Attack Strategies on Complex Networks

Lazaros K. Gallos[1], Reuven Cohen[2], Fredrik Liljeros[3], Panos Argyrakis[1], Armin Bunde[4], and Shlomo Havlin[5]

[1] Department of Physics, University of Thessaloniki, 54124 Thessaloniki, Greece
gallos@physics.auth.gr
[2] Department of Physics, Bar-Ilan University, 52900 Ramat-Gan, Israel
[3] Department of Sociology, Stockholm University 106 91 Stockholm, Sweden
[4] Institut für Theoretische Physik III, Justus-Liebig-Universität Giessen
Heinrich-Buff-Ring 16, 35392 Giessen, Germany

Abstract. In this work, we estimate the resilience of scale-free networks on a number of different attack methods. We study a number of different cases, where we assume that a small amount of knowledge on the network structure is available, or can be approximately estimated. We also present a class of real-life networks that prove to be very resilient on intentional attacks, or equivalently much more difficult to immunize completely than most model scale-free networks.

1 Introduction

A large number of diverse systems in society, nature and technology can be described by the concept of a network [1, 2]. In a network the form of inter-relations between the system parts determines many structural and dynamic properties of the system. One such property that has received considerable attention is the robustness of a network under failures[3, 4] or intentional attack [3, 5, 6]. Equivalently, from a sociological point of view, the robustness of a network can be related to an immunization process, where immunized nodes no longer transmit a disease, and thus the destruction of a spanning cluster in the network means that the population is immune to a disease, which will soon die out because it will encounter non-susceptible nodes. In such cases we are mainly interested in using the lowest possible number of vaccinations, either for reasons of increased cost or unavailability of a large number of vaccines. These strategies strongly influence the form of the resulting network, which in turn affects important dynamic properties, such as delivery time in the Internet, delays in information or virus spreading, etc [7].

In the course of an intentional attack nodes of the network are removed in decreasing order of their degree (number of connections to other nodes). This is considered to be the most harmful type of attack on a network, since the removal of the hubs results in the largest possible damage. This removal process has many and important implications, since depending on the application, it may describe the resilience of a network, such as the Internet, or the required number of vaccinations for immunization considerations, etc. For a scale-free network, where the probability that a node has a given number of links decays

V.N. Alexandrov et al. (Eds.): ICCS 2006, Part III, LNCS 3993, pp. 1048–1055, 2006.

as a power-law $P(k) \sim k^{-\gamma}$, it has been shown that the critical percentage f_c of removed nodes that results in network desintegration is very low (less than $f_c = 0.07$) [5, 6]. It is, thus, a well-established fact, supported by analytic results and simulations on model and real-life networks, that a scale-free network is very vulnerable to intentional attacks (where f_c is close to 0), although the same network is extremely robust under random node failures (where $f_c \simeq 1$) [4].

2 Attacks with Limited Network Knowledge

In many cases, it is possible that the robustness of a node depends on its connectivity, i.e. the probability of damaging a node either by failure or by an external attack depends on the degree k of the node. We simulate this situation by using the probability $W(k) \sim k^{\alpha}$ for a node with degree k to become inactive. The parameter α can be regarded as a measure of our knowledge on the network structure. When $\alpha < 0$ nodes with low degree are more vulnerable, while for $\alpha > 0$ high-degree nodes are removed with higher probability than the low degree nodes. The cases $\alpha = 0$ and $\alpha \to \infty$ represent the known random removal and targeted intentional attack, respectively.

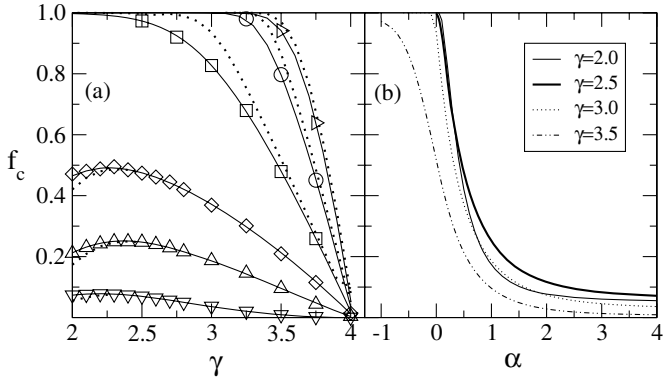

Fig. 1. (a) Values of f_c vs γ for different α values: (bottom to top) $\alpha = 4, 1, 0.5, 0, -0.5, -1$. Symbols represent simulation data ($N = 10^6$ nodes) from $100-300$ different network realizations. Solid lines are the theoretical predictions for finite-size networks, while dashed lines correspond to infinite-size networks. (b) Values of f_c vs α for infinite-size networks and different γ values.

Results of simulations on model networks for the critical fraction f_c of the nodes that need to be removed before the destruction of the spanning cluster are presented in Fig. 1. A theoretical result for this problem has been presented in [8], which can be seen in the figure to be in excellent agreement with the simulations. For $\gamma < 3$, f_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 $\alpha = \gamma - 1$, f_c saturates, which means that the knowledge available to the attacker in this case is sufficient to destroy the network most efficiently.

Our results show that little knowledge on the highly connected nodes in an intentional attack reduces the threshold drastically compared to the random case. Thus, a large network can be damaged efficiently even when only a small fraction of hubs is known to the attacker. For immunization of populations this means that if we are able to identify (and immunize) even with small probability the virus spreaders we can significantly reduce the spreading threshold.

3 Acquaintance Immunization with Limited Knowledge

Another method of attacking the system, with application in immunization strategies, is the acquaintance immunization method [9]. According to this scheme a random node is selected and then points to a random acquaintance along one of its links. The node at the other end of the link is the one to be immunized (removed). This method achieves great efficiency in lowering the percolation threshold. Here, we consider the probability of further improving the efficiency of this process, provided that we use partial knowledge on the network structure.

We select a percentage p of a network comprising N nodes and ask them to direct us to a random acquaintance of theirs, which will then be immunized. The initial selection of the pN nodes is based on partial information on their connectivity, so that a node i with k_i links has a probability $W(k_i) \sim k_i^{\alpha}$ of being approached (the parameter α has the same meaning as in the previous section). We increase the value of p up to a value p_c where the fraction f_c of actually immunized nodes results to the arrest of the epidemic (or equivalently to the desctruction of the spanning cluster).

We present a theoretical analysis of the problem, based on the arguments presented in Ref. [9]. We consider a network where a fraction p of its nodes have been selected and have pointed to a fraction f of unique nodes that have been immunized. We assume that in the network there are no degree-degree correlations, so that the probability of a node with k links to be connected to a node with k' links is independent of k, $\phi(k') \equiv p(k'|k) = k'P(k')/\langle k \rangle$.

For our derivation we assume that loops can be neglected and the nodes are located on layers, l, from a chosen origin. We denote the number of nodes with degree k in the layer l by $n_l(k)$. The event of a node with degree k being susceptible (not immunized) is denoted by s_k. As a starting point for the calculation of $n_l(k)$ we use Eq. (1) from Ref. [9]

$$n_{l+1}(k') = \sum_{k=1}^{k_{\max}} n_l(k)(k-1)p(k'|k,s_k)p(s_{k'}|k',k,s_k)\,. \tag{1}$$

The upper value $k_{\max}$ is taken equal to the natural cutoff $k_{\max} = N^{1/(\gamma-1)}$, while $p(k'|k,s_k)$ denotes the probability of reaching a node of degree k' by following a

link from a susceptible node with degree k and $p(s_{k'}|k', k, s_k)$ is the probability that this k'-degree node is also susceptible.

A random node of degree k is selected with probability $k^\alpha/(N\langle k^\alpha\rangle)$ where $\langle k^\alpha\rangle = \sum k^\alpha P(k)$. In order to find the probability for a random acquaintance to be immunized we divide by $1/k$, so that the probability of immunizing this specific acquaintance is $k^{\alpha-1}/(N\langle k^\alpha\rangle)$. The probability that this node is not selected is $1 - k^{\alpha-1}/(N\langle k^\alpha\rangle)$ and after Np immunization attempts it becomes

$$\nu_p(k) = \left(1 - \frac{k^{\alpha-1}}{N\langle k^\alpha\rangle}\right)^{Np} \sim \exp\left(-\frac{k^{\alpha-1}}{\langle k^\alpha\rangle}p\right) . \tag{2}$$

Since the network is uncorrelated we consider the average value $\nu_p = \langle\nu_p(k)\rangle = \sum_k \nu_p(k)\phi(k)$, so that the probability that a node with degree k is susceptible is, in general, $p(s_k|k) = \nu_p^k$. If the degree of one neighbor is known to be k' this probability becomes $p(s_k|k, k') = \nu_p^{k-1} \exp\left(-\frac{k'^{\,\alpha-1}}{\langle k^\alpha\rangle}p\right)$. Since immunization of a node is independent of the probability that its neighbor is also immunized we also have $p(s_k|k, k') = p(s_k|k, k', s_{k'})$. Combining the above results with the Bayes rule and Eq. (2) we finally get the expression:

$$n_{l+1}(k') = n_l(k') \sum_{k=1}^{k_{\max}} \phi(k)\nu_p^{k-2}(k-1) \exp\left(-\frac{2k^{\alpha-1}}{\langle k^\alpha\rangle}p\right) . \tag{3}$$

When the sum in the above expression is greater than 1 then the number of susceptible nodes increases with increasing layer index l. When the sum is less than 1 the percolation phase dissapears. At the critical concentration p_c this sum is, thus, equal to 1, i.e.

$$\sum_{k=1}^{k_{\max}} \phi(k)\nu_{p_c}^{k-2}(k-1) \exp\left(-\frac{2k^{\alpha-1}}{\langle k^\alpha\rangle}p\right) = 1 . \tag{4}$$

Next, we obtain the critical value p_c by numerically solving this equation and we compute the critical fraction of immunized nodes, i.e. the fraction of not susceptible nodes:

$$f_c = 1 - \sum_{k=1}^{k_{\max}} P(k)p(s_k|k) = 1 - \sum_{k=1}^{k_{\max}} P(k)\nu_{p_c}^k . \tag{5}$$

The numerical solution of Eq. (5) for f_c as a function of α for networks with different γ exponents is compared in Fig. 2 with simulation data.

For networks with $\gamma < 3$ the critical threshold is minimized at values of $\alpha \simeq 1$, which is the optimum value for the presented strategy, while for $\gamma \geq 3$ the threshold presents a different behavior and decreases monotonically with increasing α values. In practice, the process at $\alpha = 1$ is equivalent to selecting a random link and immunizing one of the two nodes attached to the given link (provided the uncorrelated network hypothesis holds). It is also interesting to notice that

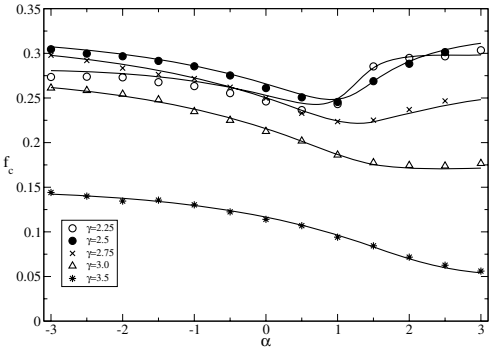

Fig. 2. Critical immunized fraction f_c of the population as a function of α for random scale-free networks with different γ exponents (shown in the plot). Network size is $N = 10^5$ nodes. Symbols are the results of simulations and lines represent the analytic solution (Eqs. 4 and 5).

up to the value $\alpha = 1$ the acquaintance immunization strategy is superior to direct immunization of the initially selected nodes, but close to this value the two methods yield a similar value for f_c. When $\alpha > 1$ the direct immunization method becomes more efficient than the acquaintance immunization strategy.

4 Robust Real-Life Scale-Free Networks

In this section we show that there exists a large class of networks, that are usually found in nature and society and have already been characterized as scale-free, but nevertheless remain robust against removal of the most connected nodes. We first present the results for real-life networks and then introduce a modified version of scale-free networks, for which our analytic and simulation treatment support these findings.

To demonstrate this issue we performed intentional attacks and random nodes removal to many different real-life networks. Although many of these systems behave in a similar way to the model network (where f_c is usually less than 10%) there is a number of networks, such as actors collaboration and science citations, where the intentional attack requires removal of a considerable portion of the network nodes, which is of the order of 65%. In order to outline the common feature of these networks, in Fig. 3 we present the degree distribution of these networks. These distributions have a flat or rising part at low-degree nodes and only after a threshold value the distribution decays as a power-law.

We use a general model for simulating similar networks. We consider networks whose degree distribution is uniform up to a threshold value k_c and for larger values decays as a power law $k^{-\gamma}$. The exact form of the distribution (plotted also for $k_c = 50$ and $\gamma = 2.5$ in Fig. 3) is

$$P(k) = \begin{cases} \frac{\gamma-1}{\gamma} k_c^{-1} & 1 < k < k_c \\ \frac{\gamma-1}{\gamma} k_c^{\gamma-1} k^{-\gamma} & k > k_c \end{cases} . \quad (6)$$

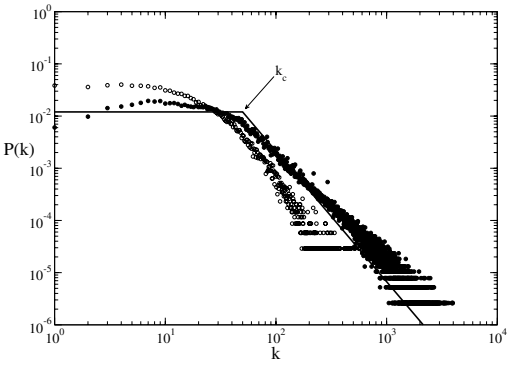

Fig. 3. Degree distributions for IMDB actors (filled symbols) and HEP citations (open symbols). The solid line represents a typical degree distribution (Eq. 6) that we used as a model.

We calculate the critical threshold f_c for such a network based on ideas introduced by Cohen et al [5] and Dorogovtsev and Mendes [11]. Nodes are removed according to their initial degree. An intentional attack results in the disruption of the network. We consider that the degrees of the nodes for the resulting network are given by the parameter $\tilde{k}$, with corresponding averages

$$\langle \tilde{k} \rangle = \int_1^{\tilde{K}} kP(k)dk \ , \ \langle \tilde{k}^2 \rangle = \int_1^{\tilde{K}} k^2 P(k)dk \, . \tag{7}$$

The effect of an intentional attack is to remove all nodes of a network whose degree is larger than a cutoff value $\tilde{K}$, i.e. $\tilde{k} \in [1, \tilde{K}]$. This also implies that f_c equals

$$f_c = 1 - \int_{\tilde{K}}^{\infty} P(k)dk \, . \tag{8}$$

At the same time, removal of a node leads to removing all its links to other nodes. We consider random networks with no correlations in the nodes connections, which means that a removal of a node results in removal of random links with probability

$$\tilde{p} = \frac{\int_{\tilde{K}}^{\infty} kP(k)dk}{\int_1^{\infty} kP(k)dk} = 1 - \frac{\langle \tilde{k} \rangle}{\langle k \rangle} \, . \tag{9}$$

It has been shown [4, 10] that a random network loses its large-scale connectivity after the removal of a critical fraction f_c of nodes, which behaves as

$$f_c = 1 - \frac{1}{\kappa - 1} \tag{10}$$

where $\kappa \equiv \langle k^2 \rangle / \langle k \rangle$ as usual. We use the above equation for the network resulting after the attack, i.e. we substitute f_c with $\tilde{p}$ from Eq. 9 and $\kappa = \langle \tilde{k}^2 \rangle / \langle \tilde{k} \rangle$. After a few trivial steps Eq. 10 becomes

$$\langle \tilde{k}^2 \rangle - \langle \tilde{k} \rangle = \langle k \rangle \, . \tag{11}$$

This formula, which is exact, has been already proved in Refs. [5, 11].

In order to use Eq. 11 we need to know whether the value of $\tilde{K}$ is larger or smaller than the threshold value of the distribution k_c. We have considered each case separately, but when $\tilde{K} > k_c$ there is no solution to the problem (which is also verified by our simulations where always $\tilde{K} < k_c$). Calculation of the involved integrals yields

$$\langle \tilde{k} \rangle \simeq \frac{\gamma - 1}{2\gamma} \frac{\tilde{K}^2}{k_c} \,, \tag{12}$$

and

$$\langle \tilde{k}^2 \rangle \simeq \frac{\gamma - 1}{3\gamma} \frac{\tilde{K}^3}{k_c} \,. \tag{13}$$

The average value of the initial degree distribution $P(k)$ (Eq. 6) can be approximated with the assumption that $k_{\max} = \infty$. However, for low γ values this assumption does not work well and we can compute the integral up to the maximum value $k_{\max} = K$, which can be computed from the relation $\int_{k_{\max}}^{\infty} P(k) = 1/N$, and is given in our case by $K = k_c N^{1/(\gamma-1)} \gamma^{1/(1-\gamma)}$. This results in a correction $x = 2N^{(2-\gamma)/(\gamma-1)} \gamma^{1/(1-\gamma)}$ to the average value of the distribution, which finally becomes

$$\langle k \rangle = \frac{(\gamma - 1) k_c}{2(\gamma - 2)} (1 - x) \,. \tag{14}$$

Combining Eqs. 11-14 we get

$$2\tilde{K}^3 - 3\tilde{K}^2 = \frac{3\gamma k_c^2}{\gamma - 2} (1 - x) \,. \tag{15}$$

We can now compute the value of $\tilde{K}$ from Eq. 15 and substitute it to Eq. 8, which can also be written as

$$f_c \simeq 1 - \frac{\gamma - 1}{\gamma} \frac{\tilde{K}}{k_c} \,. \tag{16}$$

The numerical solution of Eqs. 15 and 16 is shown in Fig. 4 as a function of γ for different values of the threshold value k_c. In the same figure we also plot results of simulations on model networks.

Comparison of the curves in Fig. 4 to the intentional attack on regular scale-free networks shows a dramatic increase in the value of f_c, over the entire γ range. Increase of the threshold value k_c enhances the stability of the network. For $k_c = 10$ the critical fraction is already above 40%, while when $k_c = 100$ the value of f_c lies in the range 70-80%. This, of course, is also a consequence of the increasing mean degree of nodes as we increase k_c, that makes all nodes in the system better connected.

These findings provide a structure that is very robust against both random failures and targeted attacks. This optimization is desirable in most cases. Such a structure, which we have seen in many instances emerges naturally, may be used to efficiently protect a network against most attacks.

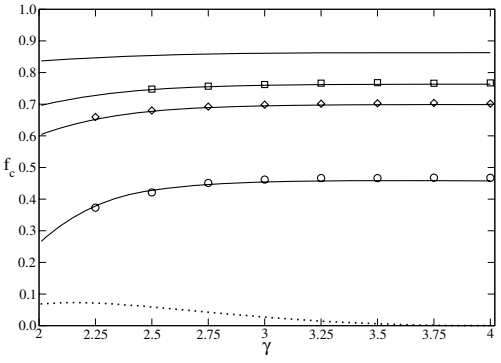

Fig. 4. Critical fraction f_c of removed nodes for networks that undergo an intentional attack, as a function of the exponent γ. From top to bottom: $k_c = 500$, 100, 50, and 10. Solid lines represent the numerical solution of Eqs. 15 and 16, while symbols are simulation results. The dashed curve corresponds to pure scale-free networks (Ref. [5]).

Acknowledgement

This work was supported by a European research NEST/PATHFINDER project DYSONET 012911.

References

1. R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).
2. S.N. Dorogovtsev and J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002).
3. R. Albert, H. Jeong, and A.L. Barabási, Nature (London) **406**, 378 (2000).
4. R. Cohen et al., Phys. Rev. Lett. **85**, 4626 (2000).
5. R. Cohen et al., Phys. Rev. Lett. **86**, 3682 (2001).
6. D.S. Callaway et al., Phys. Rev. Lett. **85**, 5468 (2000).
7. D.J. Watts, Proc. Nat. Ac. Sci. **99**, 5766 (2002).
8. L.K. Gallos et al., Phys. Rev. Lett. **94**, 188701 (2005).
9. R. Cohen, S. Havlin, and D. ben-Avraham, Phys. Rev. Lett. **91**, 247901 (2003).
10. G. Paul, S. Sreenivasan, and H.E. Stanley, preprint arxiv:cond-mat/0507202 (2005).
11. S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. Lett. **87**, 219801 (2001).