFFICIENT

To complete the characterization of the Barabási-Albert model we need to discuss the behavior of two additional measures: the network diameter and the clustering coefficient. Both quantities play an important role in comparing the model predictions to the properties of real systems.

DIAMETER

The network diameter, representing the maximum distance in the Barabási-Albert network, is predicted to follow

$$D \sim \frac{\log N}{\log \log N} . \quad (5.29)$$

a result obtained independently by Cohen and Havlin [34] and Bollobás and Riordan [35], the latter also offering an exact proof. Eq. 5.29 tells us that the network diameter grows slower than $\log N$, hence the distances in the Barabási-Albert model are smaller than the distances observed in a random graph of similar size. The difference is particularly relevant for large N . Note that while Eq. 5.29 is derived for the diameter, we expect that the average distance $\langle d \rangle$ scales in a similar fashion. The impact of the $\log \log N$ term captures the scaling of $\langle d \rangle$ with N , but for large $N (\geq 10^4)$ the impact of the logarithmic correction becomes noticeable.

CLUSTERING COEFFICIENT

The clustering coefficient of the Barabási-Albert model follows ADVANCED TOPICS 5.C.

$$C = \frac{m-1}{8} \frac{(\ln N)^2}{N} , \quad (5.30)$$

a result obtained by Klemm and Eguiluz [36], and proved by Bollobás [37]. The prediction Eq. 5.30 is quite different from the $1/N$ dependence obtained for the random network model Fig. 3.20. The difference comes in the $(\ln N)^2$ term, that increases the clustering coefficient for large N . Consequently the clustering coefficient of the Barabási-Albert model decays slower than expected for a random network, indicating that the obtained network is locally more clustered.

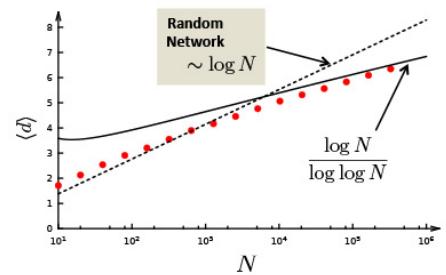


Figure 5.16
Average distance

The dependence of the average distance on the system size in the Barabási-Albert model. The continuous line corresponds to the exact result Eq. 5.29, while the dotted line corresponds to the prediction obtained in CH. 3 for a random network. Note that the analytical predictions do not provide the exact prefactors, hence the lines are not fits, but indicate only the predicted N dependent trends. The results were averaged for ten independent runs for $m = 2$.

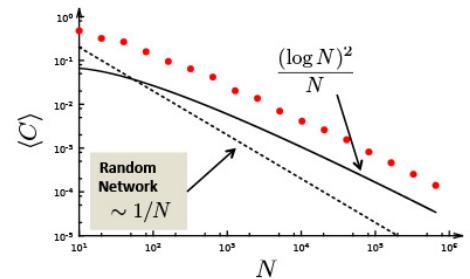


Figure 5.17
Clustering coefficient

The dependence of the average clustering coefficient on the system size N for the Barabási-Albert model. The continuous line corresponds to the analytical prediction Eq. 5.30, while the dotted line corresponds to the prediction for a random network, for which $\langle C \rangle \sim 1/N$. The results were averaged for ten independent runs for $m = 2$. The dashed and continuous curves are only drawn to indicate the N dependent trends. Hence, they do not represent a precise fit.

SUMMARY

The most important message of the Barabási-Albert model is that network structure and evolution are inseparable. Indeed, in the Erdős-Rényi, configuration or the hidden parameter models the role of the modeler is to place the links between a fixed number of nodes. Returning to our earlier analogy, the networks generated by these models relate to real networks like a photo of a painting relates to the painting itself: it may look like the real one, but the process of generating a photo is drastically different from the process of painting the original painting. The aim of the Barabási-Albert model is to capture the processes that assemble a network in the first place. Hence, it aims to paint the painting again, coming as close as feasible to the original brush strokes. Consequently, the modeling philosophy of the Barabási-Albert model is simple: *to understand the topology of a complex system, we first need to describe how it came into being.*

Dynamics and network assembly take the driving role and the structural characteristics of the network, like the degree distribution, is a by-product of this modeling philosophy. Random graphs, the configuration and the hidden parameter models will continue to play an important role as we try to understand how certain network properties deviate from our expectations. Yet, if we want to explain the origin of a particular network property, we will have to use models that capture the system's genesis. In its current form the Barabási-Albert model cannot describe the wide range of network characteristics observed in real systems. This is illustrated by the model's notable limitations:

- It predicts $\gamma=3$ while the degree exponent of real networks varies between 2 and 5 [Table 4.2](#).
- Many networks, like the WWW or citation networks, are directed, while the model generates undirected networks.
- Many processes known to occur in networks, from linking already existing nodes to the disappearance of links and nodes, are absent from the model.

- The model does not allow to distinguish between nodes based on some intrinsic characteristics, like the novelty of a research paper or the utility of a webpage. While the Barabási-Albert model is occasionally used as a model of the Internet or the cell, in reality is not designed to capture the details of any particular real network systems. It is a minimal, proof of principle model whose main purpose is to capture the basic mechanisms responsible for the emergence of the scale-free property.

Therefore, if we want to understand the evolution of systems like the Internet, the cell or the WWW, we need to incorporate the important details that contribute to the time evolution of these systems, like the directed nature of the WWW, the possibility of internal links and node and link removal. As we show in [CHAPTER 6](#), these limitations can be systematically resolved. Finally, the results discussed in this chapter allow us to formulate the next law:

The Third Law of Networks: Growth and Preferential Attachment.

Hubs and power laws are a joint consequence of growth and preferential attachment.

Let us revisit the three criteria we used earlier to establish the validity of a network law:

- (a) Quantitative formulation of the third law is provided by the Barabási-Albert model, together with its documented ability to generate scale-free networks based on growth and preferential attachment.
- (b) Universality: [SECTION 5.7](#) offers direct empirical evidence that real networks that exhibit the scale-free property are characterized by preferential attachment; [SECTION 5.2](#) offers evidence of growth.
- (c) Non-random origin: Preferential attachment is obviously absent from random networks, which is the main reason why random networks do not develop hubs and power laws.

BOX 5.6

AT A GLANCE: BARABÁSI-ALBERT MODEL

Number of nodes

$$N = t$$

Number of links

$$N = mt$$

Average Degree

$$\langle k \rangle = 2m$$

Degree dynamics

$$k_i(t) = m (t/t_i)^\beta$$

Dynamical exponent

$$\beta = 1/2$$

Degree distribution

$$p_k \sim k^\gamma$$

Degree exponent

$$\gamma = 3$$

Average distance

$$\langle d \rangle \sim \log N / \log \log N$$

Clustering coefficient

$$\langle C \rangle \sim (\ln N)^2 / N$$

HOMEWORK

1. Calculate the degree distribution of the directed Barabási-Albert model. That is, in each time a new node arrives, that connects with a directed link to a node chosen with preferential attachment Eq. 5.1, where $n(k_{in})$ depends only the node's in-degree. Discuss both the in and out-degree distribution of the resulting network.
2. Use the rate equation approach described above that the directed copying model leads to a scale-free network with the incoming degree exponent $\gamma_{in} = (2 - p) / (1 - p)$, hence the degree exponent varies between $\gamma_{in} = 2$ for $p \rightarrow 0$ and $\gamma_{in} = \infty$ for $p \rightarrow 1$.