

PHYSICS OF NETWORK SECURITY

Y.-C. LAI *

*Department of Electrical Engineering,
Department of Physics and Astronomy,
Arizona State University
Tempe, AZ 85287, USA
E-mail: yclai@chaos1.la.asu.edu*

X. WANG and C. H. LAI

*Department of Physics,
National University of Singapore, Singapore 117542*

In a scale-free network, there is a small subset of nodes with linkages far heavier than those of others in the network. Intentional attack on one or a few nodes in this group can trigger a *cascade* of node failures, leading to potential breakdown of the network on a large scale. This typically occurs for situations where the capacities of nodes in the network are small and overload results in node failure. Breakdown of the network can be understood as a phase transition that occurs when the node capacity parameter is decreased through a critical point. However, when the node capacities are sufficiently large, the network can be robust against cascading breakdown. A physical theory to account for these phenomena is reviewed. In situations where overload does not cause node failure but generate traffic congestion, attack can induce persistent oscillations of the network in the sense that its characterizing quantities, such as the diameter, oscillate in time and are never able to restore to their original values. This remarkable phenomenon of *network oscillation* is also discussed.

1. Introduction

Complex networks arising in many natural and man-made systems are scale-free in that their connectivity (or degree) distributions follow an algebraic law.^{1,2} In such a network, a small subset of nodes can be far more important than others in that the numbers of links, or the degrees, of these nodes can be significantly larger than those of the rest of the nodes in the network.

*Work partially supported by NSF under Grant No. ITR-0312131 and by AFOSR under Grant No. F49620-01-1-0317.

From the standpoint of security, this means attacks on nodes in this group can have a devastating effect on the overall integrability and function of the network.³ Assuming that the node capacity is finite and a node functions normally only when its load is smaller than the capacity, there are two distinct situations concerning the effect of attack: (1) overload causes a node to fail completely, and (2) overload does not cause node failure but instead generate traffic congestion at the node. A typical example of the former is electrical power grid, while examples of the latter include the internet and air transportation networks. The purpose of this article is to address the physics and dynamics of the security of scale-free networks mainly for the first case. [The second case will be discussed only briefly as the research is still ongoing at the National University of Singapore (NUS).]

In a scale-free network, since nodes with the higher degrees in the network typically handle more loads necessary for the normal operation of the network, an attack to disable one or few of these nodes means that their loads will be redistributed to other nodes. Because the amount of the redistributed loads can be large, this can cause other nodes in the network to fail, if their loads exceed their capacities, which in turn causes more loads to be redistributed, and so on. This cascading process can continue until the network becomes totally disintegrated. Indeed, simulations show, for instance, that for a realistic power-grid network, attack on a single node can disable more than half of the nodes, essentially shutting down the network.⁴

In Sec. 2, we describe a prototype model for attack-induced cascades on scale-free networks. In Sec. 3, we review a physical theory in terms of phase transition to understand the dynamical mechanism underlying the cascading process. In Sec. 4, we discuss a practical strategy to prevent cascading breakdown and derive theoretical criteria for designing networks that are immune to cascades. Summary and a discussion of network oscillations are presented in Sec. 5.

2. Model of cascades in complex networks

A prototype model based on load dynamics for cascading in complex networks is proposed recently.⁴ The load (or the betweenness⁶) at a node 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, which is assumed⁴ to be proportional to its initial load,

$$C_i = (1 + \alpha)L_0(i) \equiv \lambda L_0(i), \quad (1)$$

where $\alpha \geq 0$ is the capacity parameter. For a particular node, if the load on it increases and becomes larger than the capacity, the node fails. Any failure leads to a redistribution of loads over the network and, as a result, subsequent failures can occur. The failures can stop without affecting too much the network connectivity 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 $G = N'/N$, where N and N' are the numbers of nodes in the component before and after the cascade, respectively. The integrity of the network is maintained if $G \approx 1$, while breakdown at a global scale occurs if $G \approx 0$.

It is demonstrated^{4,7} that global cascades can occur if (1) the network exhibits a highly heterogeneous distribution of loads, (2) the removed node is among those with higher load, and (3) the capacity parameter is around or below a critical value. It is further demonstrated⁷ that, when the capacity parameter α is viewed as a control parameter, the occurrence of global cascades can be regarded as the consequence of a phase transition. In particular, as α is decreased through a critical value α_c , global cascades are highly probable.

Given a scale-free network, how to obtain theoretical estimate of the critical capacity-parameter value α_c for phase transition becomes an interesting issue. A formula is obtained⁷ that relates α_c to several basic quantities characterizing the network. The main idea leading to the formula will be presented in the next Section. Another interesting question concerns the robustness of the network against cascading breakdown. In particular, for sufficiently large value of α , global cascades are unlikely. The issue is to determine the lower bound α_s , where cascades are not possible for $\alpha > \alpha_s$. This may be important for network design under the constraint of limited resource, where one wishes to have networks that are immune to global cascades but at the same time do not wish to employ nodes with unnecessarily large capacities. These considerations are schematically illustrated in Fig. 1. Theoretical determination of α_s has been obtained recently,⁸ which will be discussed in Sec. 4.

3. Formula for phase-transition point α_c

It is convenient to consider the situation⁷ 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, the quantity 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

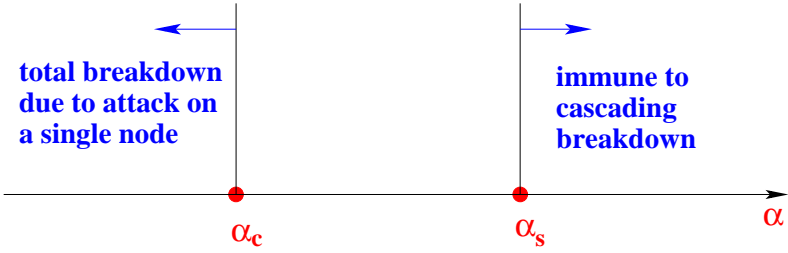


Fig. 1. Schematic illustration of the key capacity-parameter values α_c and α_s , the phase-transition point for global cascades and the lower bound for cascade-free networks, respectively.

scale-free networks, we have,^{5,9,10} $L(k) \sim k^\eta$, where $\eta > 0$ is a scaling exponent. More formally, the degree and the load distribution can be written as $P(k) = ak^{-\gamma}$ and $L(k) = bk^\eta$, respectively, where a and b are positive constants. Let k_{max} be the largest degree in the network. Before the attack, we have $\int_1^{k_{max}} P(k)dk = N$ and $\int_1^{k_{max}} P(k)L(k)dk = S$, where S is the total load of the network. Evaluating the integrals yields

$$a = \frac{(1-\gamma)N}{[k_{max}^{1-\gamma} - 1]} \quad \text{and} \quad b = \frac{\beta S}{a(1 - k_{max})^{-\beta}}, \quad (2)$$

where $\beta \equiv \gamma - \eta - 1$. After the removal of the highest degree node, the degree and the load distribution become $P'(k) = a'k^{-\gamma'}$ and $L'(k) = b'k^{\eta'}$, respectively. Since only a small fraction of nodes is 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 $C(k) - L(k) = \alpha L(k)$. A node still functions if $\alpha > (b'/b - 1)$ but it fails if $\alpha < (b'/b - 1)$. The critical value α_c of the tolerance parameter is then

$$\alpha_c = \frac{b'}{b} - 1 \approx \left\{ 1 - k_{max'}^{-\beta} \left(-1 + \left(\frac{k_{max}}{k_{max'}} \right)^{-\beta} \right) \right\} \left(\frac{S'}{S} \right) - 1, \quad (3)$$

where the fact $(k_{max'}^{1-\gamma} - 1)/(k_{max}^{1-\gamma} - 1) \approx 1$ has been used. This is so because both $k_{max'}^{1-\gamma}$ and $k_{max}^{1-\gamma}$ approach zero for $N \rightarrow \infty$ and $\gamma > 1$. In the limit $N \rightarrow \infty$, we have $k_{max'}^{-\beta} \sim 0$, $k_{max}/k_{max'} \sim \text{constant}$, and

$S'/S \rightarrow 1$, so $\alpha_c \approx 0$, indicating that an infinite scale-free network cannot be brought down by a single attack if $\alpha > 0$. 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. (3) 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 pre-designed capacity-parameter value α can be made. If α_c shows a tendency of increase and approaches α , early warning can be issued to signal an immediate danger of network breakdown. The validity of Eq. (3) has been established through extensive numerical computations.⁷

4. Determination of α_s : criterion for designing cascade-free networks

Theoretical determination of α_s is made possible by considering the parallel problem of how to prevent catastrophic cascades caused by attacks. A simple and intuitive method is to lower the average loads present in the network. This can be achieved by removing a small set of nodes that contribute to the loads in the network but they themselves otherwise process little load.¹¹ Removal of these nodes and all links connected to them will not affect the functioning of the network but will help enhance the load tolerance for each remaining node. When an intentional attack occurs to disable one or few influential nodes in the network, the load to each remaining node will increase but, because of the extra capacity gained through control, failure is less likely. For scale-free networks, cascades can be prevented or their sizes can be reduced significantly by intentionally removing carefully selected, a few percent of the nodes.¹¹

To formulate the problem quantitatively, we let ρ be the fraction of intentionally removed nodes. As ρ is increased from zero, the network becomes more robust against global cascading. However, this trend cannot continue indefinitely, for the network will become disintegrated (even without any attack) if ρ is too large. There exists then a critical value ρ_c for which the network's ability to sustain attack-induced cascading breakdown reaches maximum. By hypothesizing the equivalence of the cascade-control problem to the problem of designing cascade-free network, both α_s and ρ_c can be obtained in a single theoretical framework.⁸

The starting point of the analysis is again the observation that, for a scale-free network, its load distribution obeys^{5,7,10} algebraic scaling with the degree variable k . $L(k) \sim k^\eta$, where η and b are positive constants. After removing a small fraction ρ of low-degree nodes, the average connectivity

of the network changes little. Moreover, the degree distribution remains to be algebraic with approximately the same scaling exponent. The next step is to determine the relation between the load distributions before and after removing ρ fraction of low-degree nodes. For convenience, all nodes in the network are labeled by integers from 1 to N , while the removed nodes are labeled by $(1 - \rho)N$ to N . The total load before the removal can be written as $S = \sum_{i=1}^{(1-\rho)N} L_i + \sum_{i=N(1-\rho)}^N L_i \equiv S_0 + S_1$, where S_0 is the sum of loads of the remaining nodes before the removal and S_1 is the total load of the removed nodes. Because the removed nodes have relatively low degrees, we have $S_0 \gg S_1$ and, hence, $S \approx S_0 = \sum_{i=1}^{N(1-\rho)} L_i$. After the removal, the total load of the network is $S' = \sum_{i=1}^{N(1-\rho)} L'_i \approx \sum_{i=1}^{N(1-\rho)} \sigma L_i$, where $0 < \sigma < 1$ is a shifting constant. Since $S = N(N - 1)D \approx N^2 D$, $S' = N(1 - \rho)[N(1 - \rho) - 1]D' \approx (1 - \rho)^2 N^2$ and $D \approx D'$, where D and D' are the diameters of the networks before and after the removal, respectively, we have $\sigma \approx (1 - \rho)^2 \approx 1 - 2\rho$. Thus, on average, the difference between the loads of node i before and after the removal is $\Delta L_i = L_i - L'_i \approx 2\rho L_i$, which is independent of the parameter λ . Since, initially, the load tolerance of node i is $(\lambda - 1)L_i$ and the process of removal results in equivalently an extra amount of load tolerance $2\rho L_i$, the node will not fail unless the load increment due to an attack exceeds $(\lambda - 1 + 2\rho)L_i$. Controlled removal of ρ percent of low-degree nodes is thus equivalent to increasing the parameter λ to $\lambda + 2\rho$ in the original network.

The effect on G of removing a ρ fraction of low-degree nodes can be analyzed by noting that, in general, G depends on both λ and ρ : $G \equiv G(\lambda, \rho)$. However, without the controlled removal, G depends on λ only and we write $G(\lambda, 0) \equiv G^0(\lambda)$. Note that $G^0(\lambda)$ can be calculated once the network configuration is given. A detailed analysis of the relation between $G(\lambda, \rho)$ and $G^0(\lambda)$ yields⁸ the following formulas for estimating λ_s and ρ_c :

$$\frac{dG^0}{d\lambda} \Big|_{\lambda_s} \approx \frac{G^0(\lambda_s)}{2} \quad \text{and} \quad \rho_c \approx (\lambda_s - \lambda_0)/2, \quad (4)$$

where λ_0 is the initial value of the network capacity. The formulas have been verified numerically.⁸

5. Discussions

Studying attacks on complex networks can help identify the vulnerabilities of real-world networks, which can be used either for protective or destructive purposes. Examples of the former include critical infrastructural networks such as the internet and the power grid, while an example of the latter is

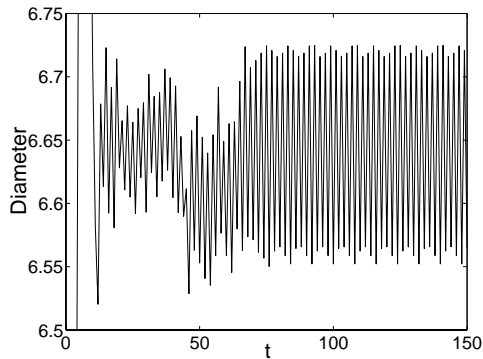


Fig. 2. An example of periodic oscillations of a scale-free network of 1000 nodes and of average degree $\langle k \rangle = 4$. To generate the oscillations, the node that handles the largest amount of load is identified and a ten-fold, sudden increase of load is applied to this node (*e.g.*, due to an attack). The diameter of the network in steady state is $D_0 \approx 5.4$. Shown are the persistent oscillations of the network diameter about a higher average value, indicating that the network is *never* able to return to its original steady state.

malfunctioned biological networks to be targeted and cured by drugs. This review article presents a theoretical approach for understanding the basic physics associated with the security of scale-free networks. In particular, we have presented a theoretical formula for estimating the critical phase-transition point with respect to the network capacity parameter, around and below which total disintegration of the network due to attack even on a single node is highly likely. We have also discussed what it takes for a scale-free network to be robust against global cascading breakdown as caused by attacks on a single node. Analyzing the dynamics of load redistribution as a result of selectively removing a small set of low-degree nodes yields a criterion that allows the lower bound of the capacity parameter for cascades-free scale-free networks and the optimal fraction of intentionally removed nodes to be determined.

An ongoing research project at NUS concerns situations in complex networks where overload does not necessarily lead to failures and an attack typically causes a large perturbation to the network. An example is the denial-of-service type of attacks on the internet, where the load of the attacked node suddenly increased to the extent that the excessive load cannot be handled in reasonable amount of time, leading to traffic congestion at the node. When this occurs, packets (or more generally, information flow) must find alternative paths in the network to reach their destinations in the fastest possible way. This can effectively change the fundamental quantities

characterizing the network, such as its diameter and betweenness centrality.⁵ The congested nodes, by design, try to process the excessive loads within their capacities by routing packets to other nodes in the network as fast as possible. Since packets are continuously generated, processed, and delivered on the network as required by its operation, routing of a large amount of excessive loads can cause other nodes in the network to be jammed. As a result, quantities such as the network diameter starts to oscillate in time and, the *network oscillates* in this sense. A remarkable recent finding at NUS is that the oscillation can be persistent in that, due to the attack, the network will never relax to its normal state prior to the attack. Depending on the specifics of the attack and the network state, the oscillation can be periodic or completely random. An example of the oscillation of a scale-free network is shown in Fig. 2. The striking feature is that the oscillations, periodic or random, are caused solely by the interplay between the network topology and the traffic flow protocol, regardless of the local node dynamics. In fact, no apparent node dynamics, linear or nonlinear, is assumed, except for the simple rule that it “holds” and causes the traffic to wait when overloaded. This may have wide implications. For instance, it can provide an alternative explanation for the recently observed chaotic oscillations in real internet traffic flow.¹² From the point of view of security, persistent oscillations the network away from its normal state could cause serious disruption to its function and therefore could be of significant concern.

References

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. R. Albert, H. Jeong, and A.-L. Barabási, *Nature* **406**, 378 (2000).
4. A. E. Motter and Y.-C. Lai, *Phys. Rev. E* **66**, 065102(R) (2002).
5. K.-I. Goh, B. Kahng, and D. Kim, *Phys. Rev. Lett.* **87**, 278701 (2001).
6. M. E. J. Newman, *Phys. Rev. E* **64**, 016132 (2001).
7. L. Zhao, K. Park, and Y.-C. Lai, *Phys. Rev. E* **70**, 035101(R) (2004).
8. L. Zhao, K. Park, Y.-C. Lai, and N. Ye, *Phys. Rev. E* **72**, 025104(R) (2005).
9. K.-I. Goh, C.-M. Ghim, B. Kahng, and D. Kim, *Phys. Rev. Lett.* **91**, 1898041 (2003).
10. K. Park, Y.-C. Lai, and N. Ye, *Phys. Rev. E* **70**, 026109 (2004).
11. A. E. Motter, *Phys. Rev. Lett.* **93**, 098701 (2004).
12. J.-B. Gao, N. S. V. Rao, J. Hu, and J. Ai, *Phys. Rev. Lett.* **94**, 198702 (2005).