Vulnerability of complex networks under intentional attack with incomplete information

IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **40** (2007) 2665–2671 doi:10.1088/1751-8113/40/11/005

Vulnerability of complex networks under intentional attack with incomplete information

J Wu, H Z Deng, Y J Tan and D Z Zhu

Department of Management Science and Engineering, College of Information Systems and Management, National University of Defense Technology, Changsha 410073, People's Republic of China

E-mail: wujunpla@163.com

Received 30 October 2006, in final form 18 January 2007
Published 28 February 2007
Online at stacks.iop.org/JPhysA/40/2665

Abstract

We study the vulnerability of complex networks under intentional attack with incomplete information, which means that one can only preferentially attack the most important nodes among a local region of a network. The known random failure and the intentional attack are two extreme cases of our study. Using the generating function method, we derive the exact value of the critical removal fraction f_c of nodes for the disintegration of networks and the size of the giant component. To validate our model and method, we perform simulations of intentional attack with incomplete information in scale-free networks. We show that the attack information has an important effect on the vulnerability of scale-free networks. We also demonstrate that hiding a fraction of the nodes information is a cost-efficient strategy for enhancing the robustness of complex networks.

PACS numbers: 89.75.Hc, 89.75.Fb, 05.70.Jk

1. Introduction

Networks with complex topology describe a wide range of systems in nature and society. Examples include the Internet [1], the World Wide Web [2], metabolic networks [3], electric power grids [4] and many others. In the past few years, the discovery of small-world [5] and scale-free properties [6] has stimulated a great deal of interest in studying the underlying organizing principles of various complex networks [7, 8].

Because of its broad application, the vulnerability of complex networks, i.e., how random failures or intentional attacks affect the integrity and operation of the networks, has received growing attention, especially from the original work by Albert *et al* [9]. Albert *et al* have introduced models for random failure and intentional attack and suggested that scale-free networks display an exceptional robustness against random failure, but show poor performance against intentional attack.

For the study of attack vulnerability of complex networks, the selection procedure of the order in which nodes are removed is an open choice. One may of course maximize the destructive effect at any fixed number of removed nodes or edges. However, this requires the information of the whole network structure. If we can obtain the complete information of the network structure, we preferentially remove the most important nodes among the whole network according to some criterion (the most common criteria are the degree of node). This attack strategy corresponds to the intentional attack with complete information. If we have zero information of the network structure, we can only remove nodes randomly. This attack strategy corresponds to the random failure. The intentional attack with complete information and random failure are just two extremes in real-world networks. The more cases are between these two extremes, i.e. an intentional attack with incomplete information. The vulnerability of complex networks under random failure or intentional attacks with complete information has been well established [9–23]. However, only a few studies have focused on the attacks with incomplete information. Dezső [24] *et al* investigated the effect of incomplete information on the epidemic threshold for viruses spreading on scale-free networks and it is assumed that the likelihood of identifying an infected node with k edges depends on the node's degree as k^{α}. Gallos *et al* [25, 26] studied the tolerance of SF networks under systematic variation of the attack strategies, in which the probability that a given node is destroyed, depends on its degrees as $W(k) \sim k^{\alpha}$. Following the method in [25], Dall'Asta *et al* [27] studied the inhomogeneous percolation under systematic variation of the attack strategies. In fact, the incomplete information considered in [24–27] is uncertain information, which means that one can obtain the information of all nodes, but the information may be uncertain. There is another scenario for incomplete information, on which this paper focuses, i.e. one can obtain the information of partial nodes, whereas the information is certain. Using simulation method, Xiao *et al* [28] examined the robustness of complex communication networks under intentional attacks, in which two special cases of nodes information missing are considered, but no analytical result is achieved.

In this paper, we study vulnerability of complex networks under intentional attacks with incomplete information by introducing a general model for intentional attacks with a tunable attack information parameter. Our study focuses on the exact value of the critical removal fraction f_c of nodes for the disintegration of networks and the size of the giant component under intentional attack with incomplete information.

2. Model of intentional attacks with incomplete information

A complex network can be represented by a graph G with N nodes and M edges. Assume that G is an undirected and simple connected graph with uncorrelated degree distribution. Let d_i be the degree of a node v_i.

An intentional attack with incomplete information means that we can only preferentially remove the most important nodes among a local region of a network. It consists of two steps:

(1) Choose the effective attack region (EAR): select $N\alpha$ nodes randomly, where the parameter α is the measure of attack information.

(2) Attack Nf nodes in G: if $\alpha > f$, we remove Nf nodes in decreasing order of degree in EAR; if $\alpha \leqslant f$, we first remove all nodes in EAR and then remove $N(f-\alpha)$ nodes randomly besides EAR.

It is obvious that there are two extreme cases: $\alpha = 0$ and $\alpha = 1$, corresponding to random failure and intentional attack with complete information, respectively.

Let $p(k)$ $(m \leqslant k \leqslant K)$ be the degree distribution that a randomly chosen node in the network has degree k, where m is the smallest degree and K is the upper cutoff. Let $q(k)$ be the probability distribution that a node is not removed given that it has degree k.

In the case $\alpha \leqslant f$, since EAR is obtained by selecting nodes randomly, it is easy to obtain that an intentional attack with incomplete information is identical to a random failure, and then we obtain

$$q(k) = 1 - f. \tag{1}$$

In the case $\alpha > f$, let $\tilde{K}$ be the maximum degree of the remaining nodes in EAR after nodes removal. Obviously, $\tilde{K}$ is a function of α and f denoted by $\tilde{K}(\alpha, f)$. Using $\tilde{K}$, $q(k)$ can be written as

$$q(k) = \begin{cases} 1 & k \leqslant \tilde{K} \\ 1 - \alpha & k > \tilde{K}. \end{cases} \tag{2}$$

Now we derive $\tilde{K}$. Sorting all $N\alpha$ nodes in EAR in decreasing order of degree, let $R(k)$ be the rank of a node with degree k in EAR, then we obtain

$$R(k) = N\alpha \int_k^K p(t)\,\mathrm{d}t. \tag{3}$$

Note that $R(\tilde{K}) = Nf$, then

$$R(\tilde{K}) = N\alpha \int_{\tilde{K}}^K p(t)\,\mathrm{d}t = Nf. \tag{4}$$

Solving equation (4), we can obtain $\tilde{K}(\alpha, f)$. In particular, for scale-free networks with power law degree distributions $p(k) = Ck^{-\lambda}$ $(m \leqslant k \leqslant K)$, where $C \approx (\lambda - 1)m^{\lambda-1}$ and $K \approx mN^{1/(\lambda-1)}$ [12], we can obtain $\tilde{K}(\alpha, f)$ in scale-free networks as follows:

$$\tilde{K}(\alpha, f) = m\left(f + \frac{1}{N}\right)^{\frac{1}{1-\lambda}} \approx mf^{\frac{1}{1-\lambda}}. \tag{5}$$

3. Critical removal fraction and size of giant component

Now we employ the generating function formalism [11, 29] to find exact analytic solutions for the critical removal fraction f_c and the size of giant component under intentional attack with incomplete information.

The generating function of $p(k)$ $(m \leqslant k \leqslant K)$ is that

$$G_0(x) = \sum_{k=m}^{K} p(k)x^k. \tag{6}$$

Another quantity that will be important to us is the distribution of the degree of the nodes that we arrive at by choosing a random edge and following it to one of its ends. The correctly normalized distribution $r(k)$ of the remaining degree is then given by [30]

$$r(k) = \frac{(k+1)p(k+1)}{\sum_k kp(k)} = \frac{(k+1)p(k+1)}{\langle k \rangle} \tag{7}$$

where $\langle k \rangle$ is the average degree in the network. Thus the generating function of the remaining degree distribution is generated by the function

$$G_1(x) = \sum_{k=m}^{K} r(k)x^{k-1} = \frac{G_0'(x)}{\langle k \rangle}. \tag{8}$$

Let $w_0(k)$ be the probability that a randomly chosen node has degree k and is not removed, i.e. $w_0(k) = p(k)q(k)$. Let $w_1(k)$ be the probability that a node at the end of a randomly chosen edge has degree k and is not removed, i.e. $w_1(k) = r(k)q(k)$. The generating function of $w_0(k)$ and $w_1(k)$ is respectively

$$F_0(x) = \sum_{k=0}^{\infty} w_0(k)x^k = \sum_{k=m}^{K} p(k)q(k)x^k \tag{9}$$

$$F_1(x) = \sum_{k=0}^{\infty} w_1(k)x^k = \sum_{k=m}^{K} r(k)q(k)x^{k-1} = \frac{F_0'(x)}{\langle k\rangle}. \tag{10}$$

Following closely the derivation in [11, 29], the generating function for the distribution of the size of component by following a randomly chosen edge is

$$H_1(x) = 1 - F_1(1) + xF_1[H_1(x)]. \tag{11}$$

The probability distribution for the size of component to which a randomly chosen node belongs is similarly generated $H_0(x)$, where

$$H_0(x) = 1 - F_0(1) + xF_0[H_1(x)]. \tag{12}$$

Although it is not usually possible to find a closed-form expression for the complete distribution of component size in a network, we can derive a closed-form expression for the average component size $\langle s\rangle$ and the relative size of giant component S (the fraction of nodes in the giant component) from equation (11) and equation (12) as follows,

$$\langle s\rangle = H_0'(1) = F_0(1) + \frac{F_0'(1)F_1(1)}{1 - F_1'(1)} \tag{13}$$

$$S = F_0(1) - F_0(u) \tag{14}$$

where u is the smallest non-negative real solution of $u = 1 - F_1(1) + F_1(u)$. Equation (13) tells us that the phase transition at which a giant component forms takes place at $F_1'(1) = 1$, i.e.

$$F_1'(1) = \frac{\sum_k k(k-1)p(k)q(k)}{\sum_k kp(k)} = 1. \tag{15}$$

Submitting $q(k) = 1 - f$ into equation (15), we obtain

$$(1-f)\sum_{k=m}^{K} k(k-1)p(k) = \sum_{k=m}^{K} kp(k). \tag{16}$$

Using equation (16), we can obtain the critical removal fraction f_c under a random failure

$$f_c^{\text{random}} = 1 - \frac{1}{\kappa - 1} \tag{17}$$

where $\kappa = \langle k^2\rangle/\langle k\rangle$. In particular, for scale-free networks with power law degree distributions $p(k) = (\lambda - 1)m^{\lambda-1}k^{-\lambda}$ ($m \leqslant k \leqslant K$), with continuous approximation for $p(k)$, we can obtain

$$\kappa = \left(\frac{2-\lambda}{3-\lambda}\right)\frac{K^{3-\lambda} - m^{3-\lambda}}{K^{2-\lambda} - m^{2-\lambda}} \tag{18}$$

where $K \approx mN^{1/(\lambda-1)}$.

In the case $\alpha \leqslant f_c^{\text{random}}$, one must attack a fraction $f \geqslant \alpha$ of nodes to disintegrate the network, and then an intentional attack with incomplete information is equivalent to a random failure. Thus, we can obtain $f_c = f_c^{\text{random}}$.

In the case $\alpha > f_c^{\text{random}}$, submitting equation (2) into equation (15), we obtain

$$\sum_{k=m}^{\tilde{K}} k(k-1)p(k) + (1-\alpha)\sum_{k=\tilde{K}+1}^{K} k(k-1)p(k) = \sum_{k=m}^{K} kp(k). \tag{19}$$

Solving equation (19) for $\tilde{K}$, we can obtain the critical value $\tilde{K}_c$ and then the critical removal fraction f_c can be derived from $\tilde{K}(\alpha, f)$. In cases where equation (19) is not exactly solvable, we can evaluate $\tilde{K}_c$ by numerical iteration starting from a suitable initial value. In particular, for scale-free networks with power law degree distributions $p(k) = (\lambda-1)m^{\lambda-1}k^{-\lambda}$ $(m \leqslant k \leqslant K)$, with continuous approximation for $p(k)$, equation (19) takes the form

$$\frac{\alpha\tilde{K}^{2-\lambda} - 2m^{2-\lambda} + (2-\alpha)K^{2-\lambda}}{2-\lambda} = \frac{\alpha\tilde{K}^{3-\lambda} - m^{3-\lambda} + (1-\alpha)K^{3-\lambda}}{3-\lambda} \tag{20}$$

where $K \approx mN^{1/(\lambda-1)}$. In the extreme case $\alpha = 1$, i.e. the intentional attack with complete information, equation (20) is in agreement with the result in [13].

From equation (20) and equation (5), we can see that for all $\lambda > 3$ there exists a phase transition at a finite $f_c < 1$. However, for $2 < \lambda < 3$ and $\alpha < 1$, we can obtain $K^{3-\lambda} \to \infty$ as $N \to \infty$ and so $f_c \to 1$ as $N \to \infty$. In other words, for $2 < \lambda < 3$, a giant component exists for arbitrarily fractions of removal ($f < 1$) as $N \to \infty$ if we can hide a fraction of the node degrees. For $\alpha = 1$, i.e. the intentional attack with complete information, there exists a phase transition at a finite $f_c < 1$ for all $\lambda > 2$. This was already argued in [13].

4. Simulation results

To validate our model and method, we perform simulations of intentional attack with incomplete information in scale-free networks. We generate scale-free networks with degree distributions $p(k) = Ck^{-\lambda}$ using the method described in [31]. We then attack the network according to the model described in section 2 with different α and f. We choose $\kappa \equiv \langle k^2\rangle/\langle k\rangle < 2$ as the criterion for the disintegration of networks [12, 31]. After each node is removed, we calculate κ. When κ becomes less than 2, we record the number of nodes t removed up to that point. This process is performed for ten realizations with a specified degree distribution and, for each network, for ten different realizations of the selection of EAR. The threshold f_c is defined as $f_c = \langle t\rangle/N$.

To study the effect of attack information on the vulnerability of complex networks, we plot the size of giant component S as a function of f for different α shown in figure 1. It is clear that the attack information has an important effect on S. In the case $\alpha = 0.6$, even if 35% of the nodes are removed, there still exists a giant component containing most of nodes. However, if we can obtain the attack information with $\alpha = 0.85$, after 10% of nodes are removed, the network is fragmented. Figure 2 shows the critical removal fraction f_c from our simulations, along with the exact solution. The agreement between the two is good. We can find that the increase of attack information reduces the robustness of networks. In other words, we can enhance the robustness of networks by hiding the information of networks. For example, for the case of the Internet ($\lambda = 2.5$), if we hide 15% ($\alpha = 0.85$) of nodes randomly, the critical removal fraction f_c can increase from 0.05 to 0.978. It means that it is a cost-efficient strategy for enhancing the robustness of complex networks to hide the information of networks. In addition, we can observe the sudden drops of the critical removal fraction at different values

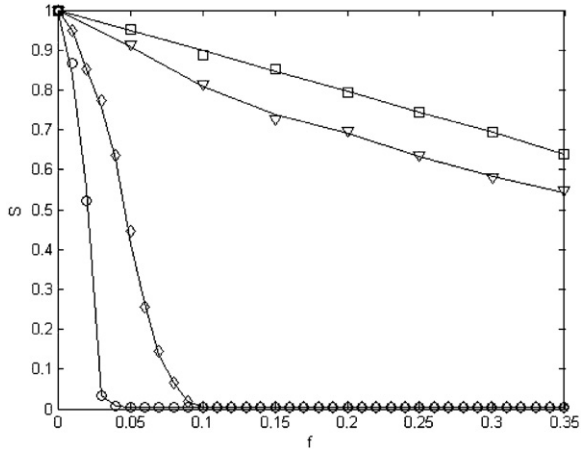

Figure 1. Size of the giant component S versus f for $\alpha = 1$ (circles), $\alpha = 0.85$ (diamonds), $\alpha = 0.6$ (triangles), $\alpha = 0$ (squares), where $N = 10^6$, $\lambda = 3$, $m = 1$. The solid lines represent the analytical results.

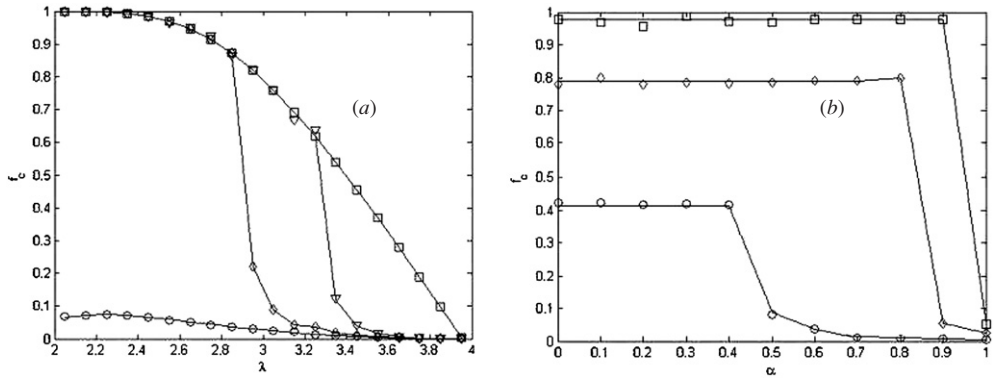

Figure 2. (*a*) Critical removal fraction f_c versus λ for $\alpha = 1$ (circles), $\alpha = 0.85$ (diamonds), $\alpha = 0.6$ (triangles), $\alpha = 0$ (squares), where $N = 10^6$, $m = 1$. (*b*) Critical removal fraction f_c versus α for $\lambda = 3.5$ (circles), $\lambda = 3$ (diamonds), $\lambda = 2.5$ (squares), where $N = 10^6$, $m = 1$. The solid lines represent the analytical results.

of α and λ in figure 2. It can be explained from the fact that if we cannot obtain adequate information $\left(\alpha \leqslant f_c^{\text{random}}\right)$, the critical removal fraction under incomplete information is just equal to the case of random failure.

5. Conclusions

We have introduced a model of intentional attack with incomplete information. The known random failure and the intentional attack with complete information are two extreme cases of our model. Using the generating function method, we have derived the exact value of the critical removal fraction f_c of nodes for the disintegration of networks and the size of the giant component S. Our analytical results allow us to make predictions on the vulnerability of complex networks under intentional attack with incomplete information.

We have shown that the attack information parameter α has an important effect on the critical removal fraction f_c of scale-free networks. The increase of attack information can strongly reduce the robustness of networks. Hiding just a small fraction of nodes can prevent the network to breakdown under intentional attack to the hubs. Remarkably, for $2 < \lambda < 3$, a giant component exists for arbitrarily fractions of removal as $N \to \infty$ if we can hide a fraction of the node degrees. It is a surprising result, since random hiding a fraction of nodes in a scale-free network should correspond to hide preferentially low-degree nodes, and then one would expect that an intentional attack should still damage the hubs of the network. It can be explained that for scale-free network with inhomogeneous degree distribution, there is few highly connected hubs which dominate a networks, so even hiding few hubs, which corresponds to hiding a small fraction of nodes randomly, can protect the whole network.

Acknowledgment

This work was supported by the National Science Foundation of China under grant no. 70501032.

References

[1] Vázquez A, Pastor-Satorras R and Vespignani A 2002 *Phys. Rev.* E **65** 066130
[2] Adamic L A and Huberman B A 2000 *Science* **287** 2115
[3] Jeong H, Tombor B, Albert R, Oltvai Z N and Barabási A-L 2000 *Nature* **407** 651–4
[4] Albert R, Albert I and Nakarado G L 2004 *Phys. Rev.* E **69** 025103
[5] Watts D J and Strogatz S H 1998 *Nature* **393** 440–2
[6] Barabási A-L and Albert R 1999 *Science* **286** 509–12
[7] Albert R and Barabási A-L 2002 *Rev. Mod. Phys.* **74** 47–51
[8] Newman M E J 2003 *SIAM Rev.* **45** 167–256
[9] Albert R, Jeong H and Barabási A-L 2000 *Nature* **406** 378–82
[10] Bollobás B and Riordan O 2003 *Int. Math.* **1** 1–35
[11] Callaway D S, Newman M E J, Strogatz S H and Watts D J 2000 *Phys. Rev. Lett.* **85** 5468–71
[12] Cohen R, Erez K, ben-Avraham D and Havlin S 2000 *Phys. Rev. Lett.* **85** 4626–8
[13] Cohen R, Erez K and ben-Avraham D 2001 *Phys. Rev. Lett.* **86** 3682–5
[14] Crucitti P, Latora V, Marchiori M and Rapisarda A 2003 *Physica* A **320** 622–42
[15] Crucitti P, Latora V, Marchiori M and Rapisarda A 2004 *Physica* A **340** 388–94
[16] Holme P, Kim B J, Yoon C N and Han S K 2002 *Phys. Rev.* E **65** 056109
[17] Lai Y C 2005 *Pramana J. Phys.* **64** 483–502
[18] Paul G, Sreenivasan S and Stanley H E 2005 *Phys. Rev. E* **72** 056130
[19] Paul G, Tanizawa T, Havlin S and Stanley H E 2004 *Eur. Phys. J.* B **38** 187–91
[20] Tanizawa T, Paul G, Cohen R, Havlin S and Stanley H E 2005 *Phys. Rev.* E **71** 047101
[21] Valente A X C N, Sarkar A and Stone H A 2004 *Phys. Rev. Lett.* **92** 118702
[22] Vázquez A and Moreno Y 2003 *Phys. Rev.* E **67** 015101
[23] Wang B, Tang H W, Guo C H and Xiu Z L 2005 *Physica* A **363** 591–6
[24] Dezső Z and Barabási A-L 2002 *Phys. Rev.* E **65** 055103
[25] Gallos L K, Argyrakis P, Bunde A, Cohen R and Havlin S 2004 *Physica* A **344** 504–9
[26] Gallos L K, Cohen R, Argyrakis P, Bunde A and Havlin S 2005 *Phys. Rev. Lett.* **94** 188701
[27] Dall'Asta L 2005 *J. Stat. Mech.* P08011
[28] Xiao S and Xiao G 2006 *Preprint* cs.NI/0609077
[29] Newman M E J, Strogatz S H and Watts D J 2001 *Phys. Rev.* E **64** 26118
[30] Newman M E J 2002 *Phys. Rev. Lett.* **89** 208701
[31] Molloy M and Reed B 1995 *Random Struct. Algorithms* **6** 161–79