PHYSICAL REVIEW E **70**, 035101(R) (2004)

Attack vulnerability of scale-free networks due to cascading breakdown

Liang Zhao,[1,2] Kwangho Park,[1] and Ying-Cheng Lai[1,3]
[1]*Department of Mathematics and Statistics, Arizona State University, Tempe, Arizona 85287, USA*
[2]*Institute of Mathematics and Computer Science, University of São Paulo, São Paulo, Brazil*
[3]*Departments of Electrical Engineering and Physics, Arizona State University, Tempe, Arizona 85287, USA*
(Received 28 April 2004; published 3 September 2004)

The possibility that a complex network can be brought down by attack on a *single* or a very few nodes through the process of cascading failures is of significant concern. Here we investigate a recent model for cascading failures in complex networks and uncover a phase-transition phenomenon in terms of the key parameter characterizing the node capacity. For parameter value below the phase-transition point, cascading failures can cause the network to disintegrate almost entirely. We obtain a theoretical estimate for the phase-transition point and provide numerical support.

DOI: 10.1103/PhysRevE.70.035101 PACS number(s): 89.75.Hc, 89.20.Hh, 05.10.−a

Complex networks arise in natural systems and they are also an essential part of modern society. Many real complex networks were found to be heterogeneous with power-law degree distribution [1–3]: $P(k) \sim k^{-\gamma}$, where k is the number of links of a randomly chosen node in the network and γ is the scaling exponent. This power-law, or algebraic, distribution means that the probability for a subset of nodes to possess a large number of links is not exponentially small, in contrast to random networks. Mathematically, the power-law distribution means that statistical moments of the degree variable are generally not defined, hence the name of scale-free networks. Because of the ubiquity of scale-free networks in natural and manmade systems, the security of these networks, i.e., how failures or attacks affect the integrity and operation of the networks, has been of great interest since the discovery of the scale-free property. The work by Albert *et al.* [4] demonstrated that scale-free networks possess the robust-yet-fragile property, in the sense that they are robust against random failures of nodes but fragile to intentional attacks. However, the term fragility here means that a scale-free network can become disintegrated under attacks on a small but still appreciable set of nodes that include a substantial fraction of links in the network [5]. An attack on a single or a very few nodes will, in general, not bring down the network. This interesting result was actually obtained based purely on the scale-free architecture of the network. In other words, dynamics in the network, i.e., how information or load is distributed in the network, was not taken into account.

An intuitive reasoning based on the load distribution would suggest that, for a scale-free network, the possibility of breakdown triggered by attack on or failure of even only a single node cannot be ignored. Imagine such a network that transports some physical quantities, or load. Nodes with large numbers of links receive a relatively heavier load. Each node, however, has a finite capacity to process or transport load. In order for a node to function properly, its load must be less than the capacity at all times; otherwise the node fails. If a node fails, its load will be directed to other nodes, causing a redistribution of load in the network. If the failing node deals with a small amount of load, there will be little effect on the network because the amount of load that needs to be redistributed is small. This is typically the situation of random failure of nodes. However, if the failing node carries a large amount of load, the consequence could be serious because this amount of load needs to be redistributed and it is possible that for some nodes, the new load exceeds their capacities. These nodes will then fail, causing further redistributions of load, and so on. As a consequence, a large fraction of the network can be shutdown.

Cascading failures can occur in many physical systems. In a power transmission grid, for instance, each node (a generator) deals with a load of power. Removal of nodes, in general, can cause redistribution of loads over all the network, which can trigger a cascade of overloading failures. The recent massive power blackout caused by a series of seemingly unrelated events on August 14, 2003 in the northeastern United States and Canada seemed to have the characteristics of cascading breakdown. Another example is the internet, where the load represents data packets and a node (router) is requested to transmit and overloading corresponds to congestion [6]. The rerouting of data packets from a congested router to another may spread the congestion to a large fraction of the network. Internet collapses caused by congestion have been reported [7]. With the possibility of cascading failures, a realistic concern is attacks on complex networks. In particular, for a scale-free network, the majority of the nodes deal with a small amount of load, so the probability for a node with a large amount of load to fail randomly is small. This, of course, will not be the case of intentional attacks that usually target one or a few of the most heavily linked nodes.

There have been a few recent studies on cascading failures in complex networks [8,9]. In Ref. [8], a simple mechanism was proposed to incorporate the dynamics of load in both random and scale-free networks. The model generates results that are completely consistent with the above intuition on cascading failures. For instance, it was demonstrated that random networks are robust against cascading breakdown, but it can be easily triggered by intentional attacks in scale-free networks. The existing results are, however, largely descriptive and qualitative. The purpose of this work is to address theoretically and numerically the fundamental

mechanism of cascading breakdown. To make analysis amenable, we focus on scale-free networks, use the load model in Ref. [8] that captures the essential features of cascading failures, and investigate cascades triggered by attack on a single node. Our finding is that cascading breakdown in scale-free networks can be understood in terms of a phase transition. In particular, let α be the tolerance parameter characterizing the capacity of nodes in the network. Cascading breakdown due to attack on a single node is possible only when α is below a critical value α_c. By making use of the degree distribution of scale-free networks and the concept of betweenness [10] to characterize the load distribution, we are able to derive a theoretical formula for estimating the phase-transition point α_c, which is verified by numerical experiments. In terms of practical utility, our result enables a possible implementation of predicting and preventing mechanism for cascading breakdown in scale-free networks.

The load dynamics in scale-free networks can be modeled, as follows. For a given network, suppose that at each time step one unit of the relevant quantity, which can be information, energy, etc., is exchanged between every pair of nodes and transported along the shortest path. To characterize the load distribution, the concept of betweenness is useful [10]. The load (or betweenness) at a node i is defined as the total number of shortest paths passing through this node. The capacity of a node is the maximum load that the node can handle. In manmade networks, the capacity is severely limited by cost. Thus, it is natural to assume that the capacity C_i of node i is proportional to its initial load L_i [8],

$$C_i = (1 + \alpha)L_i, \tag{1}$$

where the constant $\alpha \geqslant 0$ is the tolerance parameter. When all nodes are on, the network operates in a free-flow state insofar as $\alpha \geqslant 0$. But, the removal of nodes in general changes the distribution of shortest paths. The load at a particular node can then change. If it increases and becomes larger than the capacity, the node fails. Any failure leads to a new distribution of load and, as a result, subsequent failures can occur. The failures can stop without affecting too much of the connectivity of the network but it can also propagate and shutdown a considerable fraction of the whole network. Cascading failures can be conveniently quantified by the relative size of the largest connected component

$$G = \frac{N'}{N}, \tag{2}$$

where N and N' are the numbers of nodes in the largest component before and after the cascade, respectively. The integrity of the network is maintained if $G \approx 1$, while breakdown occurs if $G \approx 0$.

To obtain an analytic estimate of the critical value of the tolerance parameter, we focus on the situation where cascading failures are caused by attack on the node with the largest number of links and the failures lead to immediate breakdown of the network. That is, G becomes close to zero after one redistribution of the load. For a node in the network, its load is a function of the degree variable k. For scale-free networks, we have [11,12],

$$L(k) \sim k^{\eta}, \tag{3}$$

where $\eta > 0$ is a scaling exponent. To proceed, we write the degree distribution as $P(k) = ak^{-\gamma}$ and the load distribution as $L(k) = bk^{\eta}$, where a and b are positive constants. Let $k_{\max}$ be the largest degree in the network. Before the attack, we have

$$\int_1^{k_{\max}} P(k)dk = N \text{ and}$$

$$\int_1^{k_{\max}} P(k)L(k)dk = S, \tag{4}$$

where S is the total load of the network. These two equations give

$$a = \frac{(1-\gamma)N}{[k_{\max}^{1-\gamma} - 1]} \text{ and}$$

$$b = \frac{\beta S}{a(1 - k_{\max})^{-\beta}}, \tag{5}$$

where $\beta \equiv \gamma - \eta - 1$. After the removal of the highest degree node (it is only the first step of the whole cascading process), the degree and load distributions become $P'(k) = a'k^{-\gamma'}$ and $L'(k) = b'k^{\eta'}$, respectively. Since only a small fraction of nodes are removed from the network, we expect the changes in the algebraic scaling exponents of these distributions to be negligible. We thus write $P'(k) \approx a'k^{-\gamma}$ and $L'(k) \approx b'k^{\eta}$, where the proportional constants a' and b' can be calculated in the same way as for a and b. We obtain $a' = (1-\gamma)(N-1)/[k_{\max'}^{1-\gamma} - 1]$ and $b' = \beta S'/a'(1-k_{\max'})^{-\beta}$, where S' is the total load of the network after the attack. For nodes with k links, the difference in load before and after the attack can be written as $\Delta L(k) \approx (b'-b)k^{\eta} = [(b'/b) - 1]L(k)$. Given the capacity $C(k)$, the maximum load increase that the nodes can handle is $C(k) - L(k) = \alpha L(k)$. The nodes still function if $\alpha > [(b'/b) - 1]$ but they fail if $\alpha < [(b'/b) - 1]$. The critical value α_c of the tolerance parameter is then

$$\begin{aligned}
\alpha_c &= \frac{b'}{b} - 1 \\
&\approx \left(\frac{{k_{\max'}}^{1-\gamma} - 1}{{k_{\max}}^{1-\gamma} - 1}\right)\left(\frac{1 - {k_{\max}}^{-\beta}}{1 - {k_{\max'}}^{-\beta}}\right)\left(\frac{S'}{S}\right) - 1 \\
&\approx \left(\frac{1 - {k_{\max}}^{-\beta}}{1 - {k_{\max'}}^{-\beta}}\right)\left(\frac{S'}{S}\right) - 1 \\
&\approx \{1 - ({k_{\max}}^{-\beta} - {k_{\max'}}^{-\beta})\}\left(\frac{S'}{S}\right) - 1 \\
&= \left\{1 - {k_{\max'}}^{-\beta}\left[-1 + \left(\frac{k_{\max}}{k_{\max'}}\right)^{-\beta}\right]\right\}\left(\frac{S'}{S}\right) - 1,
\end{aligned} \tag{6}$$

where the third line of Eq. (6) is obtained from the second line by using the fact $({k_{\max'}}^{1-\gamma} - 1)/({k_{\max}}^{1-\gamma} - 1) \approx 1$. This is so because both ${k_{\max'}}^{1-\gamma}$ and ${k_{\max'}}^{1-\gamma}$ approach zero when

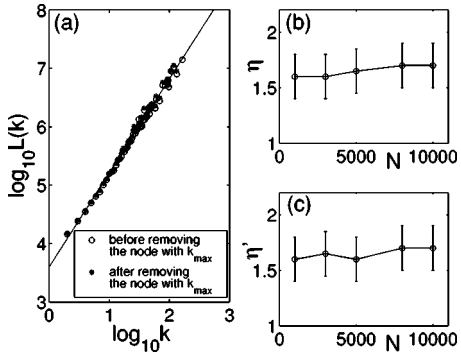

FIG. 1. (a) Algebraic scaling of the load $L(k)$ for a scale-free network of $N=10\,000$ nodes, $\gamma=3$, and $\langle k\rangle=4$. The open circles and the asterisks denote the load values before and after an intentional attack that removes the node with the maximum number of links. We have $\eta\approx\eta'\approx1.6$. (b), (c) algebraic scaling exponents of the load in relation to network size, before and after the nodes with the largest degree are removed, respectively. For each network size N, the resulting data were averaged over more than 20 realizations.

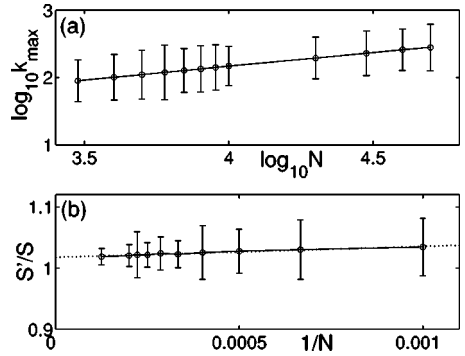

FIG. 2. (a) Algebraic scaling of $k_{\max}$ with N. For each network size N, 5000 realizations are averaged. (b) Load ratio S'/S vs $1/N$. For each network size N, more than 20 realizations were averaged.

$N\to\infty$ and $\gamma>1$. In the limit $N\to\infty$, we have $k_{\max'}^{-\beta}\sim0$, $k_{\max}/k_{\max'}\sim$ constant, and $S'/S\to1$, so $\alpha_c\approx0$, indicating that an infinite scale-free network cannot be brought down by a single attack if $\alpha>0$. On the other hand, for a finite-size network, since $k_{\max'}^{-\beta}>0$, we have $\alpha_c>0$, suggesting that breakdown can occur for $\alpha<\alpha_c$. The practical usage of Eq. (6) is that it provides a way to monitor the state of (finite) network to assess the risk of cascading breakdown. In particular, the critical value α_c can be computed in time and comparison with the predesigned tolerance parameter value α can be made. If α_c shows a tendency of increase and approaches α, an early warning can be issued to signal an immediate danger of network breakdown.

We now provide numerical support for the theoretical prediction Eq. (6). We generate scale-free networks by using the standard Barabási-Albert model [1], as detailed in Ref. [13]. The shortest paths and the load $L(k)$ are computed by using the algorithm developed by Newman [10]. Figure 1(a) shows the algebraic scaling of the load for a scale-free network. The scaling exponent of the degree distribution $P(k)$ is $\gamma\approx3$ (not shown) and the average number of links in the network is $\langle k\rangle=4$. The open circles in Fig. 1 indicate the values of the load for the original network. Apparently $L(k)$ follows the expected algebraic distribution, with exponent $\eta\approx1.6$. Figures 1(b) and 1(c) show the exponents η in relation to system size before and after the highest degree node is removed, respectively. In both cases, we obtained $\eta=\eta'=1.6(2)$. Computer simulations show that the load distribution and cascading behavior observed above hold for various $\langle k\rangle$. To simulate an intentional attack, we remove the node with the maximum number of links ($k_{\max}=81$ for the realization of the network shown in the figure). The distribution of the load is recalculated after the network stabilizes itself. That is, after the attack the load on the removed node is redistributed to the network and new load to every node is recalculated. Any node with load exceeding its capacity is removed and load is recalculated, and so on, until the process reaches a new equilibrium. The new values of the load are denoted by the asterisks in Fig. 1. We see that the distribution still follows a power law with approximately the same scaling exponent. This justifies the approximation $\eta'\approx\eta$ used in our theory.

As N is increased, we expect $k_{\max}$ to increase following an algebraic scaling law [13]. This behavior is shown in Fig. 2(a). After the attack and redistribution of load, we find that the ratio $k_{\max}/k_{\max'}$, where $k_{\max'}$ is the new value of the maximum number of links, is constant, regardless of the network size. We also numerically observed that the load ratio S'/S (before and after the attack) is approximately one for large N, as shown in Fig. 2(b).

Figure 3(a) shows cascading failures when a single node with a different degree is removed from the network. We see that, when a node with a small degree is removed, the G value remains close to one except when α is close to zero. However, when the node with the largest degree (in this case $k=81$) is removed, nearly total breakdown of the network, as represented by values of G close to zero, occurs when $\alpha<0.1$. The phase-transition point α_c is thus about 0.1. With numerical values of $k_{\max}=81$, $k_{\max'}=60$, $S\approx1.86\times10^7$, and $S'\approx1.91\times10^7$, theoretically predicted value of α_c in Eq. (6)

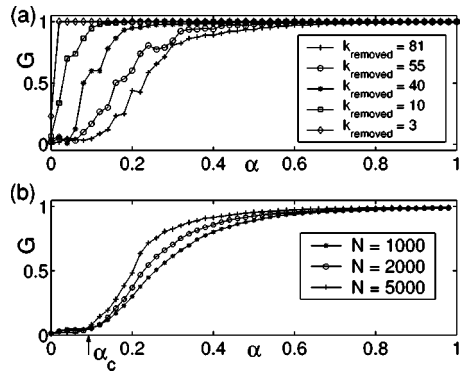

FIG. 3. Cascading failure in a scale-free network in relation to the tolerance parameter α. (a) Removal of the nodes with different number of links for $N=2000$. In the case of the removal of the node with the highest degree, the phase-transition point is $\alpha_c\approx0.1$, meaning that for $\alpha<\alpha_c$, the networks disintegrate almost entirely under intentional attack on a single node. (b) Phase transitions for networks of different sizes. The resulting data points were averaged over 30 realizations.

gives $\alpha_c \approx 0.1$, which is consistent with numerics. This phase transition phenomenon seems to be robust for different sizes of network, as shown in Fig. 3(b), G vs α for $N=1000$, $N=2000$, and $N=5000$, respectively.

What about attacks that target more than one node? In this case, we expect that the phase transition will occur for higher values of the tolerance parameter, because it becomes more difficult for the network to maintain its integrity at lower tolerance, as compared with the case of attack on a single node. Figure 4 shows G versus both α and $N_{trigger}$, the number of nodes that an attack targets. Here the removed nodes are those with the highest numbers of links. We see that, as $N_{trigger}$ is increased, the phase-transition point α_c also increases. Roughly we have $\alpha_c \sim N_{trigger}$. Note that the number of targeted nodes, while more than one, is still far smaller compared with the total number of nodes in the network. Practically, this means that, even if the network is designed to have a high tolerance by stipulating high capacities for its nodes, cascading failures triggered by attack on a very small subset of nodes are capable of bringing down the entire network.

In summary, we investigated cascading failures triggered by attacks on a single or a few nodes in scale-free networks in a more quantitative manner and focused on the fundamental and practically important question of whether such failures can lead to the disintegration of the network. Our finding is a phase-transition-like phenomenon in terms of the network tolerance parameter characterizing the node capacity, where the two distinct phases correspond to the situations

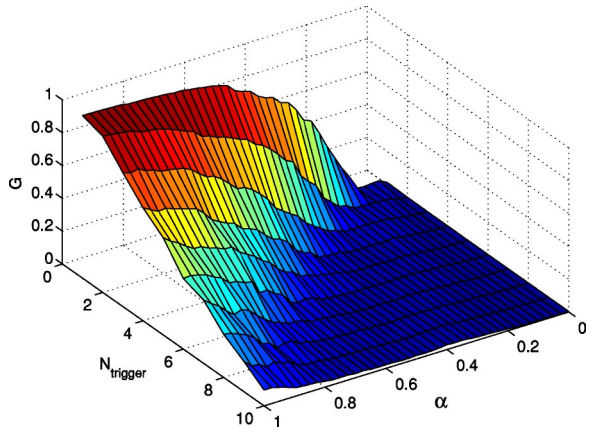

FIG. 4. (Color online) For a scale-free network of $N=2000$ nodes under attack targeting multiple nodes, G vs α and $N_{trigger}$, the number of targeted nodes. For each parameter value, G is averaged over 30 realizations.

where the network under attack remains largely integrated or disintegrated as a result of cascading failures. We obtained a theoretical estimate for the phase-transition point and provided a numerical check. These results should be useful in furthering studies in the important area of network security.

This work was supported by NSF under Grant No. ITR-0312131 and by AFOSR under Grant No. F49620-01-1-0317.

[1] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2002).
[5] R. Cohen, K. Erez, D. b-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000); *ibid.* **86**, 3682 (2001).
[6] A. Arenas, A. Días-Guilera, and R. Guimerà, Phys. Rev. Lett. **86**, 3196 (2001).
[7] V. Jacobson, Comput. Commun. Rev. **18**, 314 (1988).
[8] A. E. Motter and Y.-C. Lai, Phys. Rev. E **66**, 065102(R) (2002).
[9] P. Holme and B. J. Kim, Phys. Rev. E **65**, 066109 (2002); P. Holme, *ibid.* **66**, 036119 (2002).
[10] M. E. J. Newman, Phys. Rev. E **64**, 016132 (2001); Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[11] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001); M. Barthélemy, *ibid.* **91**, 189803 (2003); K.-I. Goh, C.-M. Ghim, B. Kahng, and D. Kim, *ibid.* **91**, 189804 (2003).
[12] K. Park, Y.-C. Lai, and N. Ye (to be published).
[13] A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272A**, 173 (1999).