Stability and Topology of Scale-Free Networks under Attack and Defense Strategies

Lazaros K. Gallos,[1] Reuven Cohen,[2] Panos Argyrakis,[1] Armin Bunde,[3] and Shlomo Havlin[2]

[1]*Department of Physics, University of Thessaloniki, 54124 Thessaloniki, Greece*

[2]*Minerva Center and Department of Physics, Bar-Ilan University, 52900 Ramat-Gan, Israel*

[3]*Institut für Theoretische Physik III, Justus-Liebig-Universität Giessen, Heinrich-Buff-Ring 16, 35392 Giessen, Germany*

(Received 16 September 2004; published 10 May 2005)

We study tolerance and topology of random scale-free networks under attack and defense strategies that depend on the degree k of the nodes. This situation occurs, for example, when the robustness of a node depends on its degree or in an intentional attack with insufficient knowledge of the network. We determine, for all strategies, the critical fraction p_c of nodes that must be removed for disintegrating the network. We find that, for an intentional attack, little knowledge of the well-connected sites is sufficient to strongly reduce p_c. At criticality, the topology of the network depends on the removal strategy, implying that different strategies may lead to different kinds of percolation transitions.

DOI: 10.1103/PhysRevLett.94.188701 PACS numbers: 89.75.Hc, 87.23.Ge, 89.20.Hh

The observation that many real networks, such as the Internet, the WWW, social and biological networks, etc., obey a power-law distribution in their nodes connectivity has inspired a new area of research [1–8]. Such a network is constructed by nodes connected with links, where the probability $P(k)$ that a node has k links is

$$P(k) \sim k^{-\gamma}, \tag{1}$$

where γ is usually found to be between $2 < \gamma < 3$.

The scale-free character of these networks, represented by having no characteristic number of nodes per link, gives rise to many different and usually unexpected results in many properties, as compared to lattice models or even to small-world networks [1]. One important feature studied is the robustness of such a network under a random node removal [9]. In general, the integrity of a network is destroyed after a critical percentage p_c of the system nodes has been removed. For scale-free networks it has been shown that $p_c = 1$; i.e., in order to destroy the network practically all the nodes have to be removed [10].

In this Letter, we consider scale-free networks where the robustness of a node depends on its connectivity. This means that the probability of damaging a node either by some kind of failure or by an external attack depends on the degree k of the node. Examples are computer networks and social networks. In a computer network such as the Internet, usually the hubs that serve many computers are built in a more robust way, so that their probability of failure is smaller than for the others. In a social network, members of a group that have more links to others have a lower probability of leaving the group. It is, however, also possible that nodes with more links are less robust. For example, traffic on a network induces high loads on highly connected nodes [11], which in turn makes them more vulnerable to failures. In some cases breakdowns are due to cascades of failures caused by the dynamics of damage spreading [12]. In computer networks many breakdowns are due to congestion building (see [13]).

The above examples represent *internal* properties of a network, where the vulnerability of each node depends on its degree. In addition, the probability of removing a node can depend on its degree also due to *external* attack strategy. For example, the most efficient attack is an intentional attack where the highest degree nodes are being removed with probability one. In this case, only a small fraction p of removed nodes is sufficient to destroy the network. This strategy, however, requires full knowledge of the network topology in order to identify the highest connected nodes. In many realistic cases, this information is not available, and only partial knowledge exists. Accordingly, in an intentional attack the high-degree nodes can be removed only with a certain probability that will depend on k. In some networks, such as terror or Mafia networks those that are higher in the hierarchy have more links and are less known, and therefore the probability to remove them is smaller than those with less links. In contrast, in normal social networks the situation is opposite: The better linked members are more visible and therefore have a higher probability to be attacked. Finally, our study also applies to immunization strategies, where the high-degree nodes are not always known in advance; see also Dezsö and Barabási [14].

In all these examples, a value $W(k_i)$ is assigned to each node, which represents the probability that a node i with k_i links becomes inactive either by failure or under some attack. Specifically, we focus on the family of functions,

$$W(k_i) = \frac{k_i^{\alpha}}{\sum_{i=1}^{N} k_i^{\alpha}}, \qquad -\infty < \alpha < \infty. \tag{2}$$

The parameter α can be the sum of two parameters, $\alpha = \alpha' + \alpha''$, which incorporates both intrinsic network vulnerability (α') and external knowledge of the system (α''). It is possible that although some highly connected nodes are intrinsically vulnerable ($\alpha' > 0$), an internal defense strategy with $\alpha'' < 0$ will give the net result $\alpha < 0$ [15]. In this case, nodes with lower k are more vulnerable, while for

 © 2005 The American Physical Society

$\alpha > 0$, nodes with larger k are more vulnerable. The cases $\alpha = 0$ and $\alpha \to \infty$ represent the known random removal [9,10] and the targeted intentional attack [9,10], respectively.

The choice of α can be related to the probability w that in each attack, one of the n highest connected nodes in a network of N nodes becomes inactive by failure or attack. By definition

$$w = \frac{\int_{k_n}^{k_{\max}} P(k)W(k)dk}{\int_{m}^{k_{\max}} P(k)W(k)dk}, \tag{3}$$

where m denotes the minimum number and $k_{\max} \sim N^{1/(\gamma-1)}$ the maximum number of links a node can have. k_n is the minimum number of links of a node that belongs to the n highest connected nodes. It is easy to verify that for $N \gg n \gg 1$, w is related to α and γ by

$$w = \frac{1 - n^{(\gamma-1-\alpha)/(\gamma-1)} m^{\alpha+1-\gamma}}{1 - N^{(\gamma-1-\alpha)/(\gamma-1)} m^{\alpha+1-\gamma}}. \tag{4}$$

Depending on the value of α, for a fixed value of γ, we can distinguish three different regimes.

(i) $\alpha > \gamma - 1$: Here, $w = 1$, and an ''attacker'' is capable of destroying all the highest nodes in the network. Even though α is finite here, this case is fully equivalent to the targeted intentional attack $\alpha \to \infty$.

(ii) $\alpha = \gamma - 1$: For this particular value of α, w depends logarithmically on n and N, $w = \ln n / \ln N$.

(iii) $\alpha < \gamma - 1$: Here, w decreases with n/N by a power law, $w = (n/N)^{1-\alpha/(\gamma-1)}$. A special case is $\alpha = 0$, where effectively nodes are picked randomly and thus $w = n/N$. In this case, it is difficult to destroy the network, and the percolation threshold p_c can even be $p_c = 1$ for well-connected networks ($\gamma < 3$) [9,10]. For $0 < \alpha \leq \gamma - 1$, the effective removal is better than random; w still approaches zero for N approaching infinity, but slower than in the random case. As is shown later, this is enough for the percolation threshold to be significantly smaller than 1. Finally, for negative values of α, the fraction w of the highest connected nodes that will be damaged decreases much faster than in the random case. This feature will have no effect on the percolation threshold for $2 < \gamma \leq 3$, where $p_c = 1$ already in the random case. But for $\gamma > 3$, p_c increases with increasing negative values of α, as we show below.

From the above discussion it is clear that the regime $\alpha > \gamma - 1$ is in the same universality class as the intentional attack, and $\alpha = 0$ is identical to the random attack. It is not clear if there are more universality classes and if not, what is the border line between them. To study this question, we have considered both numerical simulations and analytical considerations. First, we determine the percolation thresholds as a function of the network parameters γ, N, and m, and of the ''attack'' parameter α. Then we study the universality classes by analyzing the topology of the network at these thresholds.

In the numerical treatment and the analysis we use random mixing for the network construction, where no correlations exist between the connectivity of neighboring sites. We first construct the network for a given γ. We fix the number of nodes N, as well as m and assign the degree k (number of links) for each node by drawing a random number from a power-law distribution $P(k) \sim k^{-\gamma}$. We then randomly select pairs of links between nodes that have not yet reached their preassigned connectivity and have not already been directly linked to each other. We repeat this selection until the entire network has been created.

To find the percolation thresholds p_c, we choose successively nodes with probability $W(k)$ [see Eq. (2)] and remove them. When a node is removed all its links are cut. After each removal, we calculate $\langle k^2 \rangle$ and $\langle k \rangle$. If $\kappa = \langle k^2 \rangle / \langle k \rangle \geq 2$, a spanning cluster exists in the network. We repeat this procedure for a large number of configurations (typically 100–300). For each concentration p of removed nodes we determine the probability F_∞ that a spanning cluster does not exist. We obtain p_c from the condition $F_\infty = 1/2$, as is shown in Fig. 1 for $\gamma = 2.5$ and 3.5 and several α values between 4 and -1. The width of the dispersion in F_∞ gives an upper bound for the error bars for p_c.

In the analytical treatment, we assume that sites are chosen for deletion according to their initial connectivity. A site is chosen with probability $k^\alpha / N\langle k^\alpha \rangle$. After d deletion attempts, the probability that a site of connectivity k has not been deleted is

$$\rho(k) = \left(1 - \frac{k^\alpha}{N\langle k^\alpha \rangle}\right)^d \approx e^{-dk^\alpha / N\langle k^\alpha \rangle}. \tag{5}$$

The condition for the existence of a spanning cluster after

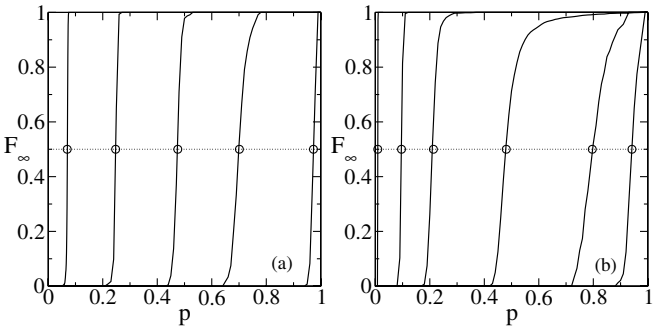

FIG. 1. The ratio of nonspanning configurations vs the fraction of removed nodes p. Lines are simulation data, from networks of $N = 10^6$ nodes, while the circles are the theoretical critical points. (a) Results from 100 different realizations of networks with $\gamma = 2.5$. Left to right: $\alpha = 4$, 1, 0.5, 0.25, and 0. (b) Results from 300 realizations of networks with $\gamma = 3.5$. Left to right: $\alpha = 4$, 1, 0.5, 0, -0.5, and -1.

the attack is as follows: a site is reached through a link with probability $kP(k)/\langle k\rangle$ and is still functional with probability $\rho(k)$. If the average number of outgoing links $(k-1)$ per site is larger than 1, a spanning cluster will exist. This consideration is formulated by

$$\sum_{k=m}^{k_{\max}} \frac{P(k)k(k-1)}{\langle k\rangle} e^{-q_c k^\alpha} = 1, \tag{6}$$

where $q \equiv d/N\langle k^\alpha\rangle$, and q_c is the value of q at criticality. To find the fraction of removed sites, one numerically solves Eq. (6) to calculate q_c and substitutes this value for calculating the critical threshold of removed sites:

$$p_c = \sum P(k)\rho(k) = \sum_{k=m}^{k_{\max}} P(k)\exp(-q_c k^\alpha). \tag{7}$$

Equation (7) describes the percolation threshold p_c for a given network of N nodes with exponent γ, as a function of the attack parameter α.

The lines in Fig. 2 represent the solutions of Eqs. (6) and (7) and are in excellent agreement with the simulations (in symbols). Remarkably, for $\gamma < 3$, p_c becomes smaller than 1 already for very small positive α values and decays rapidly with increasing α. Accordingly, by a very small preference probability to remove highly connected nodes, which arises, for example, in an intentional attack with very little knowledge of the network structure, this network can be destroyed by removing a comparatively small fraction of nodes. Above $\alpha = \gamma - 1$, p_c saturates, which means that the knowledge available to the attacker in this case is sufficient to destroy the network most efficiently. From Figs. 2(a) and 2(c) one might conclude that negative values of α might lead to $p_c = 1$ also for γ values above 3. We tested this question numerically and found that for all values of γ values between 3 and 4 and negative α, the percolation threshold p_c is below 1. For large negative values of α and $\gamma > 3$, the critical threshold can be approximated by $p_c = 1 - (\gamma-1)[2(\frac{\gamma-3}{\gamma-2})m^{2-\gamma}]^{(\gamma-1)/(\gamma-3)}$, which is below 1 for all $\gamma > 3$ and $m \geq 2$.

To study the effect of the different attack strategies on the topology of the network just before disruption, we analyzed the topology of the network at the percolation transition for different values of γ and α. We characterize the topology by the way the average shortest topological distance $\langle l\rangle$ between two nodes scales with the cluster size N_c. We expect that N_c scales with $\langle l\rangle$ as $N_c \sim \langle l\rangle^{d_\ell}$, where d_ℓ is the topological (''chemical'') dimension. Using a mean-field-type approximation [16], it has been suggested that for random removal,

$$d_\ell = \frac{\gamma-2}{\gamma-3}, \qquad 3 < \gamma < 4, \tag{8}$$

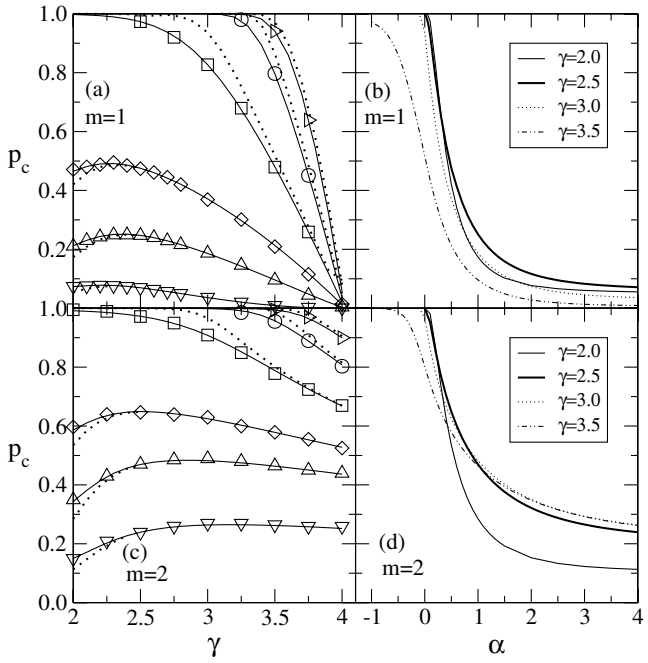

FIG. 2. (a) Values of p_c vs γ for different α values: (bottom to top) $\alpha = 4$, 1, 0.5, 0, -0.5, and -1. The symbols represent simulation data ($N = 10^6$ nodes) from 100–300 different network realizations. The solid lines are the theoretical predictions for finite-size networks, while the dashed lines correspond to infinite-size networks. Lower cutoff: $m = 1$. (b) Values of p_c vs α for infinite-size networks and different γ values. Lower cutoff $m = 1$. (c) The same as (a), with a lower cutoff $m = 2$. (d) The same as (b), with a lower cutoff $m = 2$.

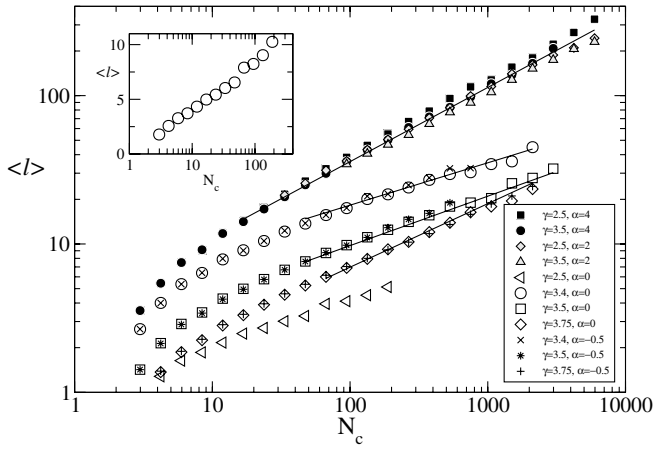

FIG. 3. Average shortest distance $\langle l\rangle$ between any two nodes of the giant cluster at criticality as a function of the cluster size N_c. The results correspond to networks of initially $N = 10^4$, 10^5, and 10^6 nodes. One thousand different configurations have been used for each N, except for $N = 10^6$ (100 configurations). The data have been logarithmically binned, and the results have been vertically shifted for presentation clarity. The values of α and γ are shown in the plot. The lines represent the theoretical slopes of (top to bottom) 1/2, 1/3.5, 1/3, and 1/2.33, respectively. Inset: semilogarithmic plot for the case of $\gamma = 2.5$ and $\alpha = 0$.

while for the intentional attack

$$d_\ell = 2, \qquad \gamma > 2. \tag{9}$$

The power-law dependence $\langle l \rangle \sim N_c^{1/d_\ell}$ is very different from the logarithmic dependence $\langle l \rangle \sim \log N$ found in scale-free networks for p well above p_c and $\gamma > 3$. Accordingly, due to the attacks, the network becomes very inefficient, since the distances between the nodes increase drastically, from a logarithmic to a power-law dependence on the total number of nodes. Equations (8) and (9) suggest that random removal and intentional attack are in different universality classes. This implies that the critical properties of the percolation transition depend on the way the transition is being approached, which is quite unusual in critical phenomena. To test these predictions and to see if there are further universality classes, we studied numerically how, at criticality, $\langle l \rangle$ scales with N_c, for different networks and different attack parameters α.

Figure 3 shows simulation results for $\langle l \rangle$ as a function of N_c for several values of γ and α. For $\alpha > 0$, the data scale quite nicely and yield a slope very close to $1/2$ (corresponding to $d_l = 2$) for all γ values, being identical to the theoretical prediction, Eq. (9), for the $\alpha = \infty$. This shows that all attacks with $\alpha > 0$ fall into the same universality class. The figure also verifies the prediction of Eq. (8) for random removal ($\alpha = 0$) when $\gamma > 3$. It shows, in addition, that attacks with $\alpha < 0$ result almost in the same network structure as for $\alpha = 0$ and yield the same topological dimension. Thus, for a given network with $\gamma > 3$ the same network can undergo transitions of two universality classes: (i) $\alpha > 0$ (universality class of the targeted intentional attack) and (ii) $\alpha \leq 0$ (universality class of random removal).

For $2 < \gamma < 3$ and $\alpha \leq 0$, the situation is less conclusive. Here, it is difficult to distinguish between a logarithmic or a power-law dependence, as can be seen from the inset of Fig. 3. Since for the pure network, $\langle l \rangle \sim \log \log N$ [16], also a simple logarithmic dependence $\langle l \rangle \sim \log N$ will lead to a strong increase of the mean distance between the nodes at criticality, as suggested by the inset.

In summary, we have studied the network tolerance under different attack strategies. We find that little knowledge on the highly connected nodes in an intentional attack reduces the threshold drastically compared with the random case. For example, when in a scale-free network with $\gamma = 2.5$ the 1% highest connected nodes are known with probability $w = 0.2$ (corresponding to $\alpha = 1$), the threshold reduces from $p_c = 1$ for the random case ($w = 0$) to $p_c \cong 0.25$. When all hubs are known ($w = 1$), p_c is close to 0.07. This shows that, for example, the Internet (see also [17]) can be damaged efficiently when only a small fraction of hubs is known to the attacker. Moreover, this result is also relevant for immunization of populations: Even if the virus spreaders are known (and immuned) with small probability, the spreading threshold can be reduced significantly. We also showed that even if the attack does not yet disintegrate the network, there is nevertheless a major damage on the network, since the distances between the nodes increase significantly and any transport process on the net may become inefficient. Our results show that the topology of the network close to criticality, characterizing the universality class of the phase transition, depends on the strategy of node removal.

This work was supported by a European research NEST Project No. DYSONET 012911, by the DAAD, and by the Israel Science Foundation.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] J. F. F. Mendes and S. N. Dorogovtsev, *Evolution of Networks: From Biological Nets to the Internet and the WWW* (Oxford University Press, Oxford, 2003).
[3] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, England, 2004).
[4] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[5] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[6] R. Albert and A.-L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000).
[7] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[8] F. Liljeros *et al.* Nature (London) **411**, 907 (2001).
[9] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000); **409**, 542 (2001).
[10] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000);**86**, 3682 (2001).
[11] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004); A. Barrat, M. Barthélemy, and A. Vespignani, Phys. Rev. Lett. **92**, 228701 (2004).
[12] A. E. Motter and Y.-C. Lai, Phys. Rev. E **66**, 065102(R) (2002).
[13] Y. Moreno, R. Pastor-Satorras, A. Vazquez, and A. Vespignani, Europhys. Lett. **62**, 292 (2003).
[14] Z. Dezsö and A.-L. Barabási, Phys. Rev. E **65**, 055103(R) (2002).
[15] An example for a negative value of α is the breakdown of computer networks, where it can be assumed that low degree computers are more prone to malfunction due to hardware failure, maintenance, viruses, power outs, etc., than high-degree computers. We may assume, for instance, that in a university network, the downtime of the main routers having hundreds of connections is about 10 times less than of end units, and the hubs for the different local area networks are somewhere in between. In this case, the situation is similar to an attack with $\alpha = -1/2$.
[16] R. Cohen, S. Havlin, and D. ben-Avraham, in *Handbook of Graphs and Networks*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, New York, 2002), Chap. 4.
[17] L. Li, D. Alderson, W. Willinger, and J. Doyle, Comput. Commun. Rev. **34**, 3 (2004).