

# Tomography and stability of complex networks

Tomer Kalisky<sup>1</sup>, Reuven Cohen<sup>1,3</sup>, Daniel ben-Avraham<sup>2</sup>, and Shlomo Havlin<sup>1</sup>

<sup>1</sup> Minerva Center and Dept. of Physics, Bar-Ilan University, Ramat-Gan, Israel

<sup>2</sup> Dept. of Physics, Clarkson University, Potsdam, NY, USA

<sup>3</sup> Dept. of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel

**Abstract.** We study the structure of generalized random graphs with a given degree distribution  $P(k)$ , and review studies on their behavior under both random breakdown of nodes and intentional attack on the most highly connected nodes. We focus on scale free networks, where  $P(k) \propto k^{-\lambda}$ , for  $m < k < K$ . We first examine the “Tomography” of these networks, i.e. the structure of layers around an arbitrary network node. It is shown that the distance distribution of all nodes from a specific network node consists of two regimes. The first is characterized by rapid growth, and the second decays exponentially. We also show analytically that the nodes degree distribution at each layer is a power law with an exponential cut-off. We then show that scale free networks with  $\lambda < 3$  are resilient to random breakdown, but sensitive to intentional attack. We also describe the behavior of the network near the phase transition and show that the critical exponents are influenced by the scale free nature of the network. We conclude that the critical exponent for the infinite cluster size behaves as  $\beta = 1/|\lambda - 3|$ , and the exponent for the finite clusters size distribution behaves as  $\tau = \frac{2\lambda-3}{\lambda-2}$ , for  $2 < \lambda < 4$ . For  $\lambda > 4$  the exponents are  $\beta = 1$  and  $\tau = 2.5$  as in normal infinite dimensional percolation. It is also shown that for all  $\lambda > 3$  the exponent for the correlation length is  $\nu = 1$  and formulas for the fractal dimensions are obtained. The size of the largest cluster at the transition point, known to scale as  $N^{2/3}$  in regular random graphs, is shown to scale as  $N^{(\lambda-2)/(\lambda-1)}$  for  $\lambda < 4$  and as  $N^{2/3}$  for  $\lambda > 4$ .

## 1 Introduction

Much attention has been focused recently on the topic of complex network behavior [1–5]. Most of the interest has been on scale-free networks, which are believed to represent many phenomena in nature. Scale-free degree distributions have been observed in the Internet [6], World Wide Web (WWW) [7], metabolic networks [8] and many others. For recent reviews see [9–12]. In this paper we review the topics of structure [13] and percolation of such networks [1–4]. Understanding network structure can help devise better networks and more efficient routing algorithms in communications networks by taking advantage of the network structure. Percolation is especially important in forecasting and preventing network malfunctions in the Internet, as well as other realistic networks, and may also be important in the understanding of the stability of biological and chemical processes [14].

Percolation theory has been studied for some decades by physicists and mathematicians. In general it deals with the dilution of a fraction  $p$  (alternatively, the occupation with a density  $q = 1 - p$ ) of the sites or bonds in a graph [15,16]. It is

known that for many graphs a finite threshold  $p_c$  exists, such that for dilution of  $p < p_c$  a spanning cluster (*i.e.* a cluster of size proportional to that of the entire network) exists. While for  $p > p_c$  the graph is fragmented into small clusters. When a spanning cluster exists, its size relative to the graph is denoted  $P_\infty(p)$ . Near the transition point  $P_\infty \sim (p_c - p)^\beta$ , where  $\beta$  (as well as other critical exponents) is universal – that is, depends only on the dimension and large scale properties of the graph and not on the local structure. At the transition point the clusters are fractals, while above and below that point the clusters are fractals up to length scale  $\xi(p)$  (the correlation length) and have the dimension of the graph above  $\xi$ . Near criticality,  $\xi \sim |p_c - p|^{-\gamma}$ .

The structure of this paper is as follows: In Section 2 we discuss general results applicable to generalized random graphs with an arbitrary degree distribution. In Section 3 we discuss networks having a scale-free degree distribution, which will be the main concern of this paper. In Section 4 we discuss the tomography of scale-free networks, that is, their partition into layers surrounding the maximally connected node at different distances. Section 5 presents the model of random breakdown in scale-free networks and analytical and numerical results for this kind of failure. Section 6 offers a similar approach for an intentional attack on the most highly connected nodes. Section 7 presents an analytical derivation of the critical exponents for the percolation transition on scale-free networks, and finally Section 8 presents conclusions and prospects.

## 2 General Results

### 2.1 Condition for a spanning cluster

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

$$\langle k_i | i \leftrightarrow j \rangle = \sum_{k_i} k_i P(k_i | i \leftrightarrow j) = 2. \quad (1)$$

Using Bayes rule we get

$$P(k_i | i \leftrightarrow j) = P(k_i, i \leftrightarrow j) / P(i \leftrightarrow j) = P(i \leftrightarrow j | k_i) P(k_i) / P(i \leftrightarrow j), \quad (2)$$

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 criteria Eq. (1) reduces to [17,2]:

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = 2, \quad (3)$$

at the critical point. A spanning cluster exists for graphs with  $\kappa > 2$ , while graphs with  $\kappa < 2$  contain only small clusters whose size is not proportional to that of the entire network. This criterion was derived earlier by Molloy and Reed [17] using somewhat different arguments.

The neglecting of loops can be justified below the threshold 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). Calculating the fraction of loops  $P_{loop}$  in the system yields:

$$P_{loop} \propto \sum_i \frac{s_i^2}{N^2} < \sum_i \frac{s_i S}{N^2} = \frac{S}{N}, \quad (4)$$

where the sum is over all clusters in the system and  $s_i$  is the size of the  $i$ th cluster. Therefore, the fraction of loops in the system is less than or proportional to  $S/N$ , where  $S$  is the size of the largest cluster. Below the critical threshold there is no spanning cluster in the system and therefore the fraction of loops is negligible. Hence, until  $\kappa = 2$  loops can be neglected. At the threshold the structure of the spanning cluster is almost a tree. Above the threshold loops can no longer be neglected, but since this only happens when a spanning cluster exists the criterion in Eq. (3) is valid as a criterion for finding the critical point. A derivation of the exact conditions under which Eq. (3) is valid can be found in [17].

## 2.2 Critical Threshold for Percolation

The above reasoning can be applied to the problem of percolation on a generalized random network. If we randomly remove a fraction  $p$  of the sites (or bonds), 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, depending on the number of removed neighbors. The new number of connections will be binomially distributed. If we begin with a distribution of degrees  $P_0(k_0)$ , the new distribution of degrees of the network will be:

$$P(k) = \sum_{k_0=k}^{\infty} P_0(k_0) \binom{k_0}{k} (1-p)^k p^{k_0-k}. \quad (5)$$

Calculating the first moment for this distribution, given  $\langle k_0 \rangle$  and  $\langle k_0^2 \rangle$  for the original distribution leads to:

$$\langle k \rangle = \sum_{k=0}^{\infty} P(k)k = (1-p)\langle k_0 \rangle. \quad (6)$$

In the same manner we can calculate the second moment:

$$\langle k^2 \rangle = \sum_{k=0}^{\infty} P(k)k^2 = (1-p)^2 \langle k_0^2 \rangle + p(1-p)\langle k_0 \rangle. \quad (7)$$

Both those quantities can be substituted into Eq. (3) to find the criterion for criticality. This yields:

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{(1-p)^2 \langle k_0^2 \rangle + p(1-p) \langle k_0 \rangle}{(1-p) \langle k_0 \rangle} = 2. \quad (8)$$

Reorganizing Eq. (8), one gets the critical threshold for percolation [2]:

$$1 - p_c = \frac{1}{\kappa_0 - 1}, \quad (9)$$

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

Eqs. (3) and (9) are valid for a wide range of generalized random graphs and distributions. For example for a Cayley tree – a graph with a fixed degree  $z$  and no loops – the criterion from Eq. (9) can be used. This yields the critical concentration  $q_c = 1 - p_c = 1/(z - 1)$ , which is well known [15]. Another example is a random Erdős-Rényi (ER) graph. In those graphs edges are distributed randomly and the resulting degree distribution is Poissonian [18]. Applying the criterion from Eq. (3) to a Poisson distribution yields:

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\langle k \rangle^2 + \langle k \rangle}{\langle k \rangle} = 2, \quad (10)$$

which reduces to  $\langle k \rangle = 1$  as known for ER graphs [18].

### 2.3 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 developed by Molloy and Reed [19]. They suggested viewing the infinite cluster as being explored and used differential equations for the number of un-exposed links and unvisited sites to find the size of the infinite cluster and the degree distribution of the residual graph (the finite clusters).

An alternative and very powerful derivation was given by Newman, Strogatz and Watts [5]. They have used the generating functions method to study the size of the infinite cluster as well as other quantities (such as the diameter and cluster size distribution). They have also applied this method to other types of graphs (directed and bipartite). Here we closely follow their derivation in order to find the size of the infinite cluster and the critical exponents.

In [5] a generating function is built for the degree distribution:

$$G_0(x) = \sum_{k=0}^{\infty} P(k)x^k. \quad (11)$$

The probability of reaching a site with degree  $k$  by following a specific link is  $kP(k)/\langle k \rangle$  [17,2,5,3], and the corresponding generating function for those probabilities is

$$G_1(x) = \frac{\sum kP(k)x^{k-1}}{\sum kP(k)} = \frac{d}{dx}G_0(x)/\langle k \rangle. \quad (12)$$

Assuming that  $H_1(x)$  is the generating function for the probability of reaching a branch of a given size by following a link, the self-consistent equation for  $H_1(x)$  is:

$$H_1(x) = xG_1(H_1(x)) . \quad (13)$$

Since  $G_0(x)$  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:

$$H_0(x) = xG_0(H_1(x)) . \quad (14)$$

Below the transition,  $H_0(1) = 1$ , since this is the probability to belong to a cluster of any size. However, above the transition this probability is no longer normalized since this does not include the infinite cluster. Then, the relative size of the giant cluster is  $P_\infty = 1 - H_0(1)$ , since  $H_0$  contains only the finite-size clusters. It follows that

$$P_\infty = 1 - \sum_{k=0}^{\infty} P(k)u^k , \quad (15)$$

where  $u \equiv H_1(1)$  is the smallest positive root (which can be found numerically) of

$$\langle k \rangle u = \sum_{k=0}^{\infty} kP(k)u^{k-1} . \quad (16)$$

This equation can be solved numerically and the solution can be substituted into Eq. (15) to calculate the size of the infinite cluster in a graph with a given degree distribution.

### 3 Scale-Free Networks

#### 3.1 Description

Our main concern in this paper will be with the behavior of scale-free networks. Scale-Free networks are networks whose degree distribution (i.e. fraction of sites with  $k$  connections) behaves as:

$$P(k) \propto k^{-\lambda}, \quad m \leq k \leq K, \quad (17)$$

where  $\lambda$  is the exponent,  $m$  is the lower cutoff, and  $K$  is the upper cutoff. There are no sites with degree below  $m$  and above  $K$ . For finite networks the upper cutoff  $K(N)$  arises naturally since the fraction of high-degree sites decays with  $k$ . An estimate of this cutoff can be found by the assumption that the tail of the distribution above  $K$  is of the order of one site [2]:

$$\sum_{k=K}^{\infty} P(k) \sim \int_K^{\infty} P(k)dk = \frac{1}{N}. \quad (18)$$

The estimate obtained this way gives:

$$K \approx mN^{1/(\lambda-1)}. \quad (19)$$

This estimate allows the derivation of finite size effects in the network and allows calculations of moments of the distribution in Eq. (17), that would otherwise diverge. Newman *et al.*[5] use an exponential cutoff rather than a sharp one, but the effect on the results is minor. Another difference between the distribution used by us and by Newman *et al.*[5] is given in Section 3.2.

The importance of scale-free networks lies in the fact that this distribution occurs in many natural and man-made networks [6,5]. An example of a scale-free network is the physical Internet structure, that is the router to router (and end-units) connectivity. This structure was studied by Faloutsos *et al.*[6]. They have found that the inter-router network can be well approximated by a non-directed scale-free network with  $\lambda \approx 2.5$ . The size of the Internet today is about  $10^7$  sites, making it a fairly large network.

A few results about the structure of scale-free networks have also been derived by Aiello *et al.*[20]. The size of the infinite cluster was calculated, and it was found that for  $\lambda \leq 2$  the infinite cluster is of the order of the size of the entire graph (i.e.  $P_\infty = 1 - o(1)$ , where  $o(1)$  is a function of the network size,  $f(N)$ , such that  $f(N) \rightarrow 0$  when  $N \rightarrow \infty$ ). For  $\lambda > \lambda_c = 3.478\dots$  there is no infinite cluster at all (since we use a somewhat different distribution Eqs. (21) and (22), we get  $\lambda_c \approx 4$ ). For  $\lambda < \lambda_c$  the second largest cluster is of order  $\ln N$ . For lower cutoff  $m \geq 2$  a spanning cluster exists for every  $\lambda$ .

The average distance between sites is also different in scale free networks from its value for normal random graphs. While for ER graphs the average distance between sites behaves as  $d \sim \ln N$  [18], for scale free graphs with  $2 < \lambda < 3$  the distance behaves as  $d \sim \ln \ln N$  [21,22], and for  $\lambda = 3$  as  $d \sim \ln N / \ln \ln N$  [23]. The reason for this short distance is the small core, containing most high degree sites, which has a very small diameter. For  $\lambda > 3$  the normal behavior  $d \sim \ln N$  is recovered.

### 3.2 Simulation

To simulate general networks we have used the following algorithm:

1. For each site choose a degree from the required distribution.
2. Create a list where each site is repeated as many times as the chosen degree.
3. Randomly choose pairs from the list and connect the chosen sites (by adding each node to the list of neighbors of the other site). Remove the chosen pair from the list (by replacing them with the last two entries in the list).

Double and self-edges are ignored, and if the number of entries in the list is odd (overall odd number of connections) one entry can also be discarded. This has a minor influence on the degree distribution if the network is large. The justification of using this algorithm for simulation and analysis as an algorithm generating the probability space of generalized random graphs can be found in [17].

To generate degrees from a scale-free distribution we generate a random number,  $u$ , between 0 and 1 from a uniform distribution, and then generate a

new number  $k$  using the formula:

$$k = \frac{m}{u^{1/(\lambda-1)}}. \quad (20)$$

This generates a random *real* number greater than  $m$ , with a distribution of  $P(k) \propto k^{-\lambda}$ . To prevent the appearance of overflows,  $u$  can be chosen such that it is large enough to yield numbers smaller than  $N - 1$  (since this is the maximum degree per site). If an upper cutoff is required a new random number is drawn whenever  $k$  is too large. After  $k$  is chosen, the closest integer is taken as the degree of the site. The resulting degree distribution is thus:

$$P(k') = \int_{k'-1/2}^{k'+1/2} ck^{-\lambda} dk, \quad (21)$$

where  $c$  is the normalization factor, except for the lower cutoff whose probability is given by:

$$P(m) = \int_m^{m+1/2} ck^{-\lambda} dk. \quad (22)$$

This is in contrast to the probability chosen by [20] and [5,3], who use a discrete scale-free distribution. The main difference is in the fraction of sites of degree  $m$  (usually  $m = 1$ ). Due to Eq. (22) the fraction of low degree sites in our derivation is lower and therefore the networks generated this way are more robust than the ones generated using the discrete distribution. However, the behavior of both distributions in the tail is approximately the same, and therefore the qualitative behavior of all phenomena influenced by the scale-free nature of the distribution should be the same.

The complexity of the algorithm depends upon the number of links in the network and therefore is of order  $O(N\langle k \rangle)$ , which is fast enough to execute for large graphs. The space needed is again of order  $O(N\langle k \rangle)$ . Those limits allow the creation of networks of about  $10^7$  sites.

An alternative method for the study of such networks is a variation of the Leath algorithm. In this method a table is created with number of sites of each degree. The Molloy-Reed construction [17] is then used to build the clusters.

### 3.3 Maximum degree

In Eq. (19) we suggest that the upper cutoff of a scale free network scales as  $K = N^{1/(\lambda-1)}$ . However, for the spatially embedded graphs [24], we find that no graph with  $\lambda < 3$  can be embedded in a lattice without sacrificing the natural cutoff. That is, the cutoff is limited to  $k \approx \sqrt{N}$ . This holds true for every  $d$ . Thus, we expect this to hold true even in the  $d \rightarrow \infty$  case. Similar results are indeed obtained for mean field (*i.e.* non-embedded) graphs [25]. On the other hand, Warren *et al* [26] find the natural cutoff even for graphs embedded in  $d = 2$  lattices.

These differences may be explained by the fact that the cutoff depends on the ensemble from which the graphs are chosen. If the ensemble is defined as all

graphs with the exact given scale free degree distribution and no self loops (loops connecting a site to itself) and no double edges (two edges or more connecting the same set of sites) then the upper cutoff can not be larger than  $\sqrt{N}$ , while if multigraphs are allowed (i.e. including self loops and double edges) the natural upper cutoff is achieved. The following two facts support this intuition:

**Almost all graphs with the natural cutoff are multigraphs:** Consider a graph with a site having degree  $K \gg \sqrt{N}$ . Since for  $\lambda > 2$  the total number of links is of order  $N$  the number of self loops of the HUB is proportional to  $K^2/cN \gg 1$ . Similar results apply for double edges between two such sites.

**Removing the double edges and self loops does not affect the behavior of the tail of the distribution:** Suppose a site has degree  $K \gg \sqrt{N}$ . The number of degree 1 sites is of order  $N$ . Therefore, the number of connections from this site to sites of degree 1 is proportional to its degree  $K$ . Since edges leading to degree 1 sites are neither self loops nor double edges, the upper cutoff is at least proportional to the natural cutoff. In reality, the deviation from the exact degree distribution is quite small.

From the above one can conclude that starting from a scale free degree distribution a multigraph can be constructed, with the given degree sequence. Converting it into a simple graph will change the degree distribution, while keeping the power law form of the tail. However, if one is forced to keep the exact degree sequence while prohibiting self-loops and double edges, the upper cutoff may change.

## 4 Tomography of Scale Free Networks

In this Section we study the network structure by describing the statistical behavior of layers surrounding the maximal connected node. First, we describe the process of generating the network, and define our terminology. Then, we analyze the degree distribution at each layer surrounding the maximally connected node. The results presented here are based on [13].

### 4.1 Description

We base our construction on the Molloy-Reed model [19], also described in Section 3. The construction process tries to gradually expose the network, following the method introduced in [21,27], and is forcing a hierarchy on the Molloy-Reed model, thus enabling us to define layers in the graph.

We start by setting the number of nodes in the network,  $N$ . We then choose the nodes degrees according to the scale-free distribution function  $P(k) = ck^{-\lambda}$ , where  $c \approx (\lambda - 1)m^{\lambda-1}$  is the normalizing constant and  $k$  is in the range  $[m, K]$ , for some chosen minimal degree  $m$  and the natural cutoff  $K = mN^{1/(\lambda-1)}$  of the distribution [2,28].

At this stage each node in the network has a given number of outgoing links, which we term *open connections*, according to its chosen degree. Let us define  $V$  as the set of  $N$  chosen nodes,  $C$  as the set of unconnected outgoing links from



the nodes in  $V$ , and  $E$  as the set of edges in the graph. Using these definitions, the set of links in  $E$  is empty at this point, while the set of outgoing open links in  $C$  contains all unconnected outgoing links in the graph. In the Molloy-Reed construction described in section 3, the links in  $C$  are randomly matched, such that at the end of the process,  $C$  is empty, and  $E$  contains all the matched links  $\langle u, v \rangle$ ,  $u, v \in V$ .

Instead, here we proceed as follows: we start from the maximal degree node, which has a degree  $K$ , and connect it randomly to  $K$  available open connections, thus removing these open connections from  $C$  (see Figure 1(a)). We have now exposed the first *layer* (or *shell*) of nodes, indexed as  $l = 1$ . We now continue to fill out the second layer  $l = 2$  in the same way: We connect all open connections emerging from nodes in layer 1 to randomly chosen open connections. These open connections may be chosen from nodes of layer 1 (thus creating a loop) or from other links in  $C$ . We continue until all open connections emerging from layer 1 have been connected, thus filling layer  $l = 2$  (see Figure 1(b)). Generally, to form layer  $l + 1$  from an arbitrary layer  $l$ , we randomly connect all open connections emerging from  $l$  to either other open connections emerging from  $l$  or chosen from the other links in  $C$  (see Figure 1(c)). Note, that when we have formed layer  $l + 1$ , layer  $l$  has no more open connections. The process continues until the set of open connections,  $C$ , is empty.

## 4.2 Theory

We proceed now to evaluate the probability for nodes with degree  $k$  to reside outside the first  $l$  layers, denoted by  $P_l(k)$ .

The number of open connections outside layer  $l$ , is given by:

$$T_l = N \sum_k k P_l(k) \quad (23)$$

Thus, we can define the probability that a detached node with degree  $k$  will be connected to an open connection emerging from layer  $l$  by  $\frac{k}{\chi_l + T_l}$ , where  $\chi_l$  is the number of open connections emerging from layer  $l$  (see Figure 1(b)).

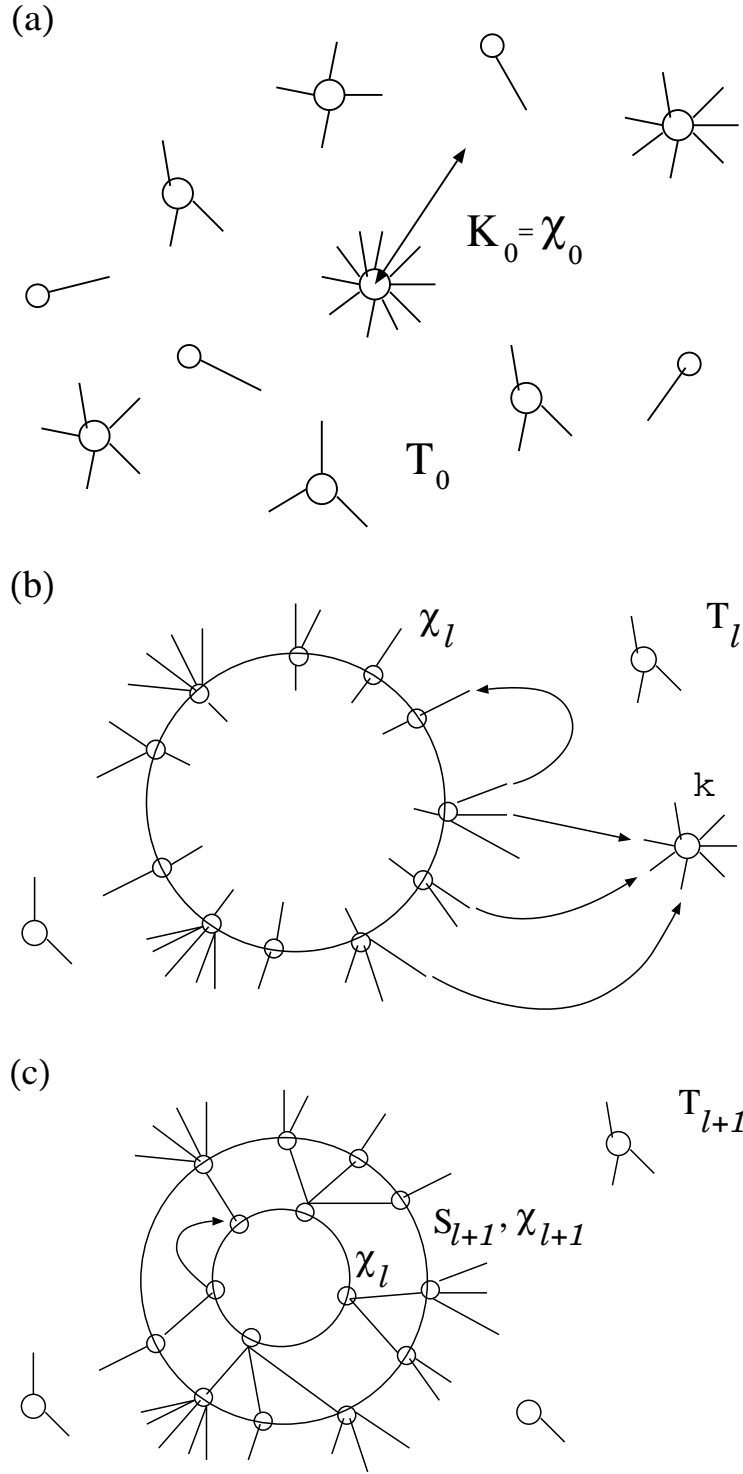
Therefore, the conditional probability for a node with degree  $k$  to be also outside layer  $l + 1$ , given that it is outside layer  $l$ , is the probability that it does not connect to *any* of the  $\chi_l$  open connection emerging from layer  $l$ , that is:

$$P(k, l + 1 | l) = \left[ 1 - \frac{k}{\chi_l + T_l} \right]^{\chi_l} \approx \exp \left( -\frac{k}{1 + \frac{T_l}{\chi_l}} \right), \quad (24)$$

for large enough values of  $\chi_l$ .

Thus, the probability that a node of degree  $k$  will be outside layer  $l + 1$  is:

$$P_{l+1}(k) = P_l(k) P(k, l + 1 | l) = P_l(k) \exp \left( -\frac{k}{1 + \frac{T_l}{\chi_l}} \right) \quad (25)$$



**Fig. 1.** Illustration of the exposure process. The large circles denote exposed layers of the giant component, while the small circles denote individual sites. The sites outside the circles have not been reached yet. (a) We begin with the highest degree node and fill out layer 1. (b) In the exposure of layer  $l + 1$  any open connection emerging from layer  $l$  may connect to any open node ( $T_l$  connections) or loop back into layer  $l$  ( $\chi_l$  connections). (c) The number of connections emerging from layer  $l + 1$  is the difference between  $T_l$  and  $T_{l+1}$  after reducing the incoming connections  $S_{l+1}$  from layer  $l$ .

Thus we derive the exponential cutoff:

$$P_l(k) = P(k) \exp\left(-\frac{k}{K_l}\right), \quad (26)$$

where :

$$\frac{1}{K_{l+1}} = \frac{1}{K_l} + \frac{1}{1 + \frac{T_l}{\chi_l}}, \quad (27)$$

gives the evolution of the cutoff with  $l$ .<sup>1</sup>

Now let us find the behavior of  $\chi_l$  and  $S_l$ , where  $S_l$  is the number of links incoming to the  $l + 1$  layer (and approximately<sup>2</sup> equals  $N_{l+1}$ , the number of nodes in the  $l + 1$  layer). The number of incoming connections to layer  $l + 1$  equals the number of connections emerging from layer  $l$ , minus the number of connections looping back into layer  $l$ . The probability for a link to loop back into layer  $l$  is:

$$P(\text{loop}|l) = \frac{\chi_l}{\chi_l + T_l} \quad (32)$$

and therefore:

<sup>1</sup> The exponential cutoff may be derived also using the following “mean field” approximation: Each node is treated independently, where the *interaction* between nodes is inserted through the expected number of incoming connections. At each node, the process is treated as equivalent to randomly distributing  $\chi_l$  independent points on a line of length  $\chi_l + T_l$  and counting the resultant number of points inside a *small* interval of length  $k$ . Thus, the number of incoming connections  $k_{in}$  from layer  $l$  to a node with  $k$  open connections is distributed according to a Poisson distribution with:

$$\langle k_{in} \rangle = \frac{k}{\chi_l + T_l} \chi_l, \quad (28)$$

and :

$$P_{l+1}(k_{in}|k) = e^{-\langle k_{in} \rangle} \frac{\langle k_{in} \rangle^{k_{in}}}{k_{in}!}. \quad (29)$$

The probability for a node with  $k$  open connections *not* to be connected to layer  $l$ , i.e. to be outside layer  $l + 1$  also, is:

$$P(k, l + 1|l) = P_{l+1}(k_{in} = 0|k) = e^{-\langle k_{in} \rangle} = \exp\left(-\frac{k}{1 + \frac{T_l}{\chi_l}}\right) \quad (30)$$

Thus the total probability to find a node of degree  $k$  outside layer  $l + 1$  is:

$$P_{l+1}(k) = P_l(k) P(k, l + 1|l) = P_l(k) \exp\left(-\frac{k}{1 + \frac{T_l}{\chi_l}}\right) \quad (31)$$

And one obtains the exponential cutoff.

<sup>2</sup> This holds true assuming that almost no site in layer  $l + 1$  is reached by two connections from layer  $l$ . This is justified in the case where  $m = 1$ , and also for the first layers in case of  $m > 1$ .

$$S_{l+1} = \chi_l \left( 1 - \frac{\chi_l}{\chi_l + T_l} \right). \quad (33)$$

The number of connections emerging from all the nodes in layer  $l + 1$  is  $T_l - T_{l+1}$ . This is the sum of the number of incoming connections from layer  $l$  into layer  $l + 1$ , which is equal to  $S_{l+1}$ , and the number of outgoing connections  $\chi_{l+1}$ . Therefore:

$$\chi_{l+1} = T_l - T_{l+1} - S_{l+1} \quad (34)$$

At this point we have the following relations:  $T_{l+1}(K_{l+1})$  Eq. (23) and Eq. (26),  $S_{l+1}(\chi_l, T_l)$  Eq. (33),  $K_{l+1}(K_l, \chi_l, T_l)$  Eq. (27), and  $\chi_{l+1}(T_l, T_{l+1}, S_{l+1})$  Eq. (34). These relations may be solved numerically. Note that approximate analytical results for the limit  $N \rightarrow \infty$  can be found in [21,27,22].<sup>3</sup>

### 4.3 Simulation

Figure 2 shows results from simulations (symbols) for the number of nodes on layer  $l$ , which can be seen to be in agreement with the analytical curves of  $S_l$  (lines). We can see that starting from a given layer  $l = L$  the number of nodes decays exponentially. We believe that the layer index  $L$  is related to the radius of the graph [21,27]. It can be seen that  $S_l$  is a good approximation for the number of nodes at layer  $l$ . This is true in cases when only a small fraction of sites in each layer  $l$  have more than one incoming connection. An example for this case is when  $m = 1$  so that most of the sites in the network have only one connection. Figure 3 shows results for  $P_l(k)$  with similar agreement. Note the exponential cutoff which becomes stronger with  $l$  (i.e.  $K_l$  is a monotonically decreasing function of  $l$ ).

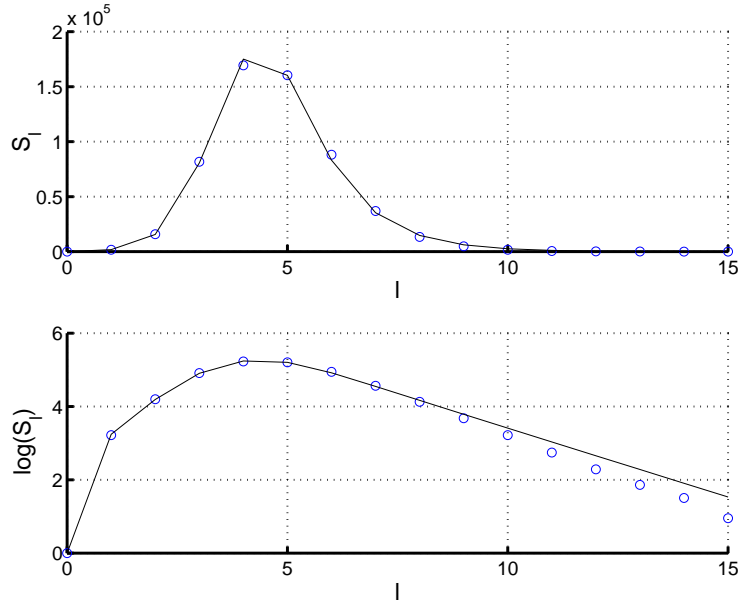
It is important to note that the simulation results give the probability distribution for the giant percolation cluster, while the analytical reconstruction gives the probability distribution for the whole graph. This may explain the difference in the probability distributions for lower degrees: many low degree nodes are not connected to the giant percolation cluster and therefore the probability distribution derived from the simulation is smaller for low  $k$ .

Similar results were found in real Internet maps and multicast trees [13]. Deviations from theory may be attributed to correlations in node degrees [29] and hierarchical structures [30], which were observed in the Internet.

<sup>3</sup> An approximate analytical expression for the upper cutoff was found to be [21]:

$$K_l \sim A \frac{(\lambda-2)^{l-1} - 1}{3-\lambda} N^{\frac{(\lambda-2)^l}{\lambda-1}}. \quad (35)$$

where  $A = \langle k \rangle m^{\lambda-2} / (3 - \lambda) = \frac{(\lambda-1)m}{(\lambda-2)(3-\lambda)}$ .

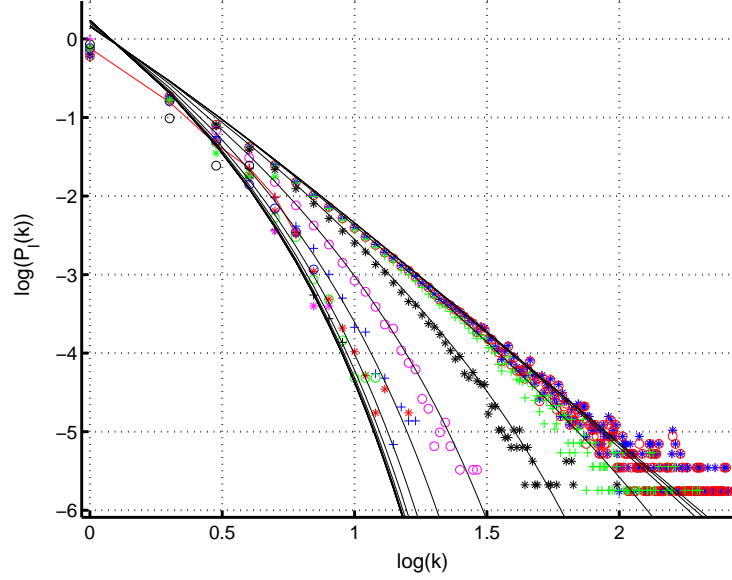


**Fig. 2.** Approximate number of nodes ( $S_l$ ) vs. layer index  $l$  for a network with  $N = 10^6$  nodes,  $\lambda = 2.85$ , and  $m = 1$ . Symbols represent simulation results while solid lines are a numerical solution for the derived recursive relations. Bottom: from the semi-log plot we see that there is an exponential decay of  $S_l$  for layers  $l > L$  starting from a given layer  $L$  which we believe is related to the radius of the graph.

#### 4.4 Bounds and Implications

The layer structure of the network has implications on several important topics. Since messages in a communication network travel between neighboring nodes the inter-node distances are important in understanding network performance and message routing. Another important subject is searching for nodes in a network. In [31], an efficient method for searching via the network HUBs is presented. This method is based on going up the degree sequence, from each site to a higher degree site, until the highest degree site is reached. Then, the search continues down the degree sequence to lower and lower degree sites. This method allows for a search much more efficient than a random one, as shown in [31]. However, as we now show, no search strategy based on local information can search a finite fraction of the network in less than  $O(N)$  steps (with possible logarithmic corrections).

Some limits on the efficiency of such techniques can be obtained by using bounds on the structure of scale free networks as in [21]. These bounds follow from the scale free degree sequence of the network, and are independent of the exact model for network generation. Therefore, they apply to **every** network with a scale-free degree sequence.



**Fig. 3.** Log-log plot of  $P_l(k)$  for different layers  $l = 0, 1, 2, \dots$  (from top to bottom), for a network with  $N = 10^6$  nodes,  $\lambda = 2.85$ , and  $m = 1$ . Symbols represent simulation results while solid lines are a numerical solution for the derived recursive relations.

If we consider the fraction,  $A$ , of the most most highly connected sites, their fraction of neighbors (relative to the network size),  $n_1(A)$ , can not exceed  $\int_f^K ck^{1-\lambda} \approx f^{2-\lambda} m^{\lambda-1} (\lambda-1)/(\lambda-2)$ , where  $f$  satisfies  $N \int_f^K ck^{-\lambda} = A$ . Hence,  $f = mA^{1/(\lambda-1)}$ , and  $n_1(A) \leq mA^{(\lambda-2)/(\lambda-1)} (\lambda-1)/(\lambda-2)$ . Similarly, the number of second neighbors of these sites can not exceed  $n_2(A) \leq n_1(n_1(A)) \leq (m(\lambda-1)/(\lambda-2))^{(2\lambda-3)/(\lambda-1)} A^{(\lambda-2)^2/(\lambda-1)^2}$ . One can continue for the third nearest neighbors and so on.

Since the number of  $l$ th nearest neighbors of a group of  $AN$  sites behaves as  $NA^{\frac{\lambda-2}{\lambda-1}}$  it follows that for  $A \sim N^{-\epsilon}$  for any  $\epsilon$  no finite number of layers can contain  $O(N)$  sites in the limit  $N \rightarrow \infty$ . One can therefore conclude that the average distance between sites in a scale-free networks **can not be a constant**, and also that **no searching of  $O(N)$  sites using less than  $O(N)$  steps is possible**. The second conclusion is true, of course, only if no information other than  $l$ th nearest neighbors is allowed, and does not apply to methods such as the one suggested in [32], which can search in less than  $O(N)$  step owing to the knowledge of some distance metric.

## 5 Random Breakdown

### 5.1 Description

Albert *et al.*[1] suggested to model the Internet as a scale-free network (in their original model  $\lambda \approx 3$ ). They suggested a scenario in which nodes in the network fail randomly (due to random error or an external cause like power-failures etc.). To model this scenario, they suggested random removal of sites from the network, after which they calculated the size of the largest remaining cluster. They have compared the results of applying this process to a scale-free network to the same model on a random Erdős-Rényi (ER) graph. They found numerically that scale-free networks with  $\lambda = 3$  are much more resilient to this kind of failure than ER graphs. In a previous paper [2] we have studied this problem analytically and numerically, and the results are summarized below.

### 5.2 Theory

In the calculation of the threshold for random breakdown, the key parameter, according to (9), is the ratio second- to first-moment,  $\kappa_0$ , which we compute by approximating the distribution (17) to a continuum. (This approximation becomes exact for  $1 \ll m \ll K$ , and it preserves the essential features of the transition even for small  $m$ . Furthermore, for our continuous distribution this approximation is fairly accurate for all values of the cutoff):

$$\kappa_0 = \left( \frac{2 - \lambda}{3 - \lambda} \right) \frac{K^{3-\lambda} - m^{3-\lambda}}{K^{2-\lambda} - m^{2-\lambda}}. \quad (36)$$

When  $K \gg m$ , this may be approximated as:

$$\kappa_0 \rightarrow \left| \frac{2 - \lambda}{3 - \lambda} \right| \times \begin{cases} m, & \text{if } \lambda > 3; \\ m^{\lambda-2} K^{3-\lambda}, & \text{if } 2 < \lambda < 3; \\ K, & \text{if } 1 < \lambda < 2. \end{cases} \quad (37)$$

We see that for  $\lambda > 3$  the ratio  $\kappa_0$  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  $\kappa_0$  diverges with  $K$  and so  $p_c \rightarrow 1$  when  $K \rightarrow \infty$  (or  $N \rightarrow \infty$ ). The percolation transition does not take place: a spanning cluster exists for arbitrarily large fractions of breakdown,  $p < 1$ . 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  $\kappa$  shows that it is lower than 2 even before the breakdown occurs. For  $\lambda > 4$  and  $m = 1$  the network will consist of only finite clusters and no spanning cluster exists (This is reminiscent of the result for  $\lambda > 3.478\dots$  found in [20], where the different threshold stems from the different distribution taken). For  $m \geq 2$ , a spanning cluster exists for every  $\lambda$ .

The size of the spanning cluster can also be measured using the methods suggested in [5]. The distribution (5) can be substituted into Eq. (11) and the calculation of the other generating functions can be done using this distribution, giving the size of the spanning cluster relative to the undisturbed network. An alternative method [3] is to build a new generating function, taking the fraction of removed sites into account:

$$G_0(x) = \sum_{k=0}^{\infty} P(k)q(k)x^k, \quad (38)$$

where  $q(k) = 1 - p(k)$  is the probability that a site of degree  $k$  is not removed. This equation replaces Eq. (11). For random breakdown  $q(k) = 1 - p$  is independent of  $k$ . The size of the infinite cluster is then given by:

$$P_{\infty} = G_0(1) - G_0(u), \quad (39)$$

where  $u$  is the smallest positive solution of:

$$u = 1 - G_1(1) + G_1(u). \quad (40)$$

### 5.3 Simulation

To simulate the process of random breakdown of a fraction  $p$  of the sites,  $pN$  sites are chosen at random. Those sites are removed and all bonds connected to them are also removed. The links are followed to remove all connections to those sites from the other end of the link.

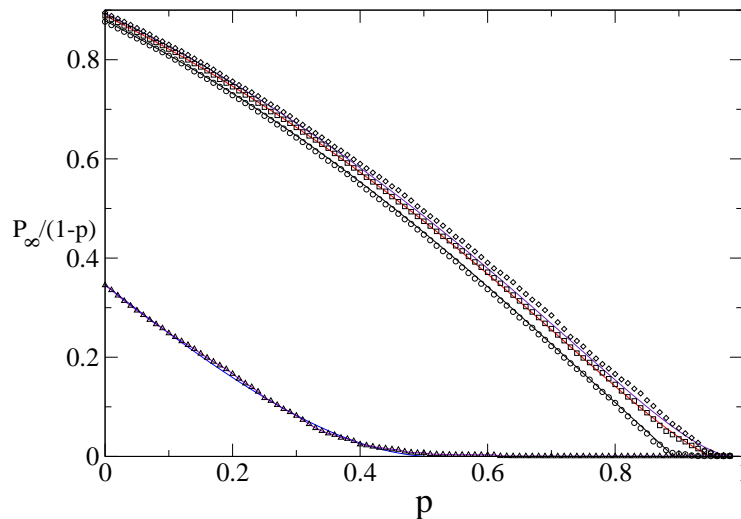
After the removal of the sites, the size of the spanning cluster (if exists) and the finite clusters is measured. The measurement is performed using breadth first search (BFS) from each site which hasn't been marked as probed. This method uncovers the graph cluster by cluster, and can also be used to find the distance (i.e. the shortest path) between a site on the cluster and all the other sites on that cluster. Numerical results for the largest cluster in random breakdown can be seen in Fig. 4. Numerical and analytical results of  $p_c$ , for different values of  $\lambda$  and  $m$ , are shown in figure 5.

## 6 Intentional Attack

### 6.1 Description

Another model suggested in [1] is that of intentional attack on the most highly connected nodes in the network. In this model an attacker (e.g. computer hackers trying to cause damage to the network) leases by some means an intentional attack on the most highly connected sites in the network, causing breakdown of those sites. An attack such as this will expectedly cause more extensive damage than a random attack, and as will be shown below, can cause even networks resilient to random breakdown to collapse.





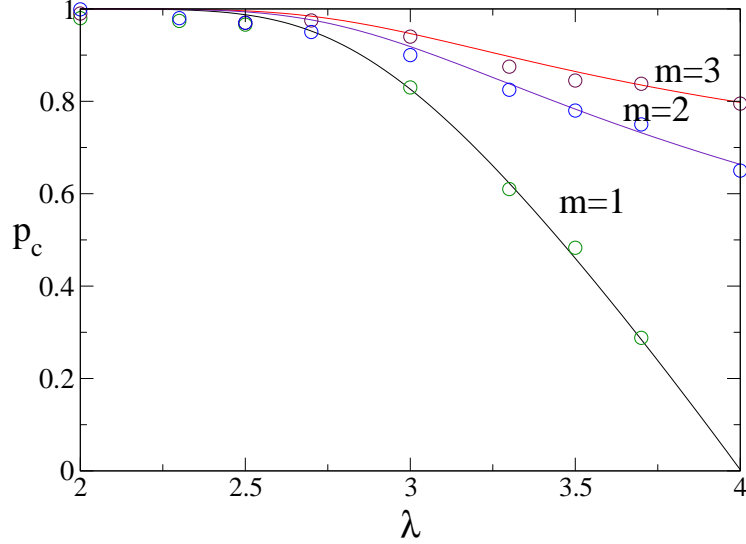
**Fig. 4.** Percolation transition for networks with power-law degree distribution. Plotted is the fraction of nodes that remain in the spanning cluster after breakdown of a fraction  $p$  of all nodes,  $P_\infty(p)/P_\infty(0)$ , as a function of  $p$ , for  $\lambda = 3.5$  (crosses) and  $\lambda = 2.5$  (other symbols), as obtained from computer simulations of up to  $N = 10^6$ . In the former case, it can be seen that for  $p > p_c \approx 0.5$  the spanning cluster disintegrates and the network becomes fragmented. However, for  $\lambda = 2.5$  (the case of the Internet), the spanning cluster persists up to nearly 100% breakdown. The different curves for  $K = 25$  (circles), 100 (squares), and 400 (triangles) illustrate the finite size-effect: the transition exists only for finite networks, while the critical threshold  $p_c$  approaches 100% as the networks grow in size.

## 6.2 Theory

Consider now intentional attack, or sabotage [1], whereby a fraction  $p$  of the sites with the highest degree is removed. (The links emanating from the sites are removed as well.) This has the following effect: (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 the original distribution, but is changed, because of the removal of many of their links. The upper cutoff  $K$  before the attack may be estimated from Eq. (19). Similarly, the new cutoff  $\tilde{K}$ , after the attack, can be estimated from

$$\sum_{k=\tilde{K}}^K P(k) = \sum_{k=\tilde{K}}^{\infty} P(k) - \frac{1}{N} = p. \quad (41)$$

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:



**Fig. 5.** The Percolation threshold  $p_c$  for different values of  $\lambda$  and  $m$ , for random breakdown of nodes. Circles represent simulation results for  $N = 5 \cdot 10^5$  nodes, while solid lines give analytical results. The deviation from  $p_c = 1$  as  $\lambda \rightarrow 3$  is due to finite size effects.

$$\tilde{K} = mp^{1/(1-\lambda)}. \quad (42)$$

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  $\tilde{p}$  of a link leading to a deleted site equals the ratio of the number of links belonging to deleted sites to the total number of links:

$$\tilde{p} = \sum_{k=\tilde{K}}^K \frac{kP(k)}{\langle k_0 \rangle}, \quad (43)$$

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

$$\tilde{p} = \left( \frac{\tilde{K}}{m} \right)^{2-\lambda} = p^{(2-\lambda)/(1-\lambda)}, \quad (44)$$

for  $\lambda > 2$ . For  $\lambda = 2$ ,  $\tilde{p} \rightarrow 1$ , 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. (43) by an integral yields  $\tilde{p} = \ln(Np/m)$ . That is, for

$\lambda = 2$ , very small values of  $p$  are needed to destroy an arbitrarily large fraction of the links as  $N \rightarrow \infty$ .

With the above results we can compute the effect of intentional attack, using the theory previously developed for random removal of sites [2]. 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. This can be seen as the result of two processes: (a) Removal of the highest degree sites reduces the upper cutoff. Since this effect changes the degree distribution,  $\kappa_0$  needs to be recalculated accordingly. (b) Removal of the links leading to the removed sites. The probability of removing a link is  $\tilde{p}$  — the probability of a randomly chosen link to lead to one of the removed sites — and all links have the same probability of being deleted. Since this effect has the influence on the probability distribution described in Eq. (5), the result in Eq. (9) can be used, with  $\tilde{p}$  replacing  $p$ . (Notice that for random site deletion the probability of a link leading to a deleted site is identical to the fraction of deleted sites.)

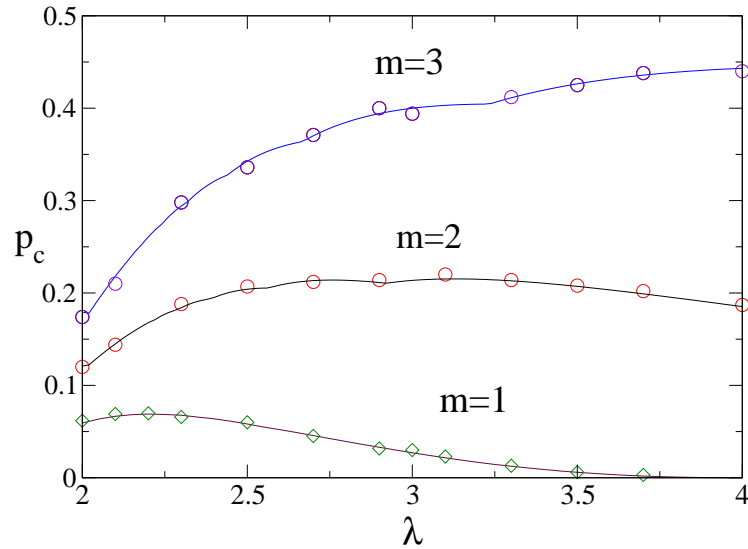
Although the number of nodes removed in intentional attack is different than in the random breakdown model, this affects the size of the spanning cluster (see below) but not the critical point. This is because the transition point is defined as the point where the spanning cluster becomes a finite fraction of the whole network. A finite fraction of the remaining nodes is also a finite fraction of the original network, so the difference has no effect on  $p_c$ .

We therefore use Eqs. (9) and (36), but with  $\tilde{p} = (\tilde{K}/m)^{2-\lambda}$  and  $\tilde{K}$  replacing  $p_c$  and  $K$ . This yields the equation:

$$(\tilde{K}/m)^{2-\lambda} - 2 = \frac{2-\lambda}{3-\lambda} m [(\tilde{K}/m)^{3-\lambda} - 1], \quad (45)$$

which can be solved numerically to obtain  $\tilde{K}(m, \lambda)$ , and then  $p_c(m, \lambda)$  can be retrieved from Eq. (42). In Fig. 6 we plot  $p_c$  — the critical fraction of sites needed to be removed in the targeted attack strategy to disrupt the network — computed in this fashion, and compared to results from numerical simulations. A phase transition exists (at a finite and small  $p_c$ ) for all  $\lambda > 2$ . The decline in  $p_c$  for large  $\lambda$  is explained from the fact that as  $\lambda$  increases the spanning cluster becomes smaller in size, even before attack. (Furthermore, for  $m < 2$  the original network is disconnected for some large enough  $\lambda$ .) The decline in  $p_c$  as  $\lambda \rightarrow 2$  results from the critically high degree of just a few sites: their removal disrupts the whole network. This was already argued in [1]. We note that for infinite systems  $p_c \rightarrow 0$  as  $\lambda \rightarrow 2$ . The critical fraction  $p_c$  is rather sensitive to the lower degree cutoff  $m$ . As shown in figure 6, for larger  $m$  the networks are more robust, though they still undergo a transition at a finite  $p_c$ .

To calculate  $P_\infty(p)$  one can use the generating function method in conjunction with  $p(k) = \Theta(\tilde{K} - k)$  [3] (to be more exact,  $p(k) = \Theta(\tilde{K} - k) + A\delta_{k, \tilde{K}}$ , where  $A$  is chosen such that  $\sum_{k=0}^{\infty} p(k) = 1 - p$ ). An alternative method is to use the generating functions in conjunction with the new distribution after the attack.



**Fig. 6.** Critical probability  $p_c$  as a function of  $\lambda$ , for networks of size  $N = 5 \cdot 10^5$  under intentional attack (circles). Solid lines represent the analytical solution, obtained from Eqs. (41-45).

### 6.3 Simulation

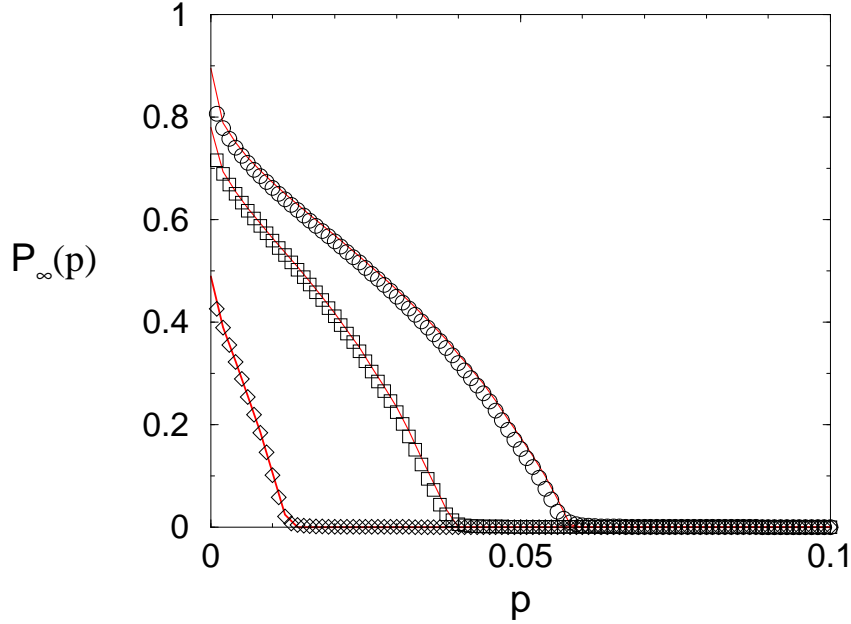
To simulate the intentional attack the same method as in random removal of sites is applied. The only difference is that the removed sites are not selected randomly, but chosen as the highest degree nodes. Sorting the sites would take  $O(N \ln N)$  operations, however since no site has degree greater than  $N$  an array of lists of sites for each degree can be produced, and the sites can be removed starting from the highest degree downwards. This only requires  $O(N)$  operations, and therefore the entire execution of the program takes only order of  $O(N)$  operations. Analytical and numerical result are shown in figure 7.

A somewhat different result would be achieved if the list is updated every time a site is deleted, since this influences the degree of the other sites. However, simulations show that this change has a small influence on the results, and the above analysis (in the previous Section) is based on the static picture. Therefore, this is the method used to obtain the results given here.

## 7 Critical Exponents

### 7.1 Introduction

In the study of percolation [16,15] and many other critical phenomena it is very common to study the behavior of the system near the critical point. Many parameters of the system (such as the order parameter – in our case  $P_\infty$ ) display



**Fig. 7.** Fraction of sites belonging to the spanning cluster,  $P_\infty$ , as a function of the fraction of removed sites,  $p$ , for networks with  $\lambda = 2.5$  (circles),  $\lambda = 2.8$  (squares), and  $\lambda = 3.3$  (diamonds). Lines represent the analytical result from Eqs. (15) and (16). Both the simulation and analysis are for system size  $N = 500,000$ .

a power-law behavior near the critical point. That is, many properties of the system behave as  $(p - p_c)^\omega$  near the transition point with a critical exponent  $\omega$ . At the transition point itself, some properties also follow a power law in another parameter.

It is well known [15,16] that many of the critical exponents are universal. That is, the value of the exponent is independent of the exact microscopic setup, and depends only on the large scale topological properties of the problem (*e.g.* the embedding dimension). In many systems there is an upper critical dimension above which the behavior of the critical exponents is independent of the dimension and all obtain their “mean-field” values. In the case of percolation the critical dimension is  $d_c = 6$  [15,16].

In the following we calculate the critical exponents for percolation in scale free networks and show that those networks, despite their mean-field nature, present a new universality class [33,27] (in fact, a new set of universality classes).

## 7.2 Infinite Cluster Size

In [3,5] the generating functions  $G_0(x)$  and  $G_1(x)$  are built for various distributions [Eqs. (11) and (12)]. Let  $H_1(x)$  be the generating function for the proba-

bility of reaching a branch of a given size by following a link. After a dilution of a fraction  $p$  of the sites (the remaining concentration is  $q = 1 - p$ ),  $H_1(x)$  satisfies the self-consistent equation

$$H_1(x) = 1 - q + qxG_1(H_1(x)) . \quad (46)$$

Since  $G_0(x)$  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

$$H_0(x) = 1 - q + qxG_0(H_1(x)) . \quad (47)$$

$H_0(1)$  is the probability that a site belongs to a cluster of any *finite* size. Thus, below the percolation transition  $H_0(1) = 1$ , while above the transition there is a finite probability that a site belongs to the infinite spanning cluster:  $P_\infty = 1 - H_0(1)$ . It follows that

$$P_\infty(q) = q \left( 1 - \sum_{k=0}^{\infty} P(k) u^k \right) , \quad (48)$$

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

$$u = 1 - q + \frac{q}{\langle k \rangle} \sum_{k=0}^{\infty} k P(k) u^{k-1} . \quad (49)$$

This equation can be solved numerically and the solution may be substituted into Eq. (48), yielding the size of the spanning cluster in a network of arbitrary degree distribution, at dilution  $q$  [3].

We now compute the order parameter critical exponent  $\beta$ . 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  $\beta = 1$  [16,15,34]. This regular mean-field result is not always valid, however, for scale-free networks. Eq. (48) has no special behavior at  $q = q_c$ ; the singular behavior comes from  $u$ . Also, at criticality  $P_\infty = 0$  and Eq. (48) imply that  $u = 1$ . We therefore examine Eq. (49) for  $u = 1 - \epsilon$  and  $q = q_c + \delta$ :

$$1 - \epsilon = 1 - q_c - \delta + \frac{(q_c + \delta)}{\langle k \rangle} \sum_{k=0}^{\infty} k P(k) (1 - \epsilon)^{k-1} . \quad (50)$$

The sum in (50) has the asymptotic form

$$\sum_{k=0}^{\infty} k P(k) (1 - \epsilon)^{k-1} \sim \langle k \rangle - \langle k(k-1) \rangle \epsilon + \frac{1}{2} \langle k(k-1)(k-2) \rangle \epsilon^2 + \dots + c \Gamma(2 - \lambda) \epsilon^{\lambda-2} , \quad (51)$$

where the highest-order analytic term is  $O(\epsilon^n)$ ,  $n = \lfloor \lambda - 2 \rfloor$ . Using this in Eq. (50), with  $q_c = 1 / (\kappa - 1) = \langle k \rangle / \langle k(k-1) \rangle$ , we get

$$\frac{\langle k(k-1) \rangle^2}{\langle k \rangle} \delta = \frac{1}{2} \langle k(k-1)(k-2) \rangle \epsilon + \dots + c \Gamma(2 - \lambda) \epsilon^{\lambda-3} . \quad (52)$$

The divergence of  $\delta$  as  $\lambda < 3$  confirms the vanishing threshold of the phase transition in that regime. Thus, in the case  $\lambda > 3$ , keeping only the dominant term as  $\epsilon \rightarrow 0$ , Eq. (52) implies

$$\epsilon \sim \begin{cases} \left( \frac{\langle k(k-1) \rangle^2}{c \langle k \rangle \Gamma(2-\lambda)} \right)^{\frac{1}{\lambda-3}} \delta^{\frac{1}{\lambda-3}} & 3 < \lambda < 4, \\ \frac{2 \langle k(k-1) \rangle^2}{\langle k \rangle \langle k(k-1)(k-2) \rangle} \delta & \lambda > 4. \end{cases} \quad (53)$$

Returning to  $P_\infty$ , Eq. (48), we see that the singular contribution in  $\epsilon$  is dominant only for the range of  $\lambda < 2$ . For  $\lambda > 3$ , we find  $P_\infty \sim q_c \langle k \rangle \epsilon \sim (q - q_c)^\beta$ . Comparing this to (53) we finally obtain

$$\beta = \begin{cases} \frac{1}{\lambda-3} & 3 < \lambda < 4, \\ 1 & \lambda > 4. \end{cases} \quad (54)$$

We see that the order parameter exponent  $\beta$  attains its regular mean-field value only for  $\lambda > 4$ . Moreover, since  $\beta > 1$  for  $\lambda < 4$  the percolation transition is smooth and is higher than 2nd-order: for  $3 + \frac{1}{n-1} < \lambda < 3 + \frac{1}{n-2}$  the transition is of the  $n$ th-order (since all derivatives up to the  $n - 1$ th are continuous). The result (54) has been reported before in [4], and also found independently in a different but related model of virus spreading [35,36]. The existence of an infinite-order phase transition at  $\lambda = 3$  for growing networks of the Albert-Barabási model, has been reported elsewhere [37,38]. These examples suggest that the critical exponents are universal and not model-dependent but depend only on  $\lambda$ .

For networks with  $\lambda < 3$  the transition still exists, though at a vanishing threshold,  $q_c = 0$ . The sum in Eq. (50) becomes:

$$\sum_{k=0}^{\infty} k P(k) u^{k-1} \sim \langle k \rangle + c \Gamma(2-\lambda) \epsilon^{\lambda-2}. \quad (55)$$

Using this in conjunction with Eq. (49), and remembering that here  $q_c = 0$  and therefore  $q = \delta$ , leads to

$$\epsilon = \left( \frac{-c \Gamma(2-\lambda)}{\langle k \rangle} \right)^{\frac{1}{3-\lambda}} \delta^{\frac{1}{3-\lambda}}, \quad (56)$$

which implies

$$\beta = \frac{1}{3-\lambda}, \quad 2 < \lambda < 3. \quad (57)$$

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  $\lambda$ -dependent in  $2 < \lambda < 3$ , and the amplitude of  $P_\infty$  diverges as  $\lambda \rightarrow 2$  (but remains finite as  $\lambda \rightarrow 4$ ).

### 7.3 Finite cluster size distribution

Next, we determine the exponents  $\tau$  and  $\sigma$ , which determine  $n_s$ , the number of clusters of size  $s$ .

In [5] it was shown that for a random graph of arbitrary degree distribution the finite clusters follow the usual scaling form:

$$n_s \sim s^{-\tau} e^{-s/s^*} . \quad (58)$$

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  $\tau$ . The probability that a site belongs to an  $s$ -cluster is  $p_s = sn_s \sim s^{1-\tau}$ , and is generated by  $H_0$ :

$$H_0(x) = \sum p_s x^s . \quad (59)$$

The singular behavior of  $H_0(x)$  stems from  $H_1(x)$ , as can be seen from Eq. (47).  $H_1(x)$  itself can be expanded from Eq. (46), by using the asymptotic form (51) of  $G_1$ . We let  $x = 1 - \epsilon$ , as before, but analyze at the critical point,  $q = q_c$ . With the notation  $\phi(\epsilon) = 1 - H_1(1 - \epsilon)$ , we finally get (note that at criticality  $H_1(1) = 1$ ):

$$-\phi = -q_c + (1 - \epsilon)q_c \left[ 1 - \frac{\phi}{q_c} + \frac{\langle k(k-1)(k-2) \rangle}{2\langle k \rangle} \phi^2 + \dots + c \frac{\Gamma(2-\lambda)}{\langle k \rangle} \phi^{\lambda-2} \right] . \quad (60)$$

From this relation we extract the singular behavior of  $H_0$ :  $\phi \sim \epsilon^y$ . Then, using Tauberian theorems [39] it follows that  $p_s \sim s^{-1-y}$ , hence  $\tau = 2 + y$ .

For  $\lambda > 4$  the term proportional to  $\phi^{\lambda-2}$  in (60) may be neglected. The linear term  $\epsilon\phi$  may be neglected as well, due to the factor  $\epsilon$ . This leads to  $\phi \sim \epsilon^{1/2}$  and to the usual mean-field result

$$\tau = \frac{5}{2}, \quad \lambda > 4 . \quad (61)$$

For  $\lambda < 4$ , the terms proportional to  $\epsilon\phi$ ,  $\phi^2$  may be neglected, leading to  $\phi \sim \epsilon^{1/(\lambda-2)}$  and [33]

$$\tau = 2 + \frac{1}{\lambda-2} = \frac{2\lambda-3}{\lambda-2}, \quad 2 < \lambda < 4 . \quad (62)$$

Note that for  $2 < \lambda < 3$  the percolation threshold is strictly  $q_c = 0$ . In that case we analyze at  $q = \delta$  small but fixed, taking the limit  $\delta \rightarrow 0$  at the very end. For the case  $2 < \lambda < 3$ ,  $\tau$  in Eq. (62) 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. (47) leading to  $\tau = 3$  for this range.

For growing networks of the Albert-Barabási model with  $\lambda = 3$ , it has been shown that  $sn_s \propto (s \ln s)^{-2}$  [38]. This is consistent with  $\tau = 3$  plus a logarithmic correction. Related results for scale free *trees* have been presented in [40].



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

$$S \propto N^{\tau-1} = N^{(\lambda-2)/(\lambda-1)}. \quad (63)$$

For  $\lambda = 4$  this reduces to the known  $N^{2/3}$ , termed by Erdős the “double jump”, due to the transition of the largest cluster from order  $\ln N$  for  $q < q_c$ , to  $N^{2/3}$  at  $q = q_c$ , to order  $N$  at  $q > q_c$  [18]. For  $\lambda \rightarrow 3$ ,  $S \propto N^{1/2}$ . It is not yet clear whether the results have a meaningful interpretation for  $\lambda < 3$ .

#### 7.4 Finite cluster size cutoff

The critical exponent  $\sigma$ , for the cutoff cluster size, can also be derived. Finite-size scaling arguments predict [16] that

$$q_c(\infty) - q_c(N) \sim N^{-\frac{1}{d\nu}} = N^{-\frac{\sigma}{\tau-1}}, \quad (64)$$

where  $N$  is the number of sites in the network,  $\nu$  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 (17) one obtains [2]

$$\kappa \approx \left( \frac{2 - \lambda}{3 - \lambda} \right) \frac{K^{3-\lambda} - m^{3-\lambda}}{K^{2-\lambda} - m^{2-\lambda}}, \quad (65)$$

where  $K \sim N^{1/(\lambda-1)}$  is the largest site degree of the network. For  $3 < \lambda < 4$ , this and Eq. (9) yield

$$q_c(\infty) - q_c(N) \sim \Delta\kappa \sim K^{3-\lambda} \sim N^{\frac{3-\lambda}{\lambda-1}}, \quad (66)$$

which in conjunction with Eq. (64) leads to

$$\sigma = \frac{\lambda - 3}{\lambda - 2}, \quad 3 < \lambda < 4. \quad (67)$$

For  $\lambda > 4$  we recover the regular mean-field result  $\sigma = 1/2$ . Note that Eqs. (64), (54), (62) are consistent with the known scaling relation:  $\sigma\beta = \tau - 2$  [16,15,34]. For  $2 < \lambda < 3$ ,  $q_c(\infty) = 0$  and  $q_c(N) \sim K^{\lambda-3} \sim N^{(\lambda-3)/(\lambda-1)}$  and therefore

$$\sigma = \frac{3 - \lambda}{\lambda - 2}, \quad 2 < \lambda < 3, \quad (68)$$

again consistent with the scaling relation  $\sigma\beta = \tau - 2$  (cf Eq. (57)).

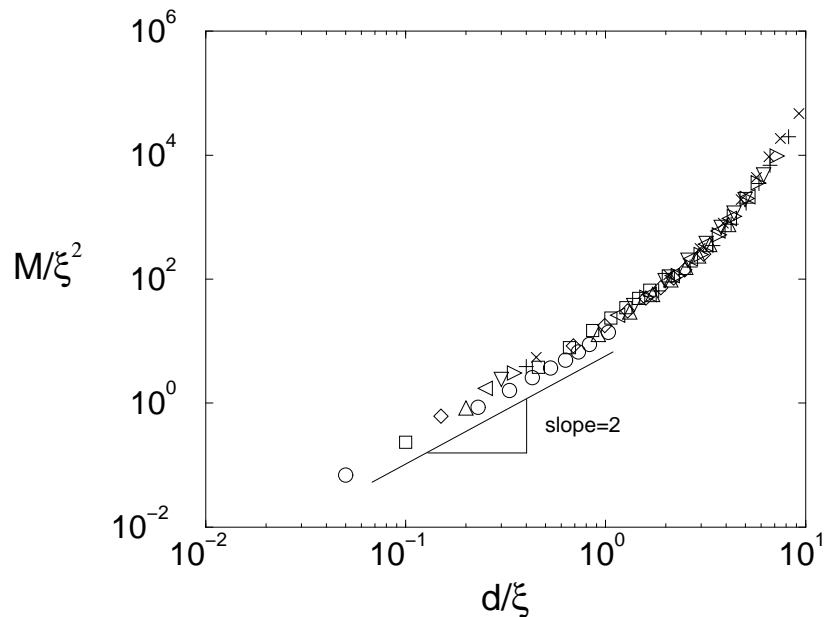
#### 7.5 Fractal Dimension

It is well known that on a random network in the well connected regime, the average distance between sites is of order  $\log_k N$  [18,41,5]. Even smaller distances

have been shown to hold for scale-free networks [21,22]. 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 a distance along the graph (which is the shortest distance along bonds). It is known from infinite-dimensional percolation theory that the chemical fractal dimension at criticality is  $d_1 = 2$  [15]. Therefore the average (chemical) distance  $d$  between pairs of sites on the spanning cluster at criticality behaves as

$$d \sim \sqrt{M}, \quad (69)$$

where  $M$  is the number of sites in the spanning cluster. This is analogous to percolation in finite dimensions, where in length-scales smaller than the correlation length the cluster is a fractal with dimension  $d_1$  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  $\xi_l \approx |p_c - p|^{-\nu_l}$ .



**Fig. 8.** Scaled mass (number of sites)  $M$  as a function of scaled distance  $d$  on the spanning cluster, for Erdos-Renyi graphs above criticality ( $p < p_c$ ,  $\langle k \rangle > 1$ ). Different symbols correspond to different values of  $\langle k \rangle$  (1.01 – 1.06). The correlation length is  $\xi = |p - p_c|^{-1}$ . Note that for  $d/\xi < 1$ , the slope is 2, corresponding to the behavior in the critical regime, while for  $d/\xi > 1$ ,  $M$  grows exponentially with  $d$ , corresponding to the well connected regime.

Next, we calculate  $\nu_l$  for scale free networks with  $3 < \lambda < 4$ . Below the transition all clusters are finite and almost all finite clusters are trees. The correlation length can be defined using the formula [15]:

$$\xi_l^2 = \frac{\sum l^2 g(l)}{\sum g(l)}. \quad (70)$$

Where  $g(l)$ , the correlation function, is the mean number of sites on the same cluster at distance  $l$  from an arbitrary occupied site. The number of sites in the  $l$  shell can be seen to be approximately  $\langle k \rangle (\kappa - 1)^{l-1}$  [5]. Since  $\kappa - 1 = (\kappa_0 - 1)q$  and  $q_c = 1/(\kappa_0 - 1)$  we get  $g(l) = c(1 - \delta)^l$ , where  $\delta = q - q_c$ . This leads to  $\xi_l \sim (q - q_c)^{-1}$ , *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

$$P_r(k) = P(k)u^k, \quad (71)$$

where  $u$  is the solution of Eq. (49). Using this distribution we can define  $\kappa_r$  for the finite clusters. This adds a term proportional to  $\epsilon^{\lambda-3}$  to the expansion of  $\xi_l$ . But, since  $\delta \propto \epsilon^{\lambda-3}$  (53), this leads again to  $\nu_l = 1$ .

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

$$d_l = \frac{\lambda - 2}{\lambda - 3}. \quad (72)$$

Similar results have been obtained by Burda *et al.*[40] for scale-free trees. Since every path when embedded in a space above the critical dimension can be seen as a random walk it is known that  $\nu = \nu_l/2$  [15]. Therefore, the fractal dimension is,

$$d_f = \frac{1}{\nu\sigma} = 2\frac{\lambda - 2}{\lambda - 3}. \quad (73)$$

The dimension of the embedding space is,

$$d_c = \frac{1}{\nu\sigma(\tau - 1)} = 2\frac{\lambda - 1}{\lambda - 3}. \quad (74)$$

Those dimensions reduce to the known 2, 4, and 6, respectively, for  $\lambda = 4$ .

A direct method for calculating the chemical dimension is also possible. Denoting the generating function of the number of sites on the  $l$ th layer of some branch, as  $N_l(x)$ , we get

$$N_{l+1}(x) = G_1(N_l(x)). \quad (75)$$

We are interested in the behavior of the average number of sites at a chemical distance  $l$  for those branches the have at least  $l$  layers. Since we expand exactly at criticality, the average branching factor is exactly 1, and therefore  $N_l(1) = 1$  for any  $l$ . Therefore,  $A_l$ , the average number of sites for surviving branches is

$$A_l = \frac{1}{1 - N_l(0)}, \quad (76)$$

since  $N_l(0)$  is the probability of the branching process to die out before the  $l$ th layer. At criticality the branching process will die out with probability  $N_l(0) \rightarrow 1$  as  $l \rightarrow \infty$ , and therefore for large  $l$  we can take  $N_l(0) = 1 + \epsilon_l$ . Expanding  $G_1$  at criticality one obtains (Eqs. (50) and (51), with  $\delta = 0$ )

$$G_1(1 - \epsilon) = 1 - \epsilon + \frac{c\Gamma(2 - \lambda)}{\langle k^2 \rangle - \langle k \rangle} \epsilon^{\lambda-2} + \dots \quad (77)$$

Substituting  $N_l(0) = 1 - \epsilon_l$  into Eq. 75 one obtains

$$1 - \epsilon_{l+1} = 1 - \epsilon_l - \frac{c\Gamma(2 - \lambda)}{\langle k^2 \rangle - \langle k \rangle} \epsilon_l^{\lambda-2} + \dots \quad (78)$$

Guessing a solution of the form  $\epsilon_l \approx Bl^{-d}$  we get

$$B(l+1)^{-d} \approx B(l^{-d} - dl^{-d-1}) = Bl^{-d} - \frac{c\Gamma(2 - \lambda)}{\langle k^2 \rangle - \langle k \rangle} (Bl^{-d})^{\lambda-2}. \quad (79)$$

implying that  $d = 1/(\lambda - 3)$ , and  $N_l(0) \sim 1 - l^{-d}$ . Noting that the mass of the branch is the sum of the layers up to the  $l$ th one, we get  $d_1 = d + 1 = (\lambda - 2)/(\lambda - 3)$ , similar to the results obtained in Eq. (72). The scaling relation therefore also proves to be correct in this case.

## 8 Conclusions

We have studied the structural properties and stability of scale-free graphs. We define a ‘‘layer’’ in a network as the set of nodes at a given distance from a chosen node. We have found that the degree distribution of the nodes of a scale free network at each layer obeys a power law with an exponential cutoff. We derived equations for this exponential cutoff and compared them with simulations. We also model the behavior of the number of nodes at each layer, and explain the observed exponential decay in the outer layers of the network. We believe our findings can have dual importance. First, they can help in devising better network algorithms that take advantage of the network structure [42]. Second, our analytical findings suggest a simple local test for the validity of the power law model as an exact model of the Internet [13].

Furthermore, we have shown that many percolation properties of those graphs are different from generalized random graphs and from the regular mean-field percolation, due to their diverging moment. We have shown that scale free graphs with  $2 < \lambda < 3$  are resilient to random breakdown of almost 100% of the nodes, making this a favorable design for unmanaged networks, like the Internet, where nodes can be disconnected or fail unexpectedly. However, those networks are sensitive to intentional attack on the most important nodes. The lack of percolation threshold also makes those networks sensitive to virus propagation [35,43], while their attack vulnerability makes them amenable to targeted immunization of the highest degree nodes [44,45]. An efficient strategy for the immunization of

complex networks without any global information on the network topology has been introduced in [46,27].

We have also studied the behavior of these networks near the percolation transition. We have shown that scale free networks with  $\lambda > 4$  near the percolation transition behave similar to infinite dimensional percolation theory, with  $M \propto l^2$ , where there is a crossover to the exponential behavior,  $M \propto k^l$ , far from the threshold. Thus, networks near the critical point become sparser, and communication becomes inefficient, as packets have to travel a long distance (many routers) on the way to their destination.

We have also shown that the critical exponents behave differently in scale-free networks than in regular networks both in the regime  $3 < \lambda < 4$  where a transition occurs in a finite  $p_c$ , as well as in the regime  $2 < \lambda < 3$  where  $p_c \rightarrow 0$ . In particular  $\beta > 1$  in both those regimes, making the transition of higher order rather than of second order.

## Acknowledgments

Support from the NSF is gratefully acknowledged (DbA). The authors wish to thank Danny Dolev, Osnat Mokryn, Yuval Shavitt and Keren Erez for useful discussions.

## References

1. R. Albert, H. Jeong, and A. L. Barabási, *Nature*, **406**, 6794, 378 (2000).
2. R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).
3. D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).
4. R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **86**, 3682 (2001).
5. M. E. J. Newman, S. H. Strogatz, and D. J. Watts *Phys. Rev. E*, **64**, 026118 (2001).
6. M. Faloutsos, P. Faloutsos, and C. Faloutsos, *ACM SIGCOMM '99 Comput. Commun. Rev.* 29, 251 (1999).
7. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, *Computer Networks* **33**, 309 (2000).
8. Jeong H., Tombor B., Albert R., Oltvai Z. N. and Barabási A. L. *Nature*, **407**, 651 (2000).
9. R. Albert and A.-L. Barabasi, *Rev. of Mod. Phys.* **74**, 47 (2002).
10. S. N. Dorogovtsev, and J. F. F. Mendes, *Adv. in Phys.*, 51 (4), (2002).
11. M. E. J. Newman, *SIAM Review* **45**, 167 (2003).
12. J. F. F. Mendes, S. N. Dorogovtsev and A. F. Ioffe, *Evolution of Networks: From Biological Nets to the Internet and WWW*, (Oxford University Press, 2003).
13. R. Cohen, D. Dolev, S. Havlin, T. Kalisky, O. Mokryn and Y. Shavitt, *cond-mat/0305582*.
14. Jeong H., Mason S., Barabási A. L. and Oltvai Z. N. *Nature*, **411**, 41 (2001).

15. A. Bunde, and S. Havlin (editors), *Fractals and Disordered System* (Springer, New York, 1996).
16. D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, 2nd edition (Taylor and Francis, London, 1991).
17. M. Molloy and B. Reed, *Random Structures and Algorithms* **6**, 161 (1995).
18. B. Bollobás, *Random Graphs*. pp. 123-136 (Academic Press, London, 1985).
19. M. Molloy and B. Reed, *Combinatorics, Probability and Computing* **7**, 295 (1998).
20. W. Aiello, F. Chung and L. Lu, *Proc. 32nd ACM Symp. Theor. Comp.*, (2000).
21. R. Cohen and S. Havlin, *Phys. Rev. Lett.*, **90**, 058701 (2003).
22. S.N. Dorogovtsev, J. F. F. Mendes, and A.N. Samukhin, *Nuclear Physics B* **653** (3), 307 (2003).
23. B. Bollobas and O. Riordan, in *Handbook of Graphs and Networks*, Eds. S. Bornholdt and H. G. Schuster, (Wiley-VCH, 2002).
24. A. F. Rozenfeld, R. Cohen, D. ben-Avraham and S. Havlin, *Phys. Rev. Lett.* **89**, 218701 (2002).
25. Z. Burda and A. Krzywicki, *Phys. Rev. E* **67**, 046118 (2003).
26. C. P. Warren, L. M. Sander and I. M. Sokolov, *Phys. Rev. E* **66**, 56105 (2002).
27. R. Cohen, S. Havlin and D. ben-Avraham, chap. 4 in *Handbook of Graphs and Networks*, Eds. S. Bornholdt and H. G. Schuster, (Wiley-VCH, 2002).
28. S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Phys. Rev. E* **63**, 062101 (2001).
29. M. E. J. Newman, *Phys. Rev. Lett* **89**, 208701 (2002).
30. A. Vazquez and R. Pastor-Satorras and A. Vespignani *cond-mat/0206084* (2002).
31. L. A. Adamic, R. M. Lukose, A. R. Punyani and B. A. Huberman, *Phys. Rev. E* **64**, 046135 (2001).
32. D. J. Watts, P. S. Dudds and M. E. J. Newman, *Science* **296**, 1302 (2002).
33. R. Cohen, D. ben-Avraham and S. Havlin, *Phys. Rev. E* **66**, 036113 (2002).
34. D. ben-Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge University Press, 2000).
35. R. Pastor-Sattoras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
36. Y. Moreno, R. Pastor-Satorras, and A. Vespignani *Eur. Phys. J. B* **26**, 521 (2002).
37. D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz, *Phys. Rev. E*, **64**, 041902 (2001).
38. S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Phys. Rev. E* **64**, 066110 (2001).
39. G. H. Weiss, *Aspects and Applications of the Random Walk* (North-Holland, Amsterdam, 1994).
40. Z. Burda, J. D. Curreira, and A. Krzywicki, *Phys. Rev. E* **64**, 046118 (2001).
41. F. Chung and L. Lu, *Adv. Appl. Math.*, **26**, 257, (2001).
42. D. Dolev, O. Mokryn and Y. Shavitt, *IEEE INFOCOM'03* (2003).
43. R. Pastor-Sattoras and A. Vespignani, *Phys. Rev. E* **63**, 066117 (2001).
44. R. Pastor-Satorras, and A. Vespignani *Phys. Rev. E* **65**, 036104 (2002).
45. Z. Dezsó, A.-L. Barabási, *Phys. Rev. E* **65**, 055103 (R) (2002).
46. R. Cohen, D. ben-Avraham and S. Havlin, *cond-mat/0207387*.