Breakdown of the Internet under Intentional Attack

Reuven Cohen,[1],* Keren Erez,[1] Daniel ben-Avraham,[2] and Shlomo Havlin[1]

[1]*Minerva Center and Department of Physics, Bar-Ilan University, Ramat-Gan, Israel*
[2]*Department of Physics, Clarkson University, Potsdam, New York 13699-5820*

We study the tolerance of random networks to intentional attack, whereby a fraction p of the most connected sites is removed. We focus on scale-free networks, having connectivity distribution $P(k) \sim k^{-\alpha}$, and use percolation theory to study analytically and numerically the critical fraction p_c needed for the disintegration of the network, as well as the size of the largest connected cluster. We find that even networks with $\alpha \leq 3$, known to be resilient to random removal of sites, are sensitive to intentional attack. We also argue that, near criticality, the average distance between sites in the spanning (largest) cluster scales with its mass, M, as $\sqrt{M}$, rather than as $\log_k M$, as expected for random networks away from criticality.

The question of stability of scale-free random networks to removal of a fraction of their sites, especially in the context of the Internet, has recently been of interest [1–3]. The Internet can be viewed as a special case of a random, scale-free network, where the probability of a site to be connected to k other sites follows a power law: $P(k) \sim k^{-\alpha}$ ($\alpha \approx 2.5$, for the Internet). It is now well established that if a fraction p of the sites is removed *randomly,* then for $\alpha > 3$ there exists a critical threshold, p_c, such that for $p > p_c$ the network disintegrates; networks with $\alpha \leq 3$ are more resilient and do not undergo this transition, although finite networks (such as the Internet) may eventually be disrupted when nearly all of their sites are removed, as shown numerically in [1,2], and analytically in [2].

Albert *et al.* [1] have introduced a model for intentional attack, or sabotage of random networks: the removal of sites is not random, but rather sites with the highest connectivity are targeted first. Their numerical simulations suggest that scale-free networks are highly sensitive to this kind of attack. In this Letter we study the problem of intentional attack in scale-free networks. Our study focuses on the exact value of the critical fraction needed for disruption and the size of the remaining largest connected cluster. We also study the distance between sites on this cluster near the transition. We find, both analytically and numerically, that scale-free networks are highly sensitive to sabotage of a small fraction of the sites, for all values of α, lending support to the view of Albert *et al.* [1].

In a recent paper [2] we have studied the properties of the percolation phase transition in scale-free random networks, and applied a general criterion for the existence of a spanning cluster (a cluster whose size is proportional to the size of the network) [2,4]:

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = 2. \qquad (1)$$

Here k is the site connectivity, and averages, indicated by angular brackets, are taken over all sites of the network. When a fraction p of the sites are randomly removed (or

a fraction p of the links are removed, or lead to deleted sites), the distribution of site connectivity is changed from the original $P(k)$ to a new distribution $\tilde{P}(k)$:

$$\tilde{P}(k) = \sum_{k_0 \geq k}^{K} P(k_0) \binom{k_0}{k} (1 - p)^k p^{k_0 - k}. \qquad (2)$$

Using this criterion together with Eq. (1), the critical threshold $p = p_c$ is found to be

$$1 - p_c = \frac{1}{\kappa_0 - 1}, \qquad (3)$$

where $\kappa_0 = \langle k_0^2 \rangle / \langle k_0 \rangle$ is calculated from the original connectivity distribution, before the removal of any sites [2].

A wide range of networks, including the Internet, have site connectivities which follow a power-law distribution [1,5,6]:

$$P(k) = ck^{-\alpha}, \qquad k = m, m + 1, \ldots, K, \qquad (4)$$

where $k = m$ is the minimal connectivity and $k = K$ is an effective connectivity cutoff present in finite networks. For the distribution (4), κ_0 can be approximated by [6]

$$\kappa_0 = \left(\frac{2 - \alpha}{3 - \alpha} \right) \frac{K^{3-\alpha} - m^{3-\alpha}}{K^{2-\alpha} - m^{2-\alpha}}. \qquad (5)$$

This, together with Eq. (2), was used to show that networks with $\alpha \leq 3$, which have a divergent second moment, are resilient to random deletion of sites [2]. Indeed, when the number of sites in such networks $N \to \infty$, then the upper cutoff $K \to \infty$, and there exists a spanning cluster for all values of $p < 1$. Another approach, based on generating functions, was introduced in [7] and was used to study a similar problem in [3].

Consider now intentional attack, or sabotage [1], whereby a fraction p of the sites with the highest connectivity is removed. (The links emanating from the sites are removed as well.) This has the following effect: (a) the cutoff connectivity K reduces to some new value, $\tilde{K} < K$, and (b) the connectivity distribution of the remaining sites

is no longer scale-free, but is changed, because of the removal of many of their links. The upper cutoff K before the attack may be estimated from

$$\sum_{k=K}^{\infty} P(k) = \frac{1}{N}, \quad (6)$$

where N is the total number of sites in the network. Similarly, the new cutoff $\tilde{K}$, after the attack, can be estimated from

$$\sum_{k=\tilde{K}}^{K} P(k) = \sum_{k=\tilde{K}}^{\infty} P(k) - \frac{1}{N} = p. \quad (7)$$

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 [6]:

$$\tilde{K} = m p^{1/(1-\alpha)}. \quad (8)$$

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 connectivity 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:

$$\tilde{p} = \sum_{k=\tilde{K}}^{K} \frac{k P(k)}{\langle k_0 \rangle}, \quad (9)$$

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

$$\tilde{p} = \left(\frac{\tilde{K}}{m}\right)^{2-\alpha} = p^{(2-\alpha)/(1-\alpha)}, \quad (10)$$

for $\alpha > 2$. For $\alpha = 2$, $\tilde{p} \to 1$, since just a few nodes of very high connectivity control the entire connectedness of the system. Indeed, consider a finite system of N sites and $\alpha = 2$. The upper cutoff $K \approx N$ must then be taken into account, and approximating Eq. (9) by an integral yields $\tilde{p} = \ln(Np/m)$. That is, for $\alpha = 2$, very small values of p are needed to destroy an arbitrarily large fraction of the links as $N \to \infty$.

With these results known 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 connectivity sites reduces the upper cutoff. Since this effect changes the connectivity distribution, κ_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. (2), the result in Eq. (3)

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 from that 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. (3) and (5), but with $\tilde{p} = (\tilde{K}/m)^{2-\alpha}$ and $\tilde{K}$ replacing p_c and K. This yields the equation

$$(\tilde{K}/m)^{2-\alpha} - 2 = \frac{2-\alpha}{3-\alpha} m[(\tilde{K}/m)^{3-\alpha} - 1], \quad (11)$$

which can be solved numerically to obtain $\tilde{K}(m, \alpha)$, and then $p_c(m, \alpha)$ can be retrieved from Eq. (8). In Fig. 1 we plot p_c—the critical fraction of sites needed to be removed in the sabotage strategy to disrupt the network—computed in this fashion, and compared to results from numerical simulations. A phase transition exists (at a finite p_c) for all $\alpha > 2$. The decline in p_c for large α is explained from the fact that as α increases the spanning cluster becomes smaller in size, even before attack. (Furthermore, for $m < 2$ the original network is disconnected for some large enough α.) The decline in p_c as $\alpha \to 2$ results from the critically high connectivity of just a few sites: their removal disrupts the whole network. This was already argued in [1]. We note that for infinite systems $p_c \to 0$ as $\alpha \to 2$. The critical fraction p_c is rather sensitive to the lower connectivity cutoff m. For larger m (the case of $m = 1$ is shown in Fig. 1) the networks are more robust, though they still undergo a transition at a finite p_c.

The size of the spanning cluster as a fraction of the number of undeleted sites, $P_\infty(p)$, can be calculated using the methods introduced in [8] and developed in [3,7]. Following closely the derivation in [7], a generating function is
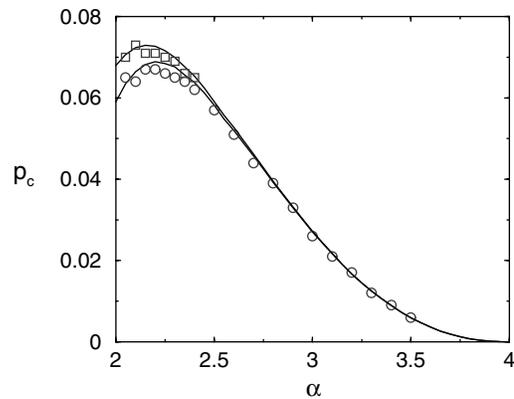

FIG. 1. Critical probability, p_c, as a function of α, for networks of size $N = 500\,000$ (circles) and $N = 64\,000$ (squares). Lines represent the analytical solution, obtained from Eqs. (7) and (9).

built for the connectivity distribution:

$$G_0(x) = \sum_{k=0}^{\infty} P(k)x^k. \tag{12}$$

The probability of reaching a site with connectivity k by following a specific link is $kP(k)/\langle k \rangle$ [2–4,7], and the corresponding generating function is

$$G_1(x) = \frac{\sum kP(k)x^{k-1}}{\sum kP(k)} = \frac{d}{dx}G_0(x)/\langle k \rangle. \tag{13}$$

Hence, the generating function for the probability of reaching a branch of a given size by following a link is

$$H_1(x) = xG_1\big(H_1(x)\big), \tag{14}$$

while the generating function for the size of a component is

$$H_0(x) = xG_0\big(H_1(x)\big). \tag{15}$$

Then, $P_\infty = 1 - H_0(1)$, since H_0 contains only the finite-size clusters. It follows that

$$P_\infty(p) = 1 - \sum_{k=0}^{\infty} \tilde{P}(k)u^k, \tag{16}$$

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

$$\langle k \rangle u = \sum_{k=0}^{\infty} kP(k)u^{k-1}. \tag{17}$$

We argue that the same holds true after attack, but the sums in Eqs. (16) and (17) should run up to only $k = \tilde{K}$, and the original distribution $P(k)$ should be replaced by the new connectivity distribution [2]:

$$\tilde{P}(k) = \sum_{k_0=m}^{\tilde{K}} P(k_0)\binom{k_0}{k}(1 - \tilde{p})^k \tilde{p}^{k_0-k}. \tag{18}$$

The actual fraction of removed sites, p, is inconsequential since we calculate the size of the infinite cluster relative to the number of *undeleted* sites. P_∞ evaluated in this fashion, and compared to numerical simulations, is shown in Fig. 2.

Since intentional attack leads always to a finite cutoff which does not scale with the system size, all the moments are finite and Eqs. (16) and (17) are well behaved. Therefore, there always exists a linear term in the series expansion of $P_\infty(p)$. Hence, near the critical point $P_\infty \sim |p_c - p|^\beta$, where $\beta = 1$ [9].

Finally, we consider the distance between sites in the spanning cluster. The behavior of this quantity in diluted networks is different from that at the highly connected regime. The average distance between two sites in the spanning cluster of a highly connected network is proportional to $\log_k N$, where k is the average connectivity and N is the number of sites [10]. This has also been shown to hold for scale-free and general networks [7]. 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
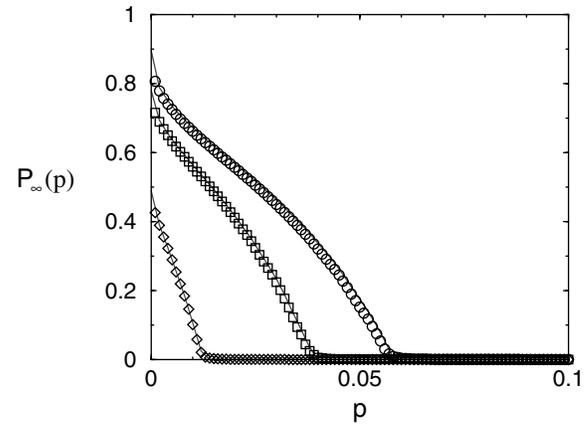

FIG. 2. Fraction of sites belonging to the spanning cluster, P_∞, as a function of the fraction of removed sites, p, for networks with $\alpha = 2.5$ (circles), $\alpha = 2.8$ (squares), and $\alpha = 3.3$ (diamonds). Lines represent the analytical result from Eqs. (16) and (17). Both the simulation and analysis are for system size $N = 500\,000$.

from infinite-dimensional percolation theory that the fractal dimension at criticality is $d_f = 2$ [11]. Therefore the average (chemical) distance d between pairs of sites on the spanning cluster at criticality behaves as

$$d \sim \sqrt{M}, \tag{19}$$

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_f 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 $\xi \approx |p_c - p|$, as can be seen in Fig. 3.
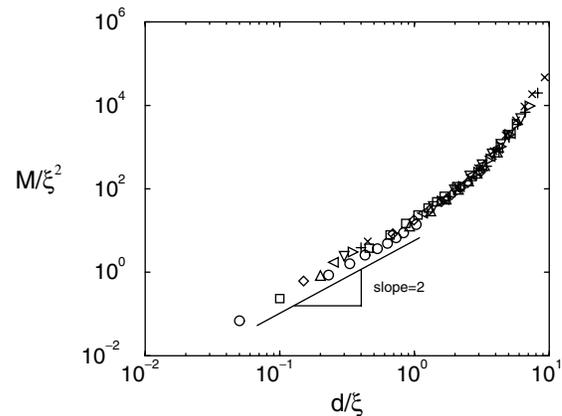

FIG. 3. Mass (number of sites), M, as a function of distance, d, on the spanning cluster. The correlation length is $\xi = |p - p_c|^{-1}$. Note that, for $d/\xi < 1$, the slope is 2, corresponding to the behavior in the critical regime, while, for $d/\xi > 1$, M grows exponentially with d, corresponding to the well connected regime.

In summary, we have shown that scale-free networks are highly sensitive to intentional attack. This is true even for networks with $\alpha < 3$, which are known to be resilient to random removal of their sites. The high sensitivity near $\alpha = 2$ results from the presence of just a few sites with connectivity comparable to the size of the system: their removal disrupts the whole network. We note that while the cutoff $\tilde{K}$ must reach a typically small number before the network is disrupted, this is achieved with a modest removal fraction p [Eq. (8)]. The effect of sabotage on the connectivity distribution of the remaining sites after the attack and thus the relation between $\tilde{p}$, $\tilde{K}$, and p [Eqs. (7) and (9)] is found explicitly in our approach. This effect is particularly important near the borderline case of $\alpha = 2$. We have also shown that the average distance between pairs of sites in the spanning cluster grows dramatically near criticality. This makes communication very inefficient, even before the spanning cluster is completely disrupted.

*Email address: cohenr@shoshi.ph.biu.ac.il

[1] R. Albert, H. Jeong, and A. L. Barabási, Nature (London) **406**, 6794 (2000); **406**, 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] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).

[5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, ACM SIGCOMM '99 Comput. Commun. Rev. **29**, 251 (1999).

[6] In practice, in our simulations we generate a *continuous* variable k, with distribution $P(k) = ck^{-\alpha}$ ($k \geq m$), and pick up the closest integer as the actual connectivity. Thus, we really have $P(k) = \int_{k-1/2}^{k+1/2} ck'^{-\alpha} dk'$. This has the effect of rendering the continuum approximation, used in our analysis, more accurate.

[7] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, cond-mat/0007235.

[8] M. Molloy and B. Reed, Comb. Probab. Comput. **7**, 295 (1998).

[9] However, for random breakdown when the upper cutoff $K \to \infty$, there are diverging moments. Using Abelian theories for calculating the sums in Eqs. (16) and (17), it can be shown that for $3 < \alpha < 4$, $\beta = 1/(\alpha - 3)$ while for $\alpha > 4$, $\beta = 1$ as in regular infinite-dimensional percolation.

[10] B. Bollobás, *Random Graphs* (Academic Press, London, 1985), pp. 123–136.

[11] *Fractals and Disordered System,* edited by A. Bunde and S. Havlin (Springer, New York, 1996).