Two-Peak and Three-Peak Optimal Complex Networks

André X. C. N. Valente,[1,*] Abhijit Sarkar,[2] and Howard A. Stone[1]
[1]*Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*
[2]*Department of Physics and Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 6 November 2003; published 19 March 2004)

A central issue in complex networks is tolerance of random failures and intentional attacks. Current literature emphasizes the dichotomy between networks with a power-law node connectivity distribution, which are robust to random failures but fragile to targeted attacks, versus networks with an exponentially decaying connectivity distribution, which are less tolerant to failures but more resilient to attacks. We prove analytically that the optimal network configuration under a classic measure of robustness is altogether different from both of the above: in all cases, failure and/or attack, there are no more than three distinct node connectivities in the optimal network.

DOI: 10.1103/PhysRevLett.92.118702 PACS numbers: 89.75.Fb, 64.60.Ak, 84.35.+i, 89.20.Hh

An outstanding issue in systems as diverse as power supply [1], transportation [2], communication [3,4], gene [5,6], metabolic [5], and ecological [7] networks is tolerance to component breakdown. Recent research on these and other complex networks has focused on abstracting from the details intrinsic to each of the systems and represents them in a unified way as a network of nodes connected by links [8–12].

Breakdown in a complex network is represented by the removal of nodes [13] and robustness refers to the ability of the surviving nodes to remain, as much as possible, interconnected. Two cases of fundamental interest are the random removal of nodes, which simulates random failures of individual elements, and the removal of the most connected nodes, which simulates a targeted attack aimed at crippling a network [4,9,14].

We report the results of an analysis of the problem of designing networks to be robust against random failures and deliberate attacks and obtain a constructive proof of the most robust network architecture. The network configurations we find are remarkably simple to describe qualitatively: They are characterized by the presence of at most three distinct node connectivities in the network.

A simple but essential measure that is used to capture the structure of a network is the node degree distribution: The degree, or connectivity, of a node is the number of links emanating from it, while the degree distribution gives the probability that a randomly chosen node has a given degree. Our analysis applies to the class of networks, known as generalized random graphs [15,16], which are random in every respect other than in their specified degree distribution. In generalized random graphs, potential node degree correlations in the network [17] are ignored. In particular, in the limit of a large number of nodes, the fraction of nodes forming loops of a given size goes to zero [15,18].

The robustness criterion we apply is the presence (or absence) of a giant connected component in the network, defined as a connected cluster of nodes whose size scales linearly with the network size N (the total number of nodes) [16]. This analysis pertains to the case where N tends to infinity. In particular, note that, given a degree distribution, N must be large enough so that any degree with nonzero probability in the distribution is present in a statistically significant number of nodes in the network. In a phase (or percolation [9]) transition, the giant connected component disappears under the removal of more than some critical fraction of the network nodes [9,15,16,19], with the size of the largest connected cluster then scaling only proportionally to $\log N$ [16]. Naturally, the percolation threshold depends on whether the nodes are removed randomly (random failure mode) or whether the most connected ones are chosen for removal first (attack mode): we denote the two distinct thresholds by f_r and f_a, respectively. This is a topological robustness criterion, as it does not take into account potential dynamical effects subsequent to the removal of a node. For an example of interplay between dynamics and robustness, see [20–22], and for other possible topological measures of robustness, see [9,14].

With the above robustness criterion in mind, we define the optimization problem as follows: Given a fixed average number of links per node, we determine the node degree distribution that maximizes the percolation threshold. An equivalent view of the problem is that of finding the degree distribution that minimizes the average number of links per node while still satisfying a minimal percolation threshold, specified *a priori,* as a desired robustness condition. We shall use these two perspectives interchangeably in the rest of the Letter. As a final constraint on the optimization problem, we require that the degree k of every node in the network satisfies $0 < k_\ell \leq k \leq k_m$, where the minimum k_ℓ and maximum k_m allowed degrees are considered given and fixed. We analyze three variations on the optimization problem: Optimization against only random failures, optimization against an intentional attack, and, finally, the most practically relevant case, optimization of a network against both random failures and intentional attacks.

We start with the case of optimization against random failures. Let $\langle k\rangle$ denote the average degree in the network and $\langle k^2\rangle$ denote the average squared degree. Then, the percolation threshold f_r for a network with a given degree distribution is given by [18,19]

$$f_r = 1 - \frac{1}{\frac{\langle k^2\rangle}{\langle k\rangle} - 1}. \tag{1}$$

Clearly, the optimal degree distribution, which maximizes f_r, maximizes $\langle k^2\rangle$ for a fixed $\langle k\rangle$. Now, the distribution that maximizes $\langle k^2\rangle$ for a fixed $\langle k\rangle$ is the one where all the nodes have either degree k_ℓ or degree k_m, the two extreme degrees. We conclude that the optimal degree distribution can be nonzero only at k_ℓ and k_m, with the choice for the exact partition of the distribution between these two degrees being determined by the value of $\langle k\rangle$. We call such a distribution with only two nonzero values a "two-peak distribution." Under this optimal two-peak distribution, f_r can be expressed as

$$f_r = 1 - \frac{\langle k\rangle}{-k_m k_\ell + \langle k\rangle(k_m + k_\ell - 1)}. \tag{2}$$

Figure 1 shows how two archetypical real networks fare by comparison: The electrical power grid of the western United States, which is a network with an exponentially decaying degree distribution, and the internet router network, which is a network with a power-law degree distribution.

We now analyze optimization against intentional attacks. In this case nodes are removed sequentially in descending order according to their degree, starting with the most connected node. Let f_a be the minimal desired percolation threshold against attack. Using the generating function formalism [15,19], we express in a useful form the condition for percolation when the fraction f_a of the most connected nodes is deleted [19,23]. Let us visualize an attack on a network as described in Fig. 2. For the purpose of determining the existence of a giant component, this view of the change in connectivity is entirely equivalent to that of deleting the attacked nodes and all links emanating from them. However, the generating functions for this network are now elementary. Let $G_0(x)$ be the generating function associated with the distribution of outgoing links of a node picked at random and let $G_1(x)$ be the generating function associated with the distribution of remaining outgoing links of a node arrived at by following a link emanating from a white node. Then, with p_k denoting the probability that a node picked at random has degree k and with q_k standing for the fraction of nodes of degree k that are white, we have

$$G_0(x) = \sum_{k=k_\ell}^{k_m} p_k(1 - q_k) + \sum_{k=k_\ell}^{k_m} p_k q_k x^k, \tag{3}$$

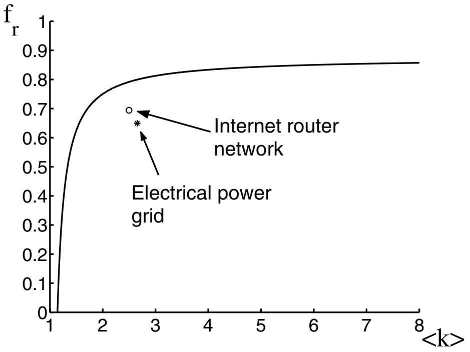

FIG. 1. Optimization against random failures. Plot of f_r the critical fraction of randomly deleted nodes where the giant connected component vanishes versus $\langle k\rangle$ the average number of links per node in the network, under an optimally chosen degree distribution—in the sense that it maximizes f_r for a given $\langle k\rangle$. A permissible minimum $k_\ell = 1$ and maximum $k_m = 8$ node degree constraint was also imposed. For $\langle k\rangle$ below 8/7 there is no percolation in the network even with no node failures. Maximum robustness occurs when all the nodes have k_m links, in which case the percolation transition occurs at $f_r = 6/7$. For comparison purposes f_r and $\langle k\rangle$ for two real networks are plotted: *, Western United States electrical power grid, an exponential network; ○, Internet router network, a power-law network. For the power grid $(k_\ell, k_m) = (1, 19)$ and for the internet $(k_\ell, k_m) = (1, 20)$ [24]. The values of f_r, $\langle k\rangle$, k_ℓ, and k_m for these real networks were computed from data in Refs. [3,25].

$$G_1(x) = \frac{\sum_{k=k_\ell}^{k_m} k p_k(1 - q_k) + \sum_{k=k_\ell}^{k_m} k p_k q_k x^{k-1}}{\sum_{k=k_\ell}^{k_m} k p_k}. \tag{4}$$

Thus, using standard methods [15,19], the percolation condition is

$$G_1'(1) \geq 1 \Rightarrow \frac{\sum_{k=k_\ell}^{k_m} k^2 p_k q_k - \sum_{k=k_\ell}^{k_m} k p_k q_k}{\sum_{k=k_\ell}^{k_m} k p_k} \geq 1. \tag{5}$$

Under attack, the function q_k takes the form

$$q_k = \begin{cases} 0, & \text{for } k > k^*, \\ q_{k^*}, & \text{for } k = k^*, \\ 1, & \text{for } k < k^*, \end{cases} \tag{6}$$

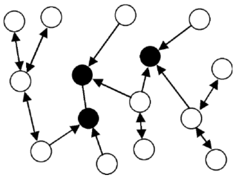

FIG. 2. Network under intentional attack. Functional nodes are white and nodes destroyed under attack are black. A link between white nodes can be traversed both ways. The nonfunctionality of a black node is characterized by having links to it become one-way incoming links only; i.e., such links work in the direction leading into the black node, but not going out of the black node (note: a link between two black nodes cannot be traversed either way).

where k^* is the largest degree a white node may have after the fraction f_a of the most connected nodes in the network has been darkened. Given this form of q_k, we express the percolation condition (5) as

$$\frac{\langle k^2\rangle_{k^*} - \langle k\rangle_{k^*}}{\langle k\rangle} \geq 1. \qquad (7)$$

This shorthand notation serves to highlight the fact that the numerator consists of the averages of k^2 and k associated with the original p_k distribution modified by transforming all nodes with degree above some k^* [including some fraction $(1 - q_{k^*})$ of the nodes of degree k^*] into nodes of degree zero.

We now argue that the optimal distribution against attack is also a two-peak distribution, namely, one where p_k can be nonzero only at k_ℓ and at k^*. Let us consider a graph with an arbitrary p_k degree distribution and with a fraction f_a of its nodes destroyed under attack. Moving all the probability in the $k > k^*$ region to the point k^* affects neither $\langle k^2\rangle_{k^*}$ nor $\langle k\rangle_{k^*}$, and yet it decreases $\langle k\rangle$. In other words, it widens the inequality (7) while decreasing the average number of links used. We conclude that, in the optimal degree distribution, $p_k = 0$ for $k > k^*$. Now let us consider the probability in the region $k < k^*$. By an argument analogous to the one used for the random failures case, $\langle k^2\rangle_{k^*}$ can be maximized while keeping $\langle k\rangle_{k^*}$ and $\langle k\rangle$ fixed by concentrating all the probability in the region $k < k^*$ at the extreme values k_ℓ and k^*. Therefore, we conclude that the optimal distribution against intentional attacks is also a two-peak distribution with the entire probability concentrated at k_ℓ and at k^*. The optimization problem is, therefore, reduced to finding the $k^* \in \{k_\ell, \ldots, k_m\}$ and $p_{k^*} \in [0, 1]$ values that minimize $\langle k\rangle$ subject to condition (7). This is further simplified by noting that, at each candidate k^*, only the p_{k^*} value that yields equality in condition (7) needs to be checked (excluding the trivial case in which the minimal distribution where all the nodes have degree k_ℓ suffices to satisfy the percolation requirement). The different k^* candidates are therefore trivially checked numerically. As an aside, note that, unlike in the random failures case, in the attack case the presence of a k_m maximum allowed degree is not strictly necessary, since the analysis will always yield a finite k^*.

Figure 3(a) shows f_a as a function of $\langle k\rangle$ for an optimal network. The values of f_a and $\langle k\rangle$ for the electrical power grid and internet router networks are provided for comparison. The characterization of the optimal network robust to a given f_a is completed by giving the values of k^* and p_{k^*}, shown in Figs. 3(b) and 3(c), respectively.

We now find the degree distribution for a network that, while minimizing the average number of links per node used, percolates under the loss of a fraction f_a of its nodes under attack and percolates under the loss of a fraction f_r of its nodes under random failure; we demand that the network satisfy the two percolation conditions separately, that is, the specification is not that the network percolates under a combined fractional loss of $f_a + f_r$. In this case we argue that the optimal degree distribution can be a three-peak distribution—its nodes have only one of three possible degrees: k_ℓ, k_m and an in-between degree k^*, where the previous definition of k^* continues to hold. Let us consider an arbitrary degree distribution p_k. Moving the entire probability in the region $k < k^*$ to the extremes k_ℓ and k^* while keeping $\langle k\rangle$ and $\langle k\rangle_{k^*}$ constant (same condition), has the effect of maximizing $\langle k^2\rangle$ and $\langle k\rangle_{k^*}$. This benefits both the robustness against attacks, Eq. (7), and the robustness against random failures, Eq. (1). We conclude that there are no nodes with a degree between k_ℓ and k^* in the optimal degree distribution. We now turn to the probability in the region $k > k^*$. As far as robustness against random failures, Eq. (1), moving all the probability to the extremes k^* and k_m while keeping $\langle k\rangle$ constant, maximizes robustness by maximizing $\langle k^2\rangle$. On the other hand, robustness against attack, Eq. (7), is not affected by the placement of the probability in the region $k > k^*$ as long as $\langle k\rangle$ is kept constant, since $\langle k^2\rangle_{k^*}$ and $\langle k\rangle_{k^*}$ are not affected. We conclude that there are no nodes with a degree between k^* and k_m in the optimal degree distribution. Therefore, we have shown that the optimal distribution can have at most three distinct degrees. The exact qualitative form of the distribution depends on the combination of f_r and f_a. The different possibilities are outlined in Fig. 4(a), using a color coding scheme. Figure 4(b) shows the $\langle k\rangle$ of the optimal network associated with each given (f_r, f_a) pair. We note that numerically only a limited number of potential solutions need to be checked to find the optimal architecture. This limited

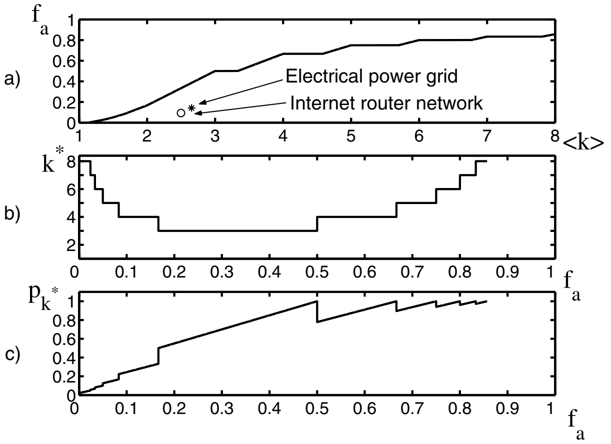

FIG. 3. Optimization against intentional attacks. (a) Analogous plot to Fig. 1, except now the nodes are deleted under attack mode, meaning the nodes are removed sequentially in descending order according to their degree, starting with the most connected node. Again, we take $(k_\ell, k_m) = (1, 8)$. *, Western United States electrical power grid; ○, Internet router network. (b) k^*, one of the two node degrees present (the other being k_ℓ) in the associated two-peak optimal degree distribution against attack. (c) p_{k^*}, the fraction of nodes with degree k^* in the optimal degree distribution. The discontinuity in the variables is a consequence of the discreteness of k^*.

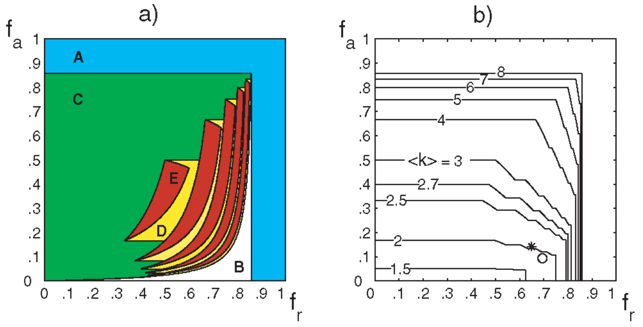

FIG. 4 (color online). Simultaneous optimization against intentional attacks and random failures. We take $(k_\ell, k_m) = (1, 8)$. To each combination of desired minimal network percolation thresholds, f_a under attack and f_r under random failures, corresponds an optimal network, i.e., one that also minimizes $\langle k \rangle$. (a) These optimal networks can be divided into different qualitative classes, illustrated using different colors: A—Robustness to these (f_a, f_r) pairs is not attainable due to the k_m constraint. B—f_r is the limiting constraint. There are two node degrees present in these networks, k_ℓ and k_m. C—f_a is the limiting constraint. There are at most two distinct node degrees in these networks, k_ℓ and k^*. D—Both f_a and f_r affect the optimal degree distribution. These networks still have just two distinct node degrees, k_ℓ and k^* (i.e., the potential third degree, k_m, turns out to have zero frequency). E—As in D, both f_a and f_r affect the optimal degree distribution but there are now three distinct node degrees in the network, k_ℓ, k^*, and k_m. (b) Contour plot of $\langle k \rangle$ for the optimal networks. The $\langle k \rangle = k_m = 8$ contour represents the maximum achievable robustness. For comparison, the (f_r, f_a) robustness thresholds of two real networks were plotted: *, Western United States power grid (exponential network); ○, Internet router (power-law network). For the power grid $\langle k \rangle = 2.7$ and for the internet $\langle k \rangle = 2.5$. Note how the points fall below the respective optimal $\langle k \rangle$ contours.

checking follows because, for the cases in which the solution is determined by a combination of f_a and f_r, it can be shown that if the optimal solution is a two-peak distribution, then equality holds in (7), while if the optimal solution is a three-peak distribution, equality (1) also holds in addition to (7). Therefore the potentially optimal values of p_{k^*} and p_{k_m} can always be expressed in terms of the candidate k^*.

In this Letter we have shown that the network configurations that maximize the percolation threshold under attack and/or random failures have at most three distinct node degrees. From a practical point of view, both engineered and naturally occurring networks have a diversity of factors influencing and constraining their ultimate configuration. Nonetheless, the optimal configurations we present provide a standard against which the robustness of real networks can be compared and act as an intuitive guide for network-robustness engineering.

A. S. deeply appreciates the guidance and support of Professor David Nelson and Professor X. Sunney Xie, as well as the financial support of NSF (DMR-0231631) and HMRL (DMR-0213805).

*Electronic address: andre@deas.harvard.edu

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440–442 (1998).
[2] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[4] R. Albert, H. Jeong, and A. Barabási, Nature (London) **406**, 378 (2000).
[5] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. Barabási, Nature (London) **407**, 651 (2000).
[6] J. Hasty, D. McMillen, and J. J. Collins, Nature (London) **420**, 224 (2002).
[7] R. V. Solé and J. M. Montoya, Proc. R. Soc. London, Ser. B **268**, 2039 (2001).
[8] S. Strogatz, Nature (London) **410**, 268 (2001).
[9] R. Albert and A. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[10] X. F. Wang, Int. J. Bifurcation Chaos Appl. Sci. Eng. **12**, 885 (2002).
[11] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[12] Z. N. Oltvai and A. Barabási, Science **298**, 763 (2002).
[13] Some problems are best modeled with the removal of links rather than nodes. The mathematical analysis for such cases parallels closely that for the case of node removal. See, for example, Ref. [26].
[14] B. Shargel, H. Sayama, I. R. Epstein, and Y. Bar-Yam, Phys. Rev. Lett. **90**, 068701 (2003).
[15] M. E. J. Newman, S. Strogatz, and D. Watts, Phys. Rev. E **64**, 026118 (2001).
[16] M. Molloy and B. Reed, Combinatorics Probab. Comput. **7**, 295 (1998).
[17] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[18] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[19] D. S. Callaway, M. E. J. Newman, S. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[20] D. J. Watts, Proc. Natl. Acad. Sci. U.S.A. **99**, 5766 (2002).
[21] A. E. Motter and Y. C. Lai, Phys. Rev. E **66**, 065102 (2002).
[22] Y. Moreno, R. Pastor-Satorras, A. Vásquez, and A. Vespignani, Europhys. Lett. **62**, 292 (2003).
[23] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).
[24] Although k_m for the power grid was 19, only 1.5% of the nodes had a degree above 8. Similarly, only 3% of the internet nodes had a degree above 8. Because of this, we considered that imposing a maximum of $k_m = 8$ in our optimal peaked networks would make the comparison fairer. Had we imposed $k_m = 19$ or 20, the robustness difference between our peaked networks and the internet and power grid networks would only increase.
[25] A. Barabási and R. Albert, Science **286**, 509 (1999).
[26] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).