Structural Properties of Scale-Free Networks

Reuven Cohen, Shlomo Havlin, and Daniel ben-Avraham

WILEY-VCH Verlag Berlin GmbH
August 18, 2002

1 Structural Properties of Scale-Free Networks

Reuven Cohen and Shlomo Havlin

Minerva Center and Department of Physics,
Bar-Ilan University, Ramat-Gan, Israel

Daniel ben-Avraham

Physics Department and Center for Statistical Physics (CISP),
Clarkson University, Potsdam NY 13699-5820, USA

Abstract

Many networks have been reported recently to follow a scale-free degree distribution in which the fraction of sites having k connections follows a power law: $P(k) = ck^{-\lambda}$. In this chapter we study the structural properties of such networks. We show that the average distance between sites in scale-free networks is much smaller than that in regular random networks, and bears an interesting dependence on the degree exponent λ. We study percolation in scale-free networks and show that in the regime $2 < \lambda < 3$ the networks are resilient to random breakdown and the percolation transition occurs only in the limit of extreme dilution. On the other hand, attack of the most highly connected nodes easily disrupts the nets. We compute the percolation critical exponents and find that percolation in scale-free networks is non-universal, i.e. depends on λ and different from the mean-field behavior in dimensions $d \geq 6$. Finally, we suggest a novel and efficient method for immunization against the spread of diseases in social networks, or the spread of viruses and worms in computer networks.

1.1 Introduction

1.1.1 Random graphs

Graph theory is rooted in the 18th century beginning with the work of Euler. A graph in its mathematical definition is a pair of sets (V, E), where V is a set of vertices (the nodes of the graph), and E is a set of edges, denoting the links between the vertices. In a directed graph, the edges are taken as ordered pairs, *i.e.*, each edge is directed from the first to the second vertex of the pair.

The work on graph theory has mainly dealt with the properties of special graphs. In the 1960s, Paul Erdős and Alfréd Rényi initiated the study of random graphs [1, 2, 3]. Random graph theory is, in fact, not the study of graphs (as there is no such thing as a "random graph"), but the study of an ensemble of graphs (or, as mathematicians prefer to call it, a *probability*

space of graphs). The ensemble is a class consisting of many different graphs, where each graph has a probability attached to it. A property studied is said to exist with probability P if the total probability of all the graphs in the ensemble of having that property is P. This structure allows the use of probability theory in conjunction with discrete mathematics for the study of graph ensembles.

Two well-studied graph ensembles are $G_{N,M}$ — the ensemble of all graphs having N vertices and M edges, and $G_{N,p}$ — consisting of graphs with N vertices, where each possible edge is realized with probability p. These two families, initially studied by Erdős and Rényi themselves, are known to be similar if $M = \binom{N}{2}p$, so long as p is not too close to 0 or 1 [4], and are referred to as ER graphs. Examples of other well-studied ensembles are the family of *regular* graphs, where all nodes have the same number of edges, $P(k) = \delta_{k,k_0}$, and the family of *unlabeled* graphs, where graphs which are isomorphic under permutations of their nodes are considered the same object.

An important attribute of a graph is the average degree, *i.e.*, the average number of edges connected to each node. We shall denote the degree of the ith node by k_i and the average degree by $\langle k \rangle$. N-vertex graphs with $\langle k \rangle = \mathcal{O}(N^0)$ are called sparse graphs. In what follows, we concern ourselves exclusively with sparse graphs.

An interesting characteristic of the ensemble $G_{N,p}$ is that many of its properties have a related threshold function, $p_t(N)$, such that if $p < p_t$ the property exists with probability 0, in the "thermodynamic limit" of $N \to \infty$, and with probability 1 if $p > p_t$. This phenomenon is similar to the physical notion of a phase transition. An example of such a property is the existence of a giant component, *i.e.*, a set of connected nodes, in the sense that a path exists between any two of them, whose size is proportional to N. Erdős and Rényi showed [2] that for ER graphs such a component exists if $\langle k \rangle > 1$. If $\langle k \rangle < 1$ only small components exist, and the size of the largest component is proportional to $\ln N$. Exactly at the threshold, $\langle k \rangle = 1$, a component of size proportional to $N^{2/3}$ emerges. This phenomenon was described by Erdős as the "double jump". Another property is the average path length distance between any two sites, which in almost every graph of the ensemble is of order $\ln N$.

1.1.2 Scale-free networks

The Erdős-Rényi model has been traditionally the dominant subject of study in the field of random graphs. Recently, however, several studies of real world networks have indicated that the ER model fails to reproduce many of their observed properties.

One of the simplest properties of a network that can be measured directly is the degree distribution, or the fraction $P(k)$ of nodes having k connections (degree k). A well-known result for ER networks is that the degree distribution is Poissonian, $P(k) = e^{-z}z^k/k!$, where $z = \langle k \rangle$ is the average degree [4].

Direct measurements of the degree distribution for networks of the Internet [5, 6], WWW (where hypertext links constitute directed edges) [7, 8], e-mail network [9], citations of scientific articles [10], metabolic networks [11, 12], trust network [13], and many more, show that the Poisson law does not apply. Rather, most often these nets exhibit a scale-free degree distribution:

$$P(k) = ck^{-\lambda}, \qquad k = m, ..., K \tag{1.1}$$

where $c \approx (\lambda - 1)m^{\lambda-1}$ is a normalization factor, and m and K are the lower and upper cutoffs for the connectivity of a node, respectively. The divergence of moments higher than $\lceil \lambda - 1 \rceil$ (as $K \to \infty$ when $N \to \infty$) is responsible for many of the special properties attributed to scale-free networks.

All real-life networks are finite (and all their moments are finite). The actual value of the cutoff K plays an important role. It may be approximated by noting that the total probability of nodes with $k > K$ is of order $1/N$ [14, 15]:

$$\int_K^\infty P(k)\, dk \sim 1/N. \tag{1.2}$$

This yields the result

$$K \sim mN^{1/(\lambda-1)}. \tag{1.3}$$

The degree distribution does not characterize the graph or ensemble in full. There are other quantities, such as the degree-degree correlation (between connected sites), the spatial correlations, etc. Several models have been presented for the evolution of scale-free networks, each of which may lead to a different ensemble. The first suggestion was the *preferential attachment* model by Barabási and Albert, which came to be known as the "Barabási-albert model" [5] . Several variants have been suggested to this model (see, *e.g.*, [16, 17]). In this Chapter we will concentrate on the "Molloy-Reed construction" [18, 19, 20], which ignores the evolution and assumes only the degree distribution and no correlations between nodes. Thus, the site reached by following a link is independent of the origin.

Scale-free distributions have been studied in physics, particularly in the context of fractals and of Lévy flights. Fractals are objects which appear similar (at least in some statistical sense) at every lengthscale [21, 22]. Many natural objects, such as mountains, clouds, coastlines and rivers, as well as the cardiovascular and nervous systems are known to be fractals. This is why we find it hard to distinguish between a photograph of a mountain and that of part of the mountain, neither can we ascertain the altitude from which a picture of a coastline had been taken. Diverse phenomena, such as the distribution of earthquakes, biological rhythms and rates of transport of data packets in communication networks, are also known to posses a scale-free distribution. They come in all sizes and rhythms, spanning many orders of magnitude.

Lévy flights were suggested by Paul Lévy [23], who was studying what is now known as Lévy stable distributions. The question he asked was, When is the length distribution of a single step in a random walk similar to that of the entire walk? Besides the known result, that of the Gaussian distribution, Lévy found an entire new family — essentially that of scale-free distributions. Stable distributions do not obey the central limit theorem (stating that for large numbers of steps the distribution of the total displacement tends to Gaussian), due to the divergence of the variance of indivual steps. Lévy walks have numerous applications [24, 25]. An interesting observation is that animal foraging patterns which follow stable distributions have been shown to be their most efficient strategy [26, 27]. For recent reviews on complex networks and in particular scale free networks see Refs. [28, 29].

1.2 Small and Ultra-Small Worlds

Regular lattices are embedded in Euclidean space, of a well-defined dimension, d. This means that $n(r)$, the number of sites within a distance r from an origin, grows as $n(r) \sim r^d$ (for large r). For fractal objects d in the last relation is non-integer and is replaced by the fractal dimension $d_{\rm f}$. Similarly, the chemical dimension, d_l, is defined by the scaling of the number of sites within l edges or less from a given site (an origin), $n(l) \sim l^{d_l}$. A third dimension, $d_{\rm min}$, relates between the chemical path (the shortest distance along edges) and Euclidean distances, $l \sim r^{d_{\rm min}}$. It satisfies $d_{\rm min} = d_{\rm f}/d_l$ [21, 22, 30].

An example of an object where these concepts fail is the Cayley tree (also known as the Bethe lattice). The Cayley tree is a regular graph, of fixed degree z, and no loops. It has been studied by physicists in many contexts, since its simplicity often allows for exact analyses. An infinite Cayley tree cannot be embedded in a Euclidean space of finite dimensionality. The number of sites at l is $n(l) \sim (z-1)^l$. Since the exponential growth is faster than any power-law, Cayley trees are referred to as infinite-dimensional systems.

In most random network models the structure is locally tree-like (since most loops occur only for $n(l) \sim N$), and, since the number of sites grows as $n(l) \sim \langle k-1 \rangle^l$, they are also infinite-dimensional. As a consequence, the diameter of such graphs (*i.e.*, the minimal path between the most distant nodes) scales like $D \sim \ln N$ [4]. This small diameter is to be contrasted with that of finite-dimensional lattices, where $D \sim N^{1/d_l}$.

Recently, a model has been suggested by Watts and Strogatz [31, 32] which retains the local high clustering of lattices while reducing the diameter to $D \sim \ln N$. This, so called, small world network is achieved by replacing a fraction ϕ of the links in a regular lattice with random links, to random distant neighbors. (In other variants of the small world model the "long range" links are simply added on, without prior removal of lattice links.) A study of scale-free networks embedded in Euclidean space (at the obvious price of a cutoff in k) which exhibit finite dimensions can be found in [33].

1.2.1 Diameter of scale-free networks

We now aim to show that scale-free networks with degree exponent $2 < \lambda < 3$ possess a diameter $D \sim \ln\ln N$, smaller even than that of ER and small world networks. If the network is fragmented, we will only be interested in the diameter of the largest cluster (assuming there is one). Our analysis of the diameter of the Molloy-Reed scale-free networks is based on [34].

We adopt a different definition of diameter: the *average* distance between any two sites on the graph. We find it easier still to focus on the radius of a graph, $L \equiv \langle l \rangle$: the average distance of all sites from the site of highest degree in the network (if there is more than one, we pick one arbitrarily). The diameter of the graph, D, is restricted to:

$$L \leq D \leq 2L, \tag{1.4}$$

and thus essentially scales like L.

1.2.2 Minimal graphs and lower bound

We begin by showing that the radius of any scale-free graph with $\lambda > 2$ has a rigorous lower bound that scales as $\ln\ln N$. It is easy to convince oneself that the smallest diameter of a graph, of a given degree distribution, is achieved by the following construction: Start with the highest degree site, then connect to each successive layer the extant sites of highest degree, until the layer is full. By construction, loops will occur only in the last layer.

Let the number of links outgoing from the lth shell (layer) be χ_l. Let K_l denote the highest degree of a site not yet reached by the lth layer. Then, for the graph of minimal diameter described above,

$$\chi_l = N \int_{K_{l+1}}^{K_l} P(k)dk \approx m^{\lambda-1} N K_{l+1}^{1-\lambda}. \tag{1.5}$$

The number of links outgoing from layer $l+1$ equals the total number of links in all the sites between K_l and K_{l+1} minus one link for each site — the one used to connect to the previous layer:

$$\chi_{l+1} = N \int_{K_{l+1}}^{K_l} (k-1)P(k)dk \approx \frac{\lambda-1}{\lambda-2} m^{\lambda-1} N K_{l+1}^{2-\lambda}. \tag{1.6}$$

Solving these recursion relations, with the initial conditions $K_0 = N^{1/(\lambda-1)}$ and $\chi_0 = K_0$, leads to:

$$\chi_l = a^{(\lambda-1)(1-u^l)} N^{1-u^{l+1}}, \tag{1.7}$$

where $a = (\lambda-1)/(\lambda-2)m$, $u = (\lambda-2)/(\lambda-1)$, and

$$K_l = m(\chi_l/N)^{\frac{1}{1-\lambda}} . \tag{1.8}$$

To bound the radius L of the graph, we will assume that the low degree sites are connected randomly to the giant cluster. We pick a site of degree $1 \ll k^* \ll (\ln\ln N)^{1/(\lambda-1)}$. Using Eq. (1.8) we can show that if $l_1 \approx \ln\ln N/\ln(\lambda-2)$ then $K_{l_1} < k^*$, so, with probability 1 all sites of degree $k \geq k^*$ lie within l_1 layers from the site we picked. On the other hand, if we start uncovering the graph from any site — provided it belongs to the giant component — then within a distance l_2 from this site there are at least l_2 bonds. The probability that none of those bonds leads to our site (of degree k^*) is $(1 - P(k^*)k^*/\langle k\rangle)^{l_2}$. That is, if $l_2 k^* P(k^*)/\langle k\rangle \gg 1$, at least one bond will lead to our site. Thus, taking $k^{*\lambda-1} \ll l_2 \ll \ln\ln N$, we will definitely reach a site of at least degree k^* in the l_2th layer from almost any site. Since $l = l_1 + l_2$, all sites are at a distance of order $\ln\ln N$ from the highest degree site, and $L \sim \ln\ln N$ is a rigorous lower bound for the diameter of scale-free networks with $\lambda > 2$.

1.2.3 The general case of random scale-free networks

We now argue that the scaling of $D \sim \ln\ln N$ is actually realized in the general case of *random* scale-free graphs with $2 < \lambda < 3$. One can view the process of uncovering the

network as actually building it, by following the links one at a time. For simplicity, let us start with the site of highest degree, $K \sim N^{1/(\lambda-1)}$ (guaranteed to belong to the giant component). Next, we expose the layers one at a time. We view the graph as built from one large developing cluster, and sites which have not yet been reached (they might or might not belong to the giant component), see Fig. 1.1. A similar consideration has been used by Molloy and Reed [19].

After l layers are explored the distribution of the yet unreached sites changes (since most high-degree sites are exposed first) to $P'(k) \approx P(k)exp(-k/K_l)$ [19].

Let us now consider layer $l+1$. A threshold function emerges: the new distribution of unvisited sites behaves like a step function — almost $P(k)$ for $k < K_{l+1}$, and 0 for $k > K_{l+1}$. The reason for this is as follows. A site with degree k has a probability of $p = k/(N\langle k\rangle)$ to be reached by following a link[1]. If there are χ_l outgoing links then for $p\chi_l > 1$ we can assume that, in the limit $N \to \infty$, the site will be reached in the next level with probability 1. Therefore, all unvisited sites with degree $k > N/\chi_l$ will be surely reached in the next chemical layer. On the other hand, almost all the unvisited sites with degree $k < N/\chi_l$ will remain unvisited in the next layer and their distribution will remain virtually unchanged. From these considerations, the highest degree of the unexplored sites in layer $l+1$ is determined by:

$$K_{l+1} \approx N/\chi_l. \tag{1.9}$$

In layer $l+1$ all sites with degree $k > N/\chi_l$ will be exposed. Since the probability of reaching a site via a link is proportional to $kP(k)$, the average degree of sites reached by following a link is $\kappa \equiv \langle k^2\rangle/\langle k\rangle$ [14]. κ for layer l can be computed from the general formula [14], valid for scale-free distributions,

$$\kappa = \left(\frac{\lambda-2}{\lambda-3}\right)\left(\frac{K^{3-\lambda}-m^{3-\lambda}}{K^{2-\lambda}-m^{2-\lambda}}\right), \tag{1.10}$$

but with the layer cutoff K of (1.9). That is, $\kappa_l \sim K_{l+1}^{3-\lambda}$.

Using the above consideration, the number of outgoing links from layer $l+1$ can be computed. Consider the total degree of all sites reached in the $l+1$ level. This includes all sites with degree k, $K_{l+1} < k < K_l$, as well as other sites with average degree proportional to $\kappa - 1$ (the -1 is due to one link going into the shell). Thus, the value of κ (the avergae number of links for a site reached via a link) is calculated using the cutoff K_{l+1}. (Loops within a layer, and multiple links connecting a site in layer $l+1$, can be neglected as long as the number of sites in the layer is less than order N, $N \to \infty$.) The two contributions can be written as the sum of two terms:

$$\chi_{l+1} \approx N\int_{K_{l+1}}^{K_l}(k-1)P(k)dk + \chi_l\left[\kappa(K_{l+1})-1\right]. \tag{1.11}$$

Noting that $P(k) \propto k^{-\lambda}$ and that $\kappa \propto K^{3-\lambda}$ [14], it follows that $\chi_{l+1} \propto NK_{l+1}^{2-\lambda}$ (note that both terms in Eq. (1.11) scale similarly). This results in a second recurrence equation:

$$\chi_{l+1} = ANK_l^{2-\lambda}, \tag{1.12}$$

where $A = \frac{\lambda-1}{\lambda-2}m^{\lambda-1} + \frac{\lambda-2}{3-\lambda}m^{\lambda-2}$.

[1] We assume that $\langle k\rangle$ for the unvisited sites is fixed, since it is dominated by the low-degree nodes, whose distribution is unchanged.

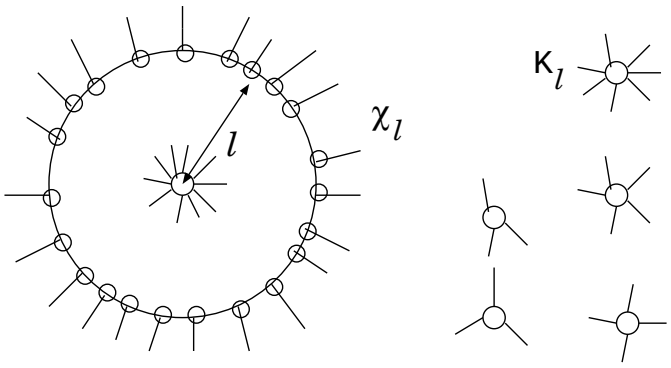

Figure 1.1: Illustration of the exposure process. The large circle denotes the exposed fraction of the giant component, while the small circles denote individual sites. The sites on the right have not been reached yet. After [34]

Solving the equations (1.9) and (1.12) yields

$$\chi_l \sim A^{\frac{(\lambda-2)^l-1}{\lambda-3}} N^{1-\frac{(\lambda-2)^l+1}{\lambda-1}}, \tag{1.13}$$

where χ_l is the number of outgoing links from the lth layer. Eq. (1.9) then leads to:

$$K_l \sim A^{\frac{(\lambda-2)^{l-1}-1}{3-\lambda}} N^{\frac{(\lambda-2)^l}{\lambda-1}}. \tag{1.14}$$

Using the same considerations that follow Eq. (1.8), one can deduce that also here

$$D \sim \ln\ln N. \tag{1.15}$$

Our result that $D \sim \ln\ln N$ is consistent with the observations that the distance in the Internet network is extremely small, and that the distance in metabolic scale-free networks is almost independent of N [11].

For $\lambda > 3$ and $N \gg 1$, κ is independent of N, and since the second term of Eq. (1.11) is dominant, Eq. (1.11) reduces to $\chi_{l+1} = (\kappa-1)\chi_l$, where κ is a constant depending only on λ. This leads to the known result $\chi_l \approx C(N,\lambda)(\kappa-1)^l$ and the radius of the network [35] is

$$L \sim \ln N. \tag{1.16}$$

For $\lambda = 3$, Eq. (1.11) reduces to $\chi_{l+1} = \chi_l \ln \chi_l$. Taking the logarithm of this equation one obtains $\ln \chi_{l+1} - \ln \chi_l = \ln\ln \chi_l$. Defining $g(l) = \ln \chi_l$ one obtains a difference equation which may be approximated (in the continuum limit) by $g' = \ln g$. Substituting $u = \ln g$, the equation reduces to

$$L = \int_{\ln\ln\sqrt{N}}^{\ln\ln N} e^{u-\ln u} du. \tag{1.17}$$

The lower bound is obtained from the highest degree site for $\lambda = 3$, having degree $K = m\sqrt{N}$. Thus, $\chi_0 = m\sqrt{N}$. The upper bound results from that l for which $\chi_l \sim N$ (and lower order corrections). The integral in Eq. (1.17) can be approximated by the steepest descent method, leading to

$$L \sim \ln N/(\ln \ln N), \tag{1.18}$$

(assuming $\ln \ln N \gg 1$).

The result of Eq. (1.18) has been obtained rigorously for the maximum distance in the Barabasi-Albert (BA) model [5], where $\lambda = 3$ (for $m \geq 2$) [36]. Although the result in [36] applies to the largest distance between two sites, their derivation leaves no doubt that the average distance would behave similarly. For $m = 1$, the graphs in the Barabasi-Albert model turn into trees, and the behavior of $D \sim \ln N$ is obtained [36, 37]. It should be noted that for $m = 1$ the giant component in the random model contains only a fraction of the sites (while for $m \geq 2$ it contains all sites — at least to leading order). This might explain why exact trees and BA trees are different from Molloy-Reed random graphs.

1.3 Percolation

Since the 1940s percolation has been the subject of intense studies among physicists and mathematicians. In site percolation, usually defined on lattices, the sites (nodes) are present (or *occupied*, or *wet*) with probability q, or equivalently, removed (blocked) with probability $p = 1 - q$. The (infinite) network undergoes a sharp phase transition at a critical threshold q_c, from a connected, or percolating phase, where a spanning cluster runs across the entire size of the system, for $q > q_c$, to a fragmented phase, where only finite clusters exist, for $q < q_c$. An introduction to the general subject of percolation can be found in [38, 22, 30][2]

The percolation transition is continuous (second order), and near the transition point many properties behave as power laws. For example, the probability for a site to be in the spanning cluster (for $q > q_c$) grows as $P_\infty \sim (q - q_c)^\beta$, and the number of clusters of size s, at $q = q_c$, is $n_s \sim s^{-\tau}$. The critical exponents β, τ, and their likes, are universal, depending only upon the dimensionality d of the lattice. For $d \geq d_c = 6$ the lattice dimension no longer plays a significant role and the critical exponents assume their mean-field values (*e.g.*, $\beta = 1$, $\tau = 5/2$). The mean-field case can be conviniently obtained from the exact solution of the percolation problem on Cayley trees (whose effective dimension is infinite). Percolation in ER graphs also follows mean-field behavior. The percolation threshold is $q_c = 1 - p_c = 1/(z-1)$ for Cayley trees, and $q_c = 1/\langle k \rangle$ for ER graphs.

The problem of percolation on scale-free networks has important practical applications. Below we explore consequences to the resilience of the Internet in the face of random breakdown of servers as well as under intentional attack, and to immunization strategies against the spread of contagious epidemics in population and computer networks.

[2] We have exchanged here the traditional roles of p and q in percolation theory, $p \leftrightarrow q$, to conform with what seems to be the norm in papers on scale-free graphs.

1.3.1 Random breakdown

For a graph having degree distribution $P(k)$ to have a spanning cluster, a site j which is reached by following a link (from site i on) the giant cluster must have at least one other link, on average, to allow the cluster to exist. For this to happen the average degree of site j must be at least 2 (one incoming and one outgoing link), given that site i is connected to j:

$$\langle k_i | i \leftrightarrow j \rangle = \sum_{k_i} k_i P(k_i | i \leftrightarrow j) = 2. \tag{1.19}$$

Using Bayes' rule we get

$$P(k_i | i \leftrightarrow j) = P(k_i, i \leftrightarrow j)/P(i \leftrightarrow j) = P(i \leftrightarrow j | k_i) P(k_i)/P(i \leftrightarrow j), \tag{1.20}$$

where $P(k_i, i \leftrightarrow j)$ is the *joint* probability that node i has degree k_i and that it is connected to node j. For randomly connected networks (neglecting loops) $P(i \leftrightarrow j) = \langle k \rangle/(N-1)$ and $P(i \leftrightarrow j | k_i) = k_i/(N-1)$, where N is the total number of nodes in the network. Using the above criterion Eq. (1.19) reduces to [18, 14]:

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = 2, \tag{1.21}$$

at the critical point. A spanning cluster exists for graphs with $\kappa > 2$, while graphs with $\kappa < 2$ contain only small clusters whose size s is negligible compared to the entire network, $\lim_{N\to\infty} s/N = 0$. The criterion (1.21) and its range of validity was derived rigorously by Molloy and Reed [18], using somewhat different arguments.

Neglecting the loops is justified below the transition, since the probability for a bond to form a loop in an s-node cluster is proportional to $(s/N)^2$ (*i.e.*, proportional to the probability of choosing two sites in that cluster). An estimate of the fraction of loops P_{loop} in the network yields

$$P_{loop} \propto \sum_i \frac{s_i^2}{N^2} < \sum_i \frac{s_i S}{N^2} = \frac{S}{N}, \tag{1.22}$$

where the sum is over all clusters in the system (s_i is the size of the ith cluster), and S is the size of the biggest cluster. Since $S \sim \ln N$ below the transition, P_{loop} is negligible when $N \to \infty$.

1.3.2 Percolation critical threshold

The above reasoning can be applied to the problem of percolation in a generalized random network [14]. If we randomly remove a fraction p of the sites (along with the emanating links), the degree distribution of the remaining sites will change. For instance, sites with initial degree k_0 will have, after the random removal of nodes, a different number of connections k, depending on the number of removed neighbors. The initial degree distribution, $P_0(k_0)$, becomes, following dilution

$$P(k) = \sum_{k_0=k}^{\infty} P_0(k_0) \binom{k_0}{k} (1-p)^k p^{k_0-k}. \tag{1.23}$$

The first two moments of the new $P(k)$ are

$$\langle k\rangle = \sum_{k=0}^{\infty} P(k)k = (1-p)\langle k_0\rangle, \tag{1.24}$$

and

$$\langle k^2\rangle = \sum_{k=0}^{\infty} P(k)k^2 = (1-p)^2\langle k_0^2\rangle + p(1-p)\langle k_0\rangle, \tag{1.25}$$

where $\langle k_0\rangle$ and $\langle k_0^2\rangle$ are computed with respect to the original distribution $P_0(k_0)$. On substituting the new moments in Eq. (1.21) we obtain the criterion for criticality, following dilution:

$$\kappa \equiv \frac{\langle k^2\rangle}{\langle k\rangle} = \frac{(1-p)^2\langle k_0^2\rangle + p(1-p)\langle k_0\rangle}{(1-p)\langle k_0\rangle} = 2. \tag{1.26}$$

This can be rearranged, to yield the critical threshold for percolation [14]:

$$1 - p_{\rm c} = \frac{1}{\kappa_0 - 1}, \tag{1.27}$$

where $\kappa_0 \equiv \langle k_0^2\rangle/\langle k_0\rangle$ is calculated using the original distribution, before the random removal of sites.

Eqs. (1.21) and (1.27) are applicable to random graphs of arbitrary degree distribution. For example, using (1.27) for Cayley trees yields the well-known threshold [38, 22] $q_{\rm c} = 1 - p_{\rm c} = 1/(z-1)$. Another example is ER graphs. Their edges are distributed randomly and the resulting degree distribution is Poissonian [4]. Applying the criterion from Eq. (1.21) to the Poisson distribution of ER graphs yields

$$\kappa \equiv \frac{\langle k^2\rangle}{\langle k\rangle} = \frac{\langle k\rangle^2 + \langle k\rangle}{\langle k\rangle} = 2, \tag{1.28}$$

which reduces to the known result [4] $\langle k\rangle = 1$.

Evidently, the key parameter governing the threshold, according to (1.27), is the ratio of second- to first-moment, κ_0. This may be estimated by approximating (1.1) to a continuous distribution (the approximation becomes exact for $1 \ll m \ll K$, and it preserves the essential features of the transition even for small m):

$$\kappa_0 = \left(\frac{2-\lambda}{3-\lambda}\right)\frac{K^{3-\lambda} - m^{3-\lambda}}{K^{2-\lambda} - m^{2-\lambda}}. \tag{1.29}$$

In the limit of $K \gg m$, we have

$$\kappa_0 \to \left|\frac{2-\lambda}{3-\lambda}\right| \times \begin{cases} m, & \lambda > 3; \\ m^{\lambda-2}K^{3-\lambda}, & 2 < \lambda < 3; \\ K, & 1 < \lambda < 2. \end{cases} \tag{1.30}$$

We see that for $\lambda > 3$ the ratio κ_0 is finite and there is a percolation transition at $1 - p_c \approx \left(\frac{\lambda-2}{\lambda-3}m - 1\right)^{-1}$: for $p > p_c$ the spanning cluster is fragmented and the network is destroyed. However, for $\lambda < 3$ the ratio κ_0 diverges with K and so $p_c \to 1$ when $K \to \infty$ (or $N \to \infty$). The percolation transition does not take place: a spanning cluster exists for arbitrarily large fractions of dilution, $p < 1$. In *finite* systems a transition is always observed, though for $\lambda < 3$ the transition threshold is exceedingly high. For the case of the Internet ($\lambda \approx 5/2$), we have $\kappa_0 \approx K^{1/2} \approx N^{1/3}$. Considering the enormous size of the Internet, $N > 10^6$, one needs to destroy over 99% of the nodes before the spanning cluster collapses. For $\lambda > 4$ calculation of κ shows that it is lower than 2 even before the breakdown occurs. For $\lambda > 4$ and $m = 1$ the network consists of only finite clusters and no spanning cluster to begin with (this is reminiscent of the result for $\lambda > 3.478...$ found in [20], where the different threshold stems from rigorous consideration of the *discrete* nature of (1.1)). Note that if $m \geq 2$, a spanning cluster exists for all values of λ.

1.3.3 Generating functions

A general method for studying the size of the infinite cluster and the residual network for a graph with an arbitrary degree distribution was first found by Molloy and Reed [19]. They consider the infinite cluster as it is being exposed, layer by layer, and develop differential equations relating the number of unexposed links and unvisited sites in subsequent shells (see Section **1.2**).

An alternative, and very powerful approach based on generating functions was advanced by Newman, Watts and Strogatz [35]. Their method is beautifully reviewed in this book, in the Chapter by Newman. Here, we follow closely in their footsteps. Our ultimate goal is to compute the size of the giant component as well as the critical exponents associated with the percolation transition in scale-free networs, resulting from random dilution (Section **1.3.5**).

In [35, 39] a generating function is constructed for the degree distribution:

$$G_0(x) = \sum_{k=0}^{\infty} P(k)x^k. \tag{1.31}$$

The probability of reaching a site with degree k by following a specific link is $kP(k)/\langle k\rangle$ [18, 14, 35], and its corresponding generating function is

$$G_1(x) = \frac{\sum kP(k)x^{k-1}}{\sum kP(k)} = \frac{d}{dx}G_0(x)/\langle k\rangle \ . \tag{1.32}$$

Let $H_1(x)$ be the generating function for the probability of reaching a branch of a given size by following a link. When a fraction $p = 1 - q$ of the sites are diluted, H_1 satisfies the self-consistent equation:

$$H_1(x) = 1 - q + qxG_1(H_1(x)) \ . \tag{1.33}$$

Given that $G_0(x)$ is the generating function for the degree of a site, the generating function for the probability of a site to belong to an n-site cluster is:

$$H_0(x) = 1 - q + qxG_0(H_1(x)) \ . \tag{1.34}$$

Below the transition, all clusters are finite and $H_0(1) = 1$. However, above the transition $H_0(1)$ is no longer normalized, since it excludes the probability of the incipient infinite cluster, P_∞. In other words, $P_\infty = 1 - H_0(1)$. It follows that

$$P_\infty(q) = q\big(1 - \sum_{k=0}^{\infty} \tilde{P}(k) u^k\big) \ , \tag{1.35}$$

where $u \equiv H_1(1)$ is the smallest positive root of

$$\langle k \rangle u = 1 - q + \frac{q}{\langle k \rangle} \sum_{k=0}^{\infty} k P(k) u^{k-1} \ . \tag{1.36}$$

This equation can be solved numerically and the solution can be substituted into Eq. (1.35) to compute the size of the infinite cluster in a random graph of arbitrary degree distribution. (The analysis neglects the presence of loops, which is well justified near criticality.)

1.3.4 Intentional attack

Another model of interest, suggested in [40], is that of intentional attack on the most highly connected nodes of the network. In this model an attacker (*e.g.*, computer hackers trying to cause damage to the network, or doctors trying to disrupt a contagious epidemic) succeeds in knocking off a fraction p of the most highly connected sites in the network. As might be expected, such a strategy is far more effective than *random* dilution. We shall see, in fact, that a *small* threshold p suffices to disrupt the net (for all λ).

Next we consider analytically the consequence of such an attack, or sabotage, on scale-free networks [41]. A different approach was given independently in [39]: (a) the cutoff degree K reduces to some new value $\tilde{K} < K$, and (b) the degree distribution of the remaining sites is no longer scale-free, but is changed due to the removal of many of their links. Recall that the upper cutoff K, before the attack, may be estimated from

$$\sum_{k=K}^{\infty} P(k) = \frac{1}{N} \ . \tag{1.37}$$

Similarly, the new cutoff $\tilde{K}$, after the attack, follows from

$$\sum_{k=\tilde{K}}^{K} P(k) = \sum_{k=\tilde{K}}^{\infty} P(k) - \frac{1}{N} = p \ . \tag{1.38}$$

If the size of the system is large, $N \gg 1/p$, the original cutoff K may be safely ignored. We can then obtain $\tilde{K}$ approximately by replacing the sum with an integral:

$$\tilde{K} = m p^{1/(1-\lambda)} \ . \tag{1.39}$$

We estimate the impact of the attack on the distribution of the remaining sites as follows. The removal of a fraction p of the sites with the highest degree results in a random removal of

links from the remaining sites — links that had connected the removed sites with the remaining sites. The probability $\tilde{p}$ for a link to lead to a deleted site equals the ratio of the number of links belonging to deleted sites to the total number of links:

$$\tilde{p} = \sum_{k=\tilde{K}}^{K} \frac{kP(k)}{\langle k_0 \rangle}, \tag{1.40}$$

where $\langle k_0 \rangle$ is the initial average degree. With the usual continuous approximation, and neglecting K, this yields

$$\tilde{p} = \left(\frac{\tilde{K}}{m}\right)^{2-\lambda} = p^{(2-\lambda)/(1-\lambda)} , \tag{1.41}$$

for $\lambda > 2$. For $\lambda = 2$, $\tilde{p} \to 1$, since just a few nodes of very high degree control the entire connectedness of the system. Indeed, consider a finite system of N sites and $\lambda = 2$. The upper cutoff $K \approx N$ must then be taken into account, and approximating Eq. (1.40) by an integral yields $\tilde{p} = \ln(Np/m)$. That is, for $\lambda = 2$, a very small value of p is needed to destroy an arbitrarily large fraction of the links as $N \to \infty$.

With the above results we can compute the effect of intentional attack, using the theory previously developed for random removal of sites [14]. Essentially, the network after attack is equivalent to a scale-free network with cutoff $\tilde{K}$, that has undergone random removal of a fraction $\tilde{p}$ of its sites. Since the latter effect influences the probability distribution as described in Eq. (1.23), the result in Eqs. (1.27) and (1.29) can be used, but with $\tilde{p} = (\tilde{K}/m)^{2-\lambda}$ and $\tilde{K}$ replacing p_c and K, respectively. This yields the equation:

$$(\tilde{K}/m)^{2-\lambda} - 2 = \frac{2-\lambda}{3-\lambda} m[(\tilde{K}/m)^{3-\lambda} - 1] , \tag{1.42}$$

which can be solved numerically to obtain $\tilde{K}(m, \lambda)$, and then $p_c(m, \lambda)$ can be retrieved from Eq. (1.39). In Fig. 1.4 we plot p_c (there denoted f_c) — the critical fraction of sites needed to be removed in the sabotage strategy to disrupt the network — computed in this fashion, and compared to results from numerical simulations. A phase transition exists (at a finite p_c) for all $\lambda > 2$. The decline in p_c for large λ is explained from the fact that as λ increases the spanning cluster becomes smaller in size, even before attack. (Furthermore, for $m < 2$ the original network is disconnected for λ large enough.) The decline in p_c as $\lambda \to 2$ results from the critically high degree of just a few sites: their removal disrupts the whole network. This was first argued in [40]. We note that for infinite systems $p_c \to 0$ as $\lambda \to 2$. The critical fraction p_c is rather sensitive to the lower degree cutoff m. For larger m the networks are more robust, though they still undergo a transition at a finite p_c. (Fig. 1.4 illustrates the case of $m = 1$.)

1.3.5 Critical exponents

The generating functions method has the advantage of turning a combinatorial problem into an algebraic one, concerning power series. The algebraic problem is often simpler.

Here we use generating functions for obtaining the percolation critical exponents [42]. These can be extracted from the resulting power series by means of appropriate Abelian and Tauberian methods [43, 44].

First, we compute the order parameter critical exponent β. Near criticality the probability of belonging to the spanning cluster behaves as $P_\infty \sim (q - q_c)^\beta$. For infinite-dimensional systems (such as a Cayley tree) it is known that $\beta = 1$ [38, 22, 30]. This regular mean-field result is not always valid, however, for scale-free networks. Eq. (1.35) has no special behavior at $q = q_c$; the singular behavior comes from u. At criticality, $P_\infty = 0$ and Eq. (1.35) imply that $u = 1$. We therefore examine Eq. (1.36) for $u = 1 - \epsilon$ and $q = q_c + \delta$:

$$1 - \epsilon = 1 - q_c - \delta + \frac{(q_c + \delta)}{\langle k \rangle} \sum_{k=0}^{\infty} kP(k)(1 - \epsilon)^{k-1}. \tag{1.43}$$

The sum in (1.43) has the asymptotic form

$$\begin{aligned} \sum_{k=0}^{\infty} kP(k)u^{k-1} &\sim \langle k \rangle - \langle k(k-1) \rangle \epsilon \\ &+ \frac{1}{2}\langle k(k-1)(k-2) \rangle \epsilon^2 + \cdots + c\Gamma(2-\lambda)\epsilon^{\lambda-2} , \end{aligned} \tag{1.44}$$

where the highest-order analytic term is $\mathcal{O}(\epsilon^n)$, $n = \lfloor \lambda - 2 \rfloor$. Using this in Eq. (1.43), with $q_c = 1/(\kappa - 1) = \langle k \rangle / \langle k(k-1) \rangle$, we get

$$\frac{\langle k(k-1) \rangle^2}{\langle k \rangle} \delta = \frac{1}{2}\langle k(k-1)(k-2) \rangle \epsilon + \cdots + c\Gamma(2-\lambda)\epsilon^{\lambda-3} . \tag{1.45}$$

The divergence of δ as $\lambda < 3$ confirms the vanishing threshold for the phase transition in that regime. Thus, limiting ourselves to $\lambda > 3$, and keeping only the dominant term as $\epsilon \to 0$, Eq. (1.45) implies

$$\epsilon \sim \begin{cases} \left(\frac{\langle k(k-1) \rangle^2}{c \langle k \rangle \Gamma(2-\lambda)} \right)^{\frac{1}{\lambda-3}} \delta^{\frac{1}{\lambda-3}} & 3 < \lambda < 4, \\ \frac{2\langle k(k-1) \rangle^2}{\langle k \rangle \langle k(k-1)(k-2) \rangle} \delta & \lambda > 4. \end{cases} \tag{1.46}$$

Returning to P_∞, Eq. (1.35), we see that the singular contribution in ϵ is dominant only for the irrelevant range of $\lambda < 2$. For $\lambda > 3$, we find $P_\infty \sim q_c \langle k \rangle \epsilon \sim (q - q_c)^\beta$.

We see that the order parameter exponent β attains its usual mean-field value only for $\lambda > 4$. Moreover, for $\lambda < 4$ the percolation transition is higher than 2nd-order: for $3 + \frac{1}{n-1} < \lambda < 3 + \frac{1}{n-2}$ the transition is of the nth-order.

For networks with $\lambda < 3$ the transition still exists, though at a vanishing threshold, $q_c = 0$. The sum in Eq. (1.43) becomes:

$$\sum_{k=0}^{\infty} kP(k)u^{k-1} \sim \langle k \rangle + c\Gamma(2-\lambda)\epsilon^{\lambda-2} . \tag{1.47}$$

Using this in conjunction with Eq. (1.36), and remembering that here $q_c = 0$ and therefore $q = \delta$, leads to

$$\epsilon = \Big(\frac{-c\Gamma(2-\lambda)}{\langle k\rangle}\Big)^{\frac{1}{3-\lambda}} \delta^{\frac{1}{3-\lambda}} \ . \tag{1.48}$$

The results can be summarized by

$$\beta = \begin{cases} \frac{1}{3-\lambda} & 2 < \lambda < 3, \\ \frac{1}{\lambda-3} & 3 < \lambda < 4, \\ 1 & \lambda > 4. \end{cases} \tag{1.49}$$

In other words, the transition in $2 < \lambda < 3$ is a mirror image of the transition in $3 < \lambda < 4$. An important difference is that $q_c = 0$ is not λ-dependent in $2 < \lambda < 3$, and the amplitude of P_∞ diverges as $\lambda \to 2$ (but remains finite as $\lambda \to 4$). Some of the results for β have been reported before in [41], and also found independently in a different but related model of virus spreading [45, 46, 47]. The existence of an infinite-order phase transition at $\lambda = 3$ for growing networks of the Albert-Barabási model has been reported in [48, 49]. These examples suggest that the critical exponents are not model-dependent but depend only on λ.

In [35] it was shown that for a random graph of arbitrary degree distribution the finite clusters (of size s) follow the usual scaling form:

$$n_s \sim s^{-\tau} e^{-s/s^*} \ . \tag{1.50}$$

At criticality $s^* \sim |q - q_c|^{-\sigma}$ diverges and the tail of the distribution behaves as a power law. We now derive the exponent τ. The probability that a site belongs to an s-cluster is $p_s = s n_s \sim s^{1-\tau}$, and is generated by H_0:

$$H_0(x) = \sum p_s x^s \ . \tag{1.51}$$

The singular behavior of $H_0(x)$ stems from $H_1(x)$, as can be seen from Eq. (1.34). $H_1(x)$ itself can be expanded from Eq. (1.33), by using the asymptotic form (1.44) of G_1. We let $x = 1-\epsilon$, as before, but work at the critical point, $q = q_c$. With the notation $\phi(\epsilon) = 1 - H_1(1-\epsilon)$, we finally get (note that at criticality $H_1(1) = 1$):

$$-\phi = -q_c + (1-\epsilon)q_c[1 - \frac{\phi}{q_c} + \frac{\langle k(k-1)(k-2)\rangle}{2\langle k\rangle}\phi^2 + \cdots + c\frac{\Gamma(2-\lambda)}{\langle k\rangle}\phi^{\lambda-2}] \ . \tag{1.52}$$

¿From this relation we extract the singular behavior of H_0: $\phi \sim \epsilon^y$. Then, using Tauberian theorems [43] it follows that $p_s \sim s^{-1-y}$, hence $\tau = 2 + y$.

For $\lambda > 4$ the term proportional to $\phi^{\lambda-2}$ in (1.52) may be neglected. The linear term $\epsilon\phi$ may be neglected as well, due to the factor ϵ. This leads to $\phi \sim \epsilon^{1/2}$ and to the usual mean-field result

$$\tau = \frac{5}{2} \ , \qquad \lambda > 4 \ . \tag{1.53}$$

For $\lambda < 4$, the terms proportional to $\epsilon\phi$, ϕ^2 may be neglected, leading to $\phi \sim \epsilon^{1/(\lambda-2)}$ and

$$\tau = 2 + \frac{1}{\lambda - 2} = \frac{2\lambda - 3}{\lambda - 2} , \qquad 2 < \lambda < 4 . \tag{1.54}$$

Note that for $2 < \lambda < 3$ the percolation threshold is strictly $q_c = 0$. In that case we work at $q = \delta$ small but fixed, taking the limit $\delta \to 0$ at the very end. For the case $2 < \lambda < 3$, τ in Eq. (1.54) represents the singularity of the distribution of branch sizes. For the distribution of cluster sizes in this range one has to consider the singularity of x in Eq. (1.34), leading to $\tau = 3$.

For growing networks of the Albert-Barabási model with $\lambda = 3$, it has been shown that $sn_s \propto (s \ln s)^{-2}$ [49]. This is consistent with $\tau = 3$ plus a logarithmic correction. Related results for scale-free trees have been presented in [50].

At the transition point the size of the largest cluster, S, can be obtained from the finite cluster distribution by taking the integral over the tail of the distribution to equal $1/N$. This results in

$$S \propto N^{\tau-1} = N^{(\lambda-2)/(\lambda-1)}. \tag{1.55}$$

For $\lambda = 4$ this reduces to the known $N^{2/3}$ dependence [4]. For $\lambda \to 3$, $S \propto N^{1/2}$. It is not yet clear whether the results have a meaningful interpretation for $\lambda < 3$.

The critical exponent σ, for the cutoff cluster size, may be also derived directly. Finite-size scaling arguments predict [38] that

$$q_c(\infty) - q_c(N) \sim N^{-\frac{1}{d\nu}} = N^{-\frac{\sigma}{\tau-1}} , \tag{1.56}$$

where N is the number of sites in the network, ν is the correlation length critical exponent: $\xi \sim (q - q_c)^{-\nu}$, and d is the dimensionality of the embedding space. Using a continuous approximation of the distribution (1.1) one obtains [14]

$$\kappa \approx \left(\frac{2-\lambda}{3-\lambda}\right) \frac{K^{3-\lambda} - m^{3-\lambda}}{K^{2-\lambda} - m^{2-\lambda}} , \tag{1.57}$$

where, as usual $K \sim N^{1/(\lambda-1)}$. For $3 < \lambda < 4$, this and Eq. (1.27) yield

$$q_c(\infty) - q_c(N) \sim \Delta\kappa \sim K^{3-\lambda} \sim N^{\frac{3-\lambda}{\lambda-1}} , \tag{1.58}$$

which in conjunction with Eq. (1.56) leads to

$$\sigma = \frac{\lambda - 3}{\lambda - 2} , \qquad 3 < \lambda < 4 . \tag{1.59}$$

For $\lambda > 4$ we recover the regular mean-field result $\sigma = 1/2$. Note that Eqs. (1.56), (1.49), (1.54) are consistent with the known scaling relation: $\sigma\beta = \tau - 2$ [38, 22, 30]. For $2 < \lambda < 3$, $q_c(\infty) = 0$ and $q_c(N) \sim K^{\lambda-3} \sim N^{(\lambda-3)/(\lambda-1)}$. Therefore

$$\sigma = \frac{3 - \lambda}{\lambda - 2} , \qquad 2 < \lambda < 3 , \tag{1.60}$$

again consistent with the scaling relation $\sigma\beta = \tau - 2$ (cf Eq. (1.49)).

1.3.6 Fractal dimension

It is well known that on a random network in the well connected regime, the average distance between sites is of order $\log_k N$ [4, 51]. This has also been shown to hold for sgeneral networks [35] and may be even lower for scale-free networks [34]. However, the diluted case is essentially the same as infinite-dimensional percolation. In this case, there is no notion of geometrical distance (since the graph is not embedded in an Euclidean space), but only of chemical distance (the smallest number of edges connecting any two nodes). It is known from infinite-dimensional percolation theory that the fractal dimension at criticality is $d_{\rm f} = 2$ [22]. Therefore the average (chemical) distance $\langle l \rangle$ between pairs of sites on the spanning cluster at criticality behaves as

$$\langle l \rangle \sim \sqrt{M} \,, \tag{1.61}$$

where M is the number of sites in the spanning cluster. This is analogous to percolation in finite dimensions, where in lengthscales smaller than the correlation length the cluster is a fractal with dimension $d_{\rm f}$ and above the correlation length the cluster is homogeneous and has the dimension of the embedding space. In our infinite-dimensional case, the crossover between these two behaviors occurs around the correlation length $\xi \approx |p_{\rm c} - p|^{-1}$.

For $3 < \lambda < 4$ the situation is somewhat different. Below the transition all clusters are finite and almost all finite clusters are trees. The correlation length can be defined using the formula [22]:

$$\xi_l^2 = \frac{\sum l^2 g(l)}{\sum g(l)}. \tag{1.62}$$

The number of sites in the l shell can be seen to be approximately $\langle k \rangle (\kappa - 1)^{l-1}$ [35]. Since $\kappa - 1 = (\kappa_0 - 1)q$ and $q_{\rm c} = 1/(\kappa_0 - 1)$ we get $g(l) = c(1-\delta)^l$, where $\delta = q - q_{\rm c}$. This leads to $\xi_l \sim (q - q_{\rm c})^{-1}$, *i.e.*, $\nu_l = 1$. Above the threshold, the finite clusters can be seen as a random graph with the residual degree distribution of sites not included in the infinite cluster [19]. That is, the degree distribution for sites in the finite clusters is

$$P_r(k) = P(k) u^k, \tag{1.63}$$

where u is the solution of Eq. (1.36). Using this distribution, we can define κ_r for the finite clusters. This adds a term proportional to $\epsilon^{\lambda-3}$ to the expansion of ξ_l. But, since $\delta \propto \epsilon^{\lambda-3}$ (1.46), this leads again to $\nu_l = 1$.

Using ν, the dimension of the network at criticality can be found. The chemical dimension $d_l = 1/\sigma\nu_l$, therefore

$$d_l = \frac{\lambda - 2}{\lambda - 3}. \tag{1.64}$$

Since every path, when embedded in a space above the critical dimension, can be seen as a random walk, it follows that $\nu = \nu_l/2$ [22]. Therefore, the fractal dimension is,

$$d_{\rm f} = \frac{1}{\nu\sigma} = 2\frac{\lambda - 2}{\lambda - 3}. \tag{1.65}$$

The dimension of the embedding space is,

$$d_c = \frac{1}{\nu\sigma(\tau-1)} = 2\frac{\lambda-1}{\lambda-3}. \tag{1.66}$$

d_l, d_f, and d_c of the embedding space reduce to the known values of 2, 4, and 6, respectively, when $\lambda = 4$. Some of the above results have been obtained also by Burda *et al.*, [50].

1.4 Percolation in Directed Networks

Many of the large complex networks of interest, such as the Internet, WWW, electric power grid, cellular, and social networks are directed [52, 28, 29]. For example, in social and economical networks if node A gains information or acquires physical goods from node B, it does not necessarily mean that node B gets similar input from node A. Likewise, most metabolic reactions are one-directional, thus changes in the concentration of molecule A affect the concentration of its product B, but the reverse is not always true. Despite the directedness of many real networks, the modeling literature, with few notable exceptions [35, 53], has focused mainly on undirected networks.

Important aspects of directed networks are captured by their degree distribution, $P(j,k)$, or the probability that an arbitrary node has j incoming and k outgoing edges. Many naturally occurring directed networks, such as the WWW, metabolic networks, citation networks, etc., exhibit a power-law, or *scale-free* degree distribution for the incoming or outgoing links:

$$P_{in(out)}(l) = cl^{-\lambda_{in(out)}}, \quad m \le l \le K\ , \tag{1.67}$$

similar to that of Eq. (1.1) [7, 8].

The structure of a directed graph has been characterized in [35, 53], and in the context of the WWW in [54, 8]. In general, a directed graph consists of a giant weakly connected component (GWCC) and several finite components. In the GWCC every site is reachable from every other, provided that the links are treated as bi-directional. The GWCC is further divided into a giant strongly connected component (GSCC), consisting of all sites reachable from each other following directed links. All the sites reachable from the GSCC are referred to as the giant OUT component, and the sites from which the GSCC is reachable are referred to as the giant IN component. The GSCC is the intersection of the IN and OUT components. Sites in the GWCC, but not in the IN and OUT components, constitute the "tendrils" (see Fig. 1.2.).

Here we repeat the analysis of Section **1.3** for the case of directed networks. We limit ourselves mostly to final results and conclusions. A detailed derivation can be found in [55].

1.4.1 Threshold

The condition for the existence of a giant component in a directed random network of arbitrary degree distribution can be deduced in a manner similar to [14]. If a site is reached following a link pointing to it, then it must have at least one outgoing link, on average, in order to be part

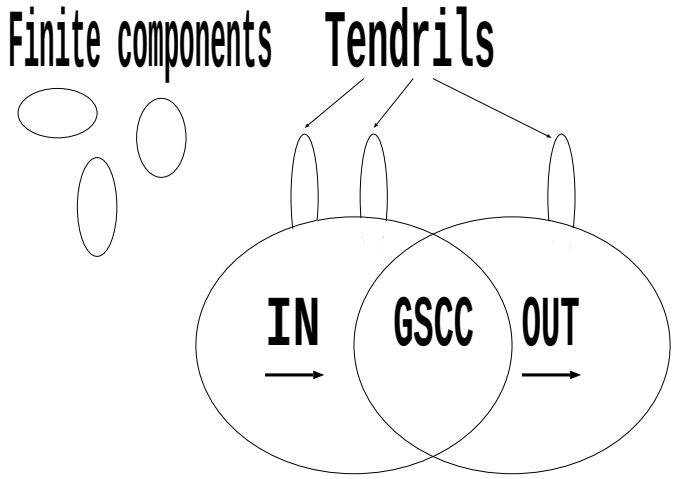

Figure 1.2: Structure of a general directed graph.

of a giant component. This condition can be written as

$$\langle k_j | i \to j \rangle = \sum_{k_i, k_j} k_j P(k_i, k_j | i \to j) = 1. \tag{1.68}$$

Reasoning as in Section **1.3.2**, the above criterion reduces to [35, 53]

$$\langle jk \rangle \geq \langle k \rangle. \tag{1.69}$$

Suppose a fraction $p = 1 - q$ of the nodes is removed from the network, then the original degree distribution, $P(j,k)$, becomes

$$P'(j,k) = \sum_{j_0, k_0}^{\infty} P(j_0, k_0) \binom{j_0}{j} (1-p)^j p^{j_0 - j} \binom{k_0}{k} (1-p)^k p^{k_0 - k} \ . \tag{1.70}$$

In view of this new distribution, Eq. (1.69) yields the percolation threshold

$$q_c = 1 - p_c = \frac{\langle k \rangle}{\langle jk \rangle} \ , \tag{1.71}$$

where averages are computed with respect to the original distribution before dilution, $P(j,k)$. Eq. (1.71) indicates that in directed scale-free networks if $\langle jk \rangle$ diverges then $q_c \to 0$ and the network is resilient to random breakdown of nodes and bonds.

The term $\langle jk \rangle$ may be dramatically influenced by the appearance of correlations between the *in*- and *out*-degrees of the nodes. This effect has been discussed in [53]. Our own studies [55] of uncorrelated and correlated distributions reveal that In the former case the threshold is simply $q_c = 1/\langle k \rangle$, while in the latter case $\langle jk \rangle$ diverges whenever

$$(\lambda_{out} - 2)(\lambda_{in} - 2) \leq 1 \ , \tag{1.72}$$

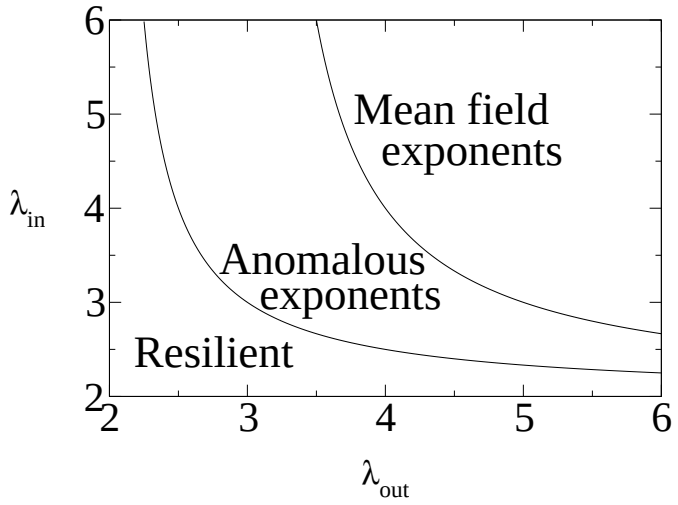

Figure 1.3: Phase diagram of the different regimes for the IN component of scale-free correlated directed networks. The boundary between Resilient and Anomalous exponents is derived from Eq. (1.72) while that between Anomalous exponents and Mean field exponents is given in table 1.1 for $\lambda^\star = 4$. For the diagram of the OUT component λ_{in} and λ_{out} change roles

causing the percolation threshold to vanish. The various regimes resulting from this observation are summarized in Fig. 1.3.

Percolation of the GWCC can be seen to be similar to percolation in the non-directed graph created from the directed graph by ignoring the directionality of the links. The threshold is obtained from the criterion (cf Eq. (1.27))

$$q_{\rm c} = \frac{\langle k \rangle}{\langle k(k-1) \rangle}. \tag{1.73}$$

Here the connectivity distribution is the convolution of the *in* and *out* distributions,

$$P'(k) = \sum_{l=0}^{k} P(l, k-l). \tag{1.74}$$

Whether the distribution is correlated or not, $P'(k)$ is always dominated by the slower decay-exponent, therefore percolation of the GWCC is the same as in non-directed scale-free networks, with $\lambda_{eff} = min(\lambda_{in}, \lambda_{out})$. Note that the percolation threshold of the GWCC may differ from that of the GSCC and the IN and OUT components [53].

1.4.2 Critical exponents

Percolation of the GSCC and IN and OUT components may be analyzed with the formalism of generating functions [44, 39, 35, 53] (see, also, the Chapter by Newman in this book). We have

	uncorrelated	*correlated*
GWCC	$min(\lambda_{out}, \lambda_{in}) + 1$	$min(\lambda_{out}, \lambda_{in})$
IN	$\lambda_{out} + 1$	$\lambda_{out} + \frac{\lambda_{in} - \lambda_{out}}{\lambda_{in} - 1}$
OUT	$\lambda_{in} + 1$	$\lambda_{in} + \frac{\lambda_{out} - \lambda_{in}}{\lambda_{out} - 1}$
GSCC	$min(\lambda_{out}, \lambda_{in}) + 1$	$min(\lambda^*_{out}, \lambda^*_{in})$

Table 1.1: Values of $\lambda^\star$ for the different network components for both correlated and uncorrelated cases.

computed the critical exponents β and τ, following the approach of Section **1.3.5**. The results are the same as for the non-directed case, Eqs. (1.49) and (1.54), but where λ is replaced by an effective λ^* whose value differs for uncorrelated and correlated distributions. The value of λ^* for the various components in both the uncorrelated and correlated scenario are summarized in table 1.1. Our findings [55] indicate that even the tiniest amount of correlation results in behavior typical to the correlated case. We may therefore conclude that in practical situations only the correlated case counts, for it is expected, to some extent, in most naturally occurring directed networks.

1.5 Efficient Immunization Strategies

It is well established that random immunization fails to prevent epidemics of diseases that spread upon contact between infected individuals. On the other hand, targeted immunization requires global knowledge of the topology of the social network in question, rendering it impractical. We propose an effective strategy, based on the immunization of a small fraction of *random acquaintances* of randomly selected individuals, that prevents epidemics without requiring global knowledge of the network [56].

Social networks are known to possess a broad distribution of the number of links (contacts) k, emanating from a node (an individual) [57, 58, 59, 10]. [3] Studies of percolation on broad distribution networks show that a large fraction f_c of the nodes need to be removed (immunized) before the integrity of the network is compromised. This is particularly true for scale-free networks with $2 < \lambda < 3$ — the case of most known networks [6, 9, 13] — where the percolation threshold $f_c \to 1$, and the network remains connected (contagious) even after immunization of most of its nodes [60, 61, 62, 63, 45, 14, 39]. In other words, with a random immunization strategy most of the population needs to be immunized before an epidemic is arrested (see Fig. 1.4).

When the most highly connected nodes are targeted first, removal of just a small fraction of the nodes results in the network's disintegration [39, 41]. This has led to the suggestion of targeted immunization of the HUBs (the most highly connected nodes in the network) [64, 65]. The main shortcoming of this approach is that it requires a complete, or at least fairly good knowledge of the connectivity of each node in the network. Such global information often proves hard to gather, and may not even be well-defined (as in social networks, where the number of social relations depends on subjective judging). Here we propose an effective

[3] Often this is the scale-free distribution $P(k) = ck^{-\lambda}$. Our results apply, however, to broad distributions in general.

immunization strategy that works at low immunization rates f, and obviates the need for global information.

1.5.1 Acquaintance immunization

In our approach, we choose a random fraction p of the population (of size N) and ask each individual to point at an acquaintance with whom they are in contact. The acquaintances, rather than the individuals themselves, are the ones immunized. The fraction f_c needed to be immunized in order to stop the epidemic can be computed analytically.

In each immunization event the probability that a node with k contacts is selected is $kP(k)/(N\langle k\rangle)$. Let $n_l(k)$ be the number of individuals in chemical shell l who are susceptible (not immunized). In the next chemical shell, $l+1$, each of those sites connects to $k-1$ neighbors (excluding the one connecting to shell $l-1$). To find out $n_{l+1}(k')$, we multiply the number of links going out of the lth layer by the probability of reaching a site of connectivity k' by following a link from a *susceptible* site, $p(k'|k \wedge s_k)$, and the probability that this site is also susceptible, $p(s_{k'}|k' \wedge k \wedge s_k)$. This gives

$$n_{l+1}(k') = \sum_k n_l(k)(k-1)p(k'|k \wedge s_k)p(s_{k'}|k' \wedge k \wedge s_k) \ . \tag{1.75}$$

From Bayes' rule,

$$p(k'|k \wedge s_k) = \frac{p(s_k|k \wedge k')p(k'|k)}{p(s_k|k)} \ . \tag{1.76}$$

$p(k'|k) = k'P(k')/\langle k\rangle$ is independent of k. $p(s_k|k \wedge k') = e^{-p/k'} \times \langle e^{-p/k}\rangle^{k-1}$, where the average is taken with respect to $p(k)$ as defined before. $p(s_k|k) = \langle \exp(-p/k)\rangle^k$, since no knowledge exists on its neighbors. Using all these relations one obtains:

$$p(k'|k \wedge s_k) = \frac{p(k')e^{-p/k'}}{\langle e^{-p/k}\rangle} \ . \tag{1.77}$$

The above results, along with (1.75) yield

$$n_{l+1}(k') = v_p^{k'-2}p(k')e^{-p/k'} \sum_k n_l(k)(k-1)e^{-p/k} \ , \tag{1.78}$$

where $v_p = \langle \exp(-p/k)\rangle$. This leads to the stable distribution of connectivity in a chemical layer: $n_l(k) = a v_p^{k-2}p(k)e^{-p/k}$, for some a. Putting this back into (1.78) results in:

$$n_{l+1}(k') = n_l(k') \sum_k p(k)(k-1)v_p^{k-2}e^{-2p/k} \ . \tag{1.79}$$

Therefore, if the sum in (1.79) is larger than 1 the population is above the percolation threshold and the epidemics would propagate, while it would be arrested if the sum is smaller than 1. Thus,

$$\sum_k p(k)(k-1)v_{p_c}^{k-2}e^{-2p_c/k} = 1 \ , \tag{1.80}$$

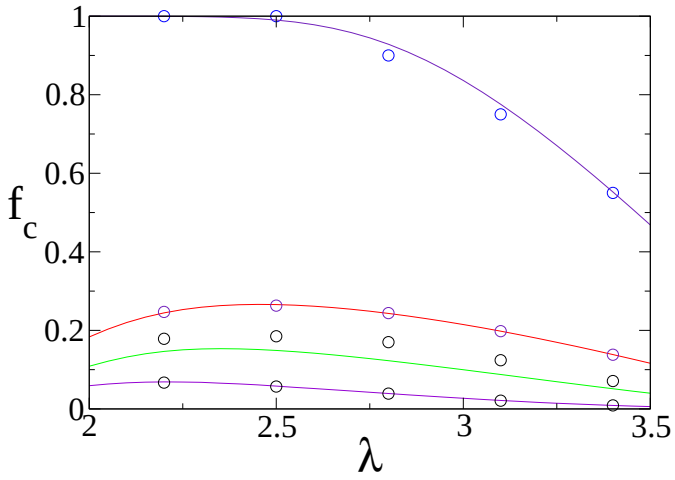

Figure 1.4: Critical probability, f_c, as a function of λ, for the random immunization (top), acquaintance immunization (middle), double acquaintance immunization (lower middle) and attack (bottom) strategies. Curves represent analytical results (approximate for double acquaintace), while data points represent simulation data, for a population $N = 10^6$.

is the condition for criticality. The desired immunization fraction then follows:

$$f_c = \sum_k P(k) v_{p_c}^k \ . \tag{1.81}$$

A related immunization strategy calls for the immunization of acquaintances referred to by at least n individuals. (Above, we specialized to $n = 1$.) The threshold is lower the larger n is, and may justify, under certain circumstances, this somewhat more involved protocol.

In Fig. 1.4, we show the immunization threshold f_c needed to stop an epidemic in networks with $2 < \lambda < 3.5$ (this covers all known cases). Plotted are curves for the (inefficient) random strategy, and the strategy advanced here, for the cases of $n = 1$ and 2. Note the dramatic decrease of f_c with the suggested strategy. Improvements can be achieved for any broad distribution.

Various immunization strategies have been proposed earlier, mainly for the case of an already spread disease and are based on tracing the chain of infection towards the superspreaders of the disease [66]. Our approach can be used even before the epidemic starts spreading, and therefore does not require any knowledge of the chain of infection.

1.6 Summary and Outlook

The main goal of this chapter has been to study the effect of the special nature of scale-free distribution on the properties of random network models. Some general methods have been presented for the study of generalized random networks. Those include methods for the study of the layer structure of the graph, the percolation threshold and the critical exponents.

The special properties of scale-free networks, in conjunction with the general method presented for the study of scale-free and other networks, might prove useful for applications such as the design of more robust networks [40], the improvement of routing [67] and search algorithms [68], and the predicting and arresting of computer and human viruses [45, 65].

Acknowledgments

We would like to thank Keren Erez, Nehemia Schwartz, Alejandro Rozenfeld and Albert-Lazslo Barabási for their collaboration, help and insights on many of the topics covered in this chapter. DbA thanks the support of the National Science Foundation (USA).

References

[1] P. Erdős and A. Rényi, On random graphs. *Publicationes Mathematicae* **6**, 290–297 (1959).

[2] P. Erdős and A. and Rényi, On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 17–61 (1960).

[3] P. Erdős, and A. Rényi, On the strength of connectedness of a random graph. *Acta Mathematica Scientia Hungary* **12**, 261–267 (1961).

[4] B. Bollobás, *Random Graphs*. Academic Press, New York (1985).

[5] A.-L. Barabási and R. Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).

[6] M. Faloutsos, P. Faloutsos and C. Faloutsos, On power-law relationships of the internet topology. *Computer Communications Review* **29**, 251–262 (1999).

[7] A.-L. Barabási, R. Albert and H. Jeong, Scale-free characteristics of random networks: the topology of the World-Wide Web *Physica A*, **281**, 69–77 (2000).

[8] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins J. Wiener, Graph structure in the web. *Computer Networks* **33**, 309–320 (2000).

[9] H. Ebel, L.-I. Mielsch and S. Bornholdt, Scale-free topology of e-mail networks. *Preprint* cond-mat/0201476 (2002).

[10] S. Redner, How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B* **4**, 131–134 (1998).

[11] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, The large-scale organization of metabolic networks *Nature*, **407**, 651, (2000).

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

[13] X. Guardiola, R. Guimera, A. Arenas, A. Diaz-Guilera, D. Streib and L. A. N. Amaral, Macro- and micro-structure of trust networks. *Preprint* cond-mat/0206240 (2002).

[14] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, Resilience of the Internet to Random Breakdown. *Phys. Rev. Lett.* **85**, 4626–4628 (2000).

[15] S. N. Dorogovtsev and J. F. F. Mendes, Natural Scale of Scale-free Networks. *Phys. Rev. E* **63**, 62101 (2001).

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

[17] P. L. Krapivsky, S. Redner and F. Leyvraz, Connectivity of Growing Random Networks. *Phys. Rev. Lett.* **85**, 4629–4632 (2000).

[18] M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms* **6**, 161–179 (1995).

[19] M. Molloy and B. Reed, The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.* **7**, 295–305 (1998).

[20] W. Aiello, F. Chung and L. Lu, A random graph model for massive graphs. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pp. 171–180. Association of Computing Machinery, New York (2000).

[21] A. Bunde and S. Havlin (eds.), *Fractals in Science*. Springer, New York (1994).

[22] A. Bunde and S. Havlin (eds.), *Fractals and Disordered System*. Springer, New York (1996).

[23] P. Lévy, *Calcul des Probabilities*. Gauthier Villars, Paris (1925).

[24] M. F. Shlesinger and J. Klafter, Accelerated Diffusion in Josephson Junctions and Related Chaotic Systems. *Phys. Rev. Lett.* **54**, 2551 (1985).

[25] J. Klafter, M. F. Shlesinger and G. Zumofen, Beyond Brownian motion. *Phys. Today* **49**, 33 (1996).

[26] G. M. Viswanathan, S. V. Buldyrev, S. Havlin, M. G. E. da Luz, E. P. Raposo and H. E. Stanley, Optimizing the success of random searches. *Nature* **401**, 911 (1999).

[27] J. M. Kleinberg, Navigation in a Small World. *Nature* **406**, 845 (2000).

[28] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).

[29] S. N. Dorogovtsev and J. F. F. Mendes, Evolution of networks. *Adv. in Phys.* **51**, 1079–1187 (2002).

[30] D. ben-Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems*. Cambridge University Press, Cambridge (2000).

[31] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).

[32] D. J. Watts, *Small Worlds*. Princton University Press, Princeton (1999).

[33] A. F. Rozenfeld, R. Cohen, D. ben-Avraham and S. Havlin, Scale-free Networks on Lattices. *Preprint* cond-mat/0205613 (2002).

[34] R. Cohen and S. Havlin, Ultra Small World in Scale-Free Networks. *Preprint* cond-mat/0205476 (2002).

[35] M. E. J. Newman, S. H. Strogatz and D. J. Watts, Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118, (2001).

[36] B. Bollobás and O. Riordan, The diameter of a scale-free random graph process *Preprint* (2001).

[37] G. Szabó, M. Alava and J. Kertész, Shortest paths and load scaling in scale-free trees *Preprint* cond-mat/0203278 (2002).

[38] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*. Taylor and Francis, London, 2nd edition (1992).

[39] D. S. Callaway, M. E. J. Newman, S. H. Strogatz and D. J. Watts, Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.* **85**, 5468–5471 (2000).

[40] R. Albert, H. Jeong and A.-L. Barabási, Attack and error tolerance of complex networks. *Nature* **406**, 378–382 (2000).

[41] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, Breakdown of the Internet under Intentional Attack. *Phys. Rev. Lett.* **86**, 3682–3685 (2001).

[42] R. Cohen, D. ben-Avraham and S. Havlin, Percolation Critical Exponents in Scale-Free Networks. Cond-mat/0202259 (2002); *Phys. Rev. E* (in press, 2002).

[43] G. H. Weiss, *Aspects and Applications of the Random Walk*. North-Holland, Amsterdam, (1994).

[44] H. S. Wilf, *Generatingfunctionology* 2nd ed. Academic Press, London, (1994).

[45] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).

[46] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E* **63**, 066117 (2001).

[47] Y. Moreno, R. Pastor-Satorras and A. Vespignani, Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B* **26**, 521–529 (2002).

[48] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman and S. H. Strogatz, Are randomly grown graphs really random? *Phys. Rev. E*, **64**, 041902 (2001).

[49] S. N. Dorogovtsev and J. F. F. Mendes, Anomalous percolating properties of growing networks *Phys. Rev. E* **64**, 066110 (2001).

[50] Z. Burda, J. D. Correia and A. Krzywicki, Statistical ensemble of scale-free random graphs. *Phys. Rev. E* **64**, 046118 (2001).

[51] F. Chung, and L. Y. Lu, The diameter of sparse random graphs. *Adv. Appl. Math.*, **26**, 257, (2001).

[52] S. H. Strogatz, Exploring complex networks. *Nature* **410**, 268–276 (2001) .

[53] S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, Giant strongly connected component of directed networks. *Phys. Rev. E* **64**, 025101 (2001).

[54] R. Albert, H. Jeong and A.-L. Barabási, Diameter of the world-wide web. *Nature* **401**, 130–131 (1999).

[55] N. Schwartz, R. Cohen, D. ben-Avraham, A.-L. Barabaśi and S. Havlin, Percolation in Directed Scale-Free Networks. *Phys. Rev. E* **66**, in press (2002).

[56] R. Cohen, D. ben-Avraham and S. Havlin, Efficient immunization of populations and computers. *Preprint* cond-mat/0207387 (2002).

[57] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley and Y. Åberg, The web of human sexual contacts. *Nature* **411**, 907–908 (2001).

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

[59] R. V. Sole and J. M. Montoya, Complexity and fragility in ecological networks *Proc. Roy. Soc. Lond. B Bio.* **268**, 2039–2045, (2001).

[60] R. M. Anderson, and R. M. May, *Infectious Diseases of Humans*. Oxford University Press, Oxford (1991).

[61] A. L. Lloyd and R. M. May, How viruses spread among computers and people. *Science* **292**, 1316–1317 (2001).

[62] C. P. Warren, L. M. Sander, I. Sokolov, C. Simon and J. Koopman, Percolation on disordered networks as a model for epidemics. *Math. Biosci.*, in press (2002).

[63] M. E. J. Newman, Exact solutions of epidemic models on networks. Preprint cond-mat/0201433 (2002).

[64] Z. Dezsö and A.-L. Barabási, Halting viruses in scale-free networks, *Preprint* cond-mat/0107420 (2001).

[65] R. Pastor-Satorras and A. Vespignani, Immunization of complex networks. *Phys. Rev. E* **65**, 036104 (2001).

[66] W. H. Wethcote and J. A. Yorke, Gonorrhea transmission dynamics and control. *Lecture notes in Biomathematics* **56**, (Springer-Verlag, 1984).

[67] K.-I. Goh, B. Kahng, D. Kim, Universal Behavior of Load Distribution in Scale-free Networks. *Phys. Rev. Lett.* **87**, 278701 (2001).

[68] L. A. Adamic, R. M. Lukose, A. R. Puniyani, B. A. Huberman, Search in Power-Law Networks. *Phys. Rev. E* **64**, 046135 (2001).