Some scale-free networks could be robust under selective node attacks

Bojin Zheng¹,² (a), Dan Huang¹, Deyi Li³ (b), Guisheng Chen²,³ and Wenfei Lan¹

¹ College of Computer Science, South-Central University for Nationalities - Wuhan 430074, China
² State key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications - Beijing 100876, China
³ School of Software, Tsinghua University - Beijing 100084, China

received 3 December 2010; accepted in final form 21 March 2011; published online 18 April 2011

PACS 89.75.Fb – Structures and organization in complex systems
PACS 89.75.Da – Systems obeying scaling laws
PACS 89.20.Hh – World Wide Web, Internet

Abstract – It is a mainstream idea that scale-free network would be fragile under the selective attacks. Internet is a typical scale-free network in the real world, but it never collapses under the selective attacks of computer viruses and hackers. This phenomenon is different from the deduction of the idea above because this idea assumes the same cost to delete an arbitrary node. Hence this paper discusses the behaviors of the scale-free network under the selective node attack with different cost. Through the experiments on five complex networks, we show that the scale-free network is possibly robust under the selective node attacks; furthermore, the more compact the network is, and the larger the average degree is, then the more robust the network is; with the same average degrees, the more compact the network is, the more robust the network is. This result would enrich the theory of the invulnerability of the network, and can be used to build robust social, technological and biological networks, and also has the potential to find the target of drugs.

Copyright © EPLA, 2011

The invulnerability of complex networks is an important issue for various real-world networks [1–3], such as the social relationship network [4], Internet [5], the World Wide Web [6], the neural networks [7] and so on. The knowledge on this topic may lead to build more robust social, biological and technological networks, to find the targets of drugs or to destroy the terrorist groups etc. Moreover, this knowledge may lead to a deeper understanding of the stability of complex systems.

The pioneer scientists Reka Albert, Hawoong Jeong and Albert-László Barabási proposed that the complex networks with scale-free property would be robust with random errors, but fragile under selective attacks, that is, the robustness and the fragility coexist [1]. This conclusion can be explained as follows, since the scale-free networks have a mass of nodes with less connectivity and a small quantity of nodes with huge connectivity, the nodes with less connectivity would be chosen with a high probability when the nodes are randomly chosen to be removed, so that the network would lose less edges and show large invulnerability; on the other hand, if the nodes are preferentially chosen, the nodes with large connectivity would be firstly removed, so that the network would be very fragile to collapse.

The following researches proved the idea of Albert et al., for examples, the studies on the protein network [8], on WWW [9], on the food chain network [10], on the E-mail network [11], on Internet [12] and so on. These researchers made the idea as a common knowledge of the scientists in this field. Holme et al. [2] summarized that the selective attacks can be categorized as node attacks and edge attacks, and for each class, there are four strategies, that is, ID (Initial Degree), RD (Recalculated Degree), IB (Initial Betweenness), RB (Recalculated Betweenness), meaning that based on the initial graphs or current graphs, one orders the degree or the betweenness of the nodes/edges descendingly, and chooses the best node to remove one by one. Holme et al. pointed out that scale-free networks are fragile under all the attack strategies.

However, some scale-free networks provide exceptional excuses. For example, Internet is regarded as a typical scale-free network, but it has a surprising stability and rarely collapses. There are at least two reasons to support

(a)E-mail: zhengbojin@gmail.com
(b)E-mail: leedeyi@tsinghua.edu.cn
its strong robustness: first, this network survives under the selective attacks of the hackers. No hacker has achieved the boastful victory although they have tried some very important nodes; second, if one treats the attacks of the hackers as random attack, although we do not agree on this treatment, the spread of computer viruses would provide strong persuasion. As concerns the nodes with large degree, they will be infected with large probability because of their numerous neighborhoods. From the point of view of statistics, this is exactly the selective attacks. Another example is the military networks. Due to the hierarchy of military groups, scale-free networks can be constructed [13,14]. According to Albert et al.’s idea, the military networks would be very fragile and the strategy “to catch bandits first catch the ringleader” would be the best. But, if this idea holds, according to the evolutionary theory, the military groups would evolve into other forms to confront the selective/military attacks; on the other hand, the strategies would not be so colorful since only one strategy is enough.

Albert et al.’s idea assumed that the probability to remove any node or edge would be the same, or say, the cost is the same. Surprisingly, few works explored the world out of this hypothesis. Actually, no matter in Internet, or in military groups, the cost to remove a node would be certainly different. Moreover, the works mainly employ the degree attacks and the betweenness attacks. Considering that “the important nodes should be important to the whole network”, the attack strategies would be based upon the definition of “the importance of nodes”. Therefore, the closeness [15] or the other definition such as the definition in PageRank [16,17] and HITS [18] also can be used as an attack factor.

Thus, when the removal cost is different for the nodes with different importance, what will happen? We can image that a mass of manpower and materials will be used to enhance the protection of the important nodes so that the important nodes could resist the attacks, thereby leading to an increase in the robustness of the whole network. Therefore, when considering the cost, the scale-free network may be robust. In this letter, we try to explore the invulnerability of complex networks with the consideration of the different costs to delete the nodes and find that some complex networks could be very robust under selective attacks.

For the quantitative validation of the robustness of scale-free networks under selective node attacks with the consideration of the cost, we define the removal cost (attack cost), the attack strategies, the attack effect based on the previous related works, and we find that the compactness and the average degree of scale-free networks are related to the robustness, furthermore, the more compact the scale-free network is, and the larger the average degree is, then the more robust the network is; for the same average degree, the more compact the scale-free network is, the more robust the network is. Five scale-free networks are used to validate our idea.

Here we firstly define the relevant concepts. Since we only consider the selective node attacks, we define the attack as the removal of a set of nodes. Since the degree measure is the most natural measure to the property of nodes, we use a function of the degree of the set of nodes as the measure of the cost of an attack. That is, for given graph $G=(V, \text{Edge})$ and the node set for the removal of nodes $Z$, 
\begin{equation}
\text{Cost}(Z) = \sum_{v \in Z} f(\text{Degree}(v)).
\end{equation}

Here, $f(x)$ can be defined as various functions, for example, $f(x)=x^2$. Under this circumstance, since the attack cost of the nodes with higher degree would be remarkably large, the best strategy would be to remove the nodes with smaller degrees firstly. If $f(x)=1/x$, since the attack cost of important nodes would be smaller, it would be efficient to remove the nodes with large degree. If $f(x)=x^0$, that is, $f(x)=1$, the results would be the same as Barabási’s. The fairest formation would be $f(x)=x$, here the cost is equivalent to the degree, it also is the simplest formation. This letter suggests this function.

For any given network, there exists an upper limitation when all the edges are removed. We normalize the cost and have (denoting the normalized cost as $C(Z)$),
\begin{equation}
C(Z) = \frac{\text{Cost}(Z)}{\text{Cost}(V)}.
\end{equation}

As to the attack strategies, due to the fact that the important nodes would be important to the whole network, we can design the attack strategies according to the definitions of the node importance. Considering the definitions in Social Network [15], we focus on three strategies, IB (Initial Betweenness), ID (Initial Degree) and IC (Initial Closeness). Focusing on the selective node attacks, the important nodes would be preferentially removed.

As to the attack effects, the researchers have proposed various measures. Commonly, it can be measured by the difference of the performance of the tested network before or after the attack, such as average geodesic length, average inverse geodesic length and the size of the giant component [2]. We use the size of the giant component after the attack $\tilde{S}(Z)$ as the measure indicator, and considering universal and convenient comparisons, we use the normalized size $E(Z)$, where,
\begin{equation}
E(Z) = \frac{\tilde{S}(Z)}{\|V\|}.
\end{equation}

Considering that there exist many kinds of attack strategies, a network is said to be robust unless it can resist multiple kinds of attack strategies. If considering the degree and the closeness only, one assumes that the degree and the closeness are negatively correlated, then, the important nodes (with larger closeness) can be removed with smaller cost (equivalent to the degrees), so that the network is fragile. Therefore, the compactness of the network, or, that every node importance be positively
Some scale-free networks could be robust under selective node attacks.

Furthermore, considering that the networks with larger average degree would have high probability to escape from collapsing, that is, they may not collapse after the removal of one or multiple most important nodes, we will investigate the relationship between the robustness and the average degree.

We use five networks to explore their behaviors under selective node attacks, whose topographies are drawn in fig. 1. Two of them are generated by the CSF algorithm [2], one is compact and shown as fig. 1(a), the other is non-compact and shown in fig. 1(b). Three widely used real-world networks, the protein network [9] shown as fig. 1(c), the political books network (Polbooks1) shown in fig. 1(d) and the scientific collaboration network [19] shown in fig. 1(e), are chosen.

Since this letter only focuses on scale-free networks, we need to validate the scale-free property of the selected networks. The common method is to check whether the degree distribution of the networks will satisfy \( P(k) \sim k^{-\gamma} \), where \( \gamma \) is a constant and \( k \) is a random variant on the degree; when plotting the equation in the log-log coordination, the curve looks like a straight line. We plot their degree distributions in fig. 2. From this figure, we can see that the curves approximate the straight line in the log-log coordination. Considering that this method somehow is risky, we also checked the scale-free property of all the selected networks by the programme written by Clauset et al. [20].

To measure the correlated coefficients between the degree and the betweenness or the closeness, we use Pearson’s correlation coefficient, which belongs to \([-1, 1]\) where −1 indicates the minimal negative correlation, and 1 indicates the maximal positive correlation. The coefficients and some statistic values of the selected networks are listed in table 1.

We choose IB, ID, IC as the attack strategies to carry out the experiments. For any given attack, that is, the set \( Z \) of nodes to be removed, we plot the corresponding normalized cost \( C(Z) \) and the corresponding normalized attack effect \( E(Z) \) under these strategies as fig. 3. To avoid the distortion of the curves, the normalized size of the set \( Z \) is used. Just as in the previous works, \( E(Z) \) drops down drastically for all these networks, meaning that the networks are very fragile. But we can also notice that \( C(Z) \) increases drastically simultaneously. Therefore, when the

---

1This network was edited by Valdis Krebs and can be downloaded from MEJ Newman’s website.
cost is taken into consideration, we cannot conclude that
the networks are fragile.

Since we cannot determine the robustness only by
$C(Z)$ or $E(Z)$, respectively, we plot the corresponding
relationship between $C(Z)$ and $E(Z)$ in fig. 4 for any
given attack $Z$. To explore the relative robustness, we
use the complete networks as the baseline. Any selective
node attack would not make this kind of networks collapse
unless all the nodes are removed. Therefore, the complete
networks could be regarded as the most robust without
the consideration of the cost. But from the viewpoint of
the cost, they may waste much resource on trivial nodes.

To explore the relationship between the robustness and
the compactness or the average degree, we categorize
the networks into two groups, of which the first
group has larger average degree (that is, CSF-Compact,
CSF-NonCompact and Polbooks) and the second with
smaller average degree (i.e., Protein and Science).

As to the first group, fig. 4 shows that all three networks
have good resistance to ID attacks in contrast to the
complete networks, because the majority of points are
located above the baseline, meaning that ID attacks on
these networks would pay a larger percentage of cost but
achieve less effects. From fig. 4(a), we can see that every
curve of the CSF-Compact network does not differ much
from the others and the majority of their parts locates
above the baseline, meaning that this network can resist
all three attack strategies, but for figs. 4(b) and (c), the
curves differ much, especially for 4(b), most points of
the IB curve locate below the baseline, meaning that the
CSF-NonCompact network and the Polbooks network
cannot perform well for IB attacks, this phenomenon must
be attributed to their smaller $\delta_h$ so that the important
nodes can be removed with small cost. As to the second
group, fig. 4 shows that both networks do not perform well
to all the attacks in contrast to the complete networks,
since the curves are totally below the baseline. But owing
to the compactness, the Protein network still performs
better than the Science network.

Two check points are chosen to quantitatively demon-
strate the attack effects on these networks. The first check
point is $E(Z) = 80\%$. Table 2 lists the attack cost under
ID, IB and IC for five networks.

Table 2: $C(Z)$ when $E(Z) = 80\%$

| Network       | ID  | IB  | IC  |
|---------------|-----|-----|-----|
| CSF-Compact   | 71.25% | 70.77% | 66.79% |
| Polbook       | 63.04% | 48.30% | 45.35% |
| CSF-NonCompact| 68.78% | 24.52% | 53.53% |
| Protein       | 19.85% | 19.77% | 24.63% |
| Science       | 18.22% | 5.51%  | 7.68%  |
IC is also small. But for the CSF-NonCompact network, the gap between IB and ID achieves 44.26%; as to the Polbooks network, the gap also achieves 14.74%, meaning that both networks have community structures and cannot resist all attack strategies. As to the Protein network and the Science network, because less average edges lead to fast collapses, less cost can achieve better attack effects.

The second check point is chosen at $E(Z) = 30\%$. Table 3 lists $C(Z)$ under ID, IB and IC strategies.

Table 3 shows that the CSF-Compact network still is very robust, achieving 30% of the size of the giant component it would spend over 90% of the cost for all the three strategies. As to the other networks, less cost means that great collapse has emerged.

The experimental results support the proposed ideas. We can further rationalize the ideas in a simple way. Here we assume that the size of the giant component is still used to evaluate the attack effects and the cost is considered.

The process is demonstrated with a simple language as follows.

**Proposition 1:** The RD strategy is the worst attack strategy when the target networks do not collapse.

**Proof:** Assume that $Z|_{RD}$ is a set to represent an RD attack, and $E(Z|_{RD}) = |N| - |Z|_{RD}$, that is, the network does not collapse. For any given $Z'$ and $|Z'| = |Z|_{RD}$, because $E(Z') \leq |N| - |Z'| = E(Z|_{RD})$, and $C(Z') \leq C(Z|_{RD})$, i.e., the RD attack pays more cost but obtains worst effects. So the RD strategy is the worst attack strategy.

According to proposition 1, if we can construct a network that does not collapse under the RD strategy, then this network would be very robust under RD selective attack.

**Proposition 2:** Any connected network ($|N| \leq 3$) would have at least a collapse under all RD attacks except the complete networks.

**Proof:** Considering the connected network with $|N| = 3$, when it is not complete, it would collapse under the RD attack. Since the connected network with $|N| = 3$ is the most foundational case, it is easy to conclude that only the complete networks would never collapse under the RD attacks.

The complete networks would never collapse, but most networks are not complete. In the real world, it would be a massive resource waste to construct such complete networks. That is, the collapses are essential. A robust network for selective attack is not a network which never collapses, but a network which resists the attacks with higher efficiency before the collapses and has endurance to collapse.

A selective attack strategy actually defines a sequence of attacks for a specific network. Each element in the sequence is a set, i.e., an attack.

**Proposition 3:** For a specific network, for any given measure if the attack strategy based upon this measure can generate an attack sequence which is the same as the sequence of RD, then this attack strategy is the worst before the collapse.

This proposition is obvious, it means that if a network is robust under the selective attacks, then this network has these features: 1) the attack strategies based upon the importance measures are the same as RD, 2) the networks can resist longer, that is, the collapses would emerge later.

As to the scale-free networks, since ID is very similar to RD, we can use ID to approximate RD. Hence, proposition 3 can be rewritten as follows,

**Proposition 3’:** For a specific scale-free network, for any given measure, if the attack strategy based upon this measure can generate an attack sequence which is the same as the sequence of ID, then this attack strategy is bad before the collapse.

Some scale-free networks, such as the CSF-Compact network, have the compactness property, that is, if one node is important in a measure, it is often also important in another measure.

Let us consider the endurance before the collapses for the scale-free networks. If the exponent of the degree distribution keeps constant, the relative important nodes, even the trivial nodes would gain more edges when the average degree increases. This fact means that the ID attack would be more difficult to divide the network.

According to the analysis above, it would be reasonable that the compact scale-free networks with higher average degree would resist the ID attack strategy, and simultaneously the other selective node attack strategies.

In general, this letter advances the current ideas on the fragility of the scale-free networks and proposes that some scale-free networks could be robust under selective node attacks when the networks are compact with a high average degree considering the cost of attacks. The experimental results on five networks validate this idea.

However, this letter spawns more problems. First, what kinds of networks would be the most robust networks under a cost constraint? This problem is related to the construction of robust social, biological and technological networks, or to the destruction of these networks, and probably it would be useful to understand the evolutionary principles of these networks. Second, what will happen when considering the vertices attacks? Third, does a phase transition exist for the average degree vs. the robustness of the networks? Fourth, which attack strategies would be
the most efficient strategies for a specific network? etc. These problems will be investigated in the future works. Besides, the conclusions in this letter could be potential to explain the evolutionary principles of biological populations and social systems which obey power law distributions. In these networks, the evolution of nodes can be due to the new nodes which could be helpful for the stability of the whole networks, that is, the nodes would obey the law that “the most stable nodes survive”, here the stability is the adaption. In the BA model, the nodes would obey the law that “the strongest nodes survive”, here the power is the adaption. By comparing to these two models, we can conclude that the previous one provides one holistic viewpoint, and the BA model is based upon the individual viewpoint.

***

Thanks to Dr J. Lü, Dr O. Burger and Dr C. L. Zhou. Supported by the National Basic Research Program of China (No. 2007CB310804), the National Natural Science Foundation of China (No. 60803095), the State key Laboratory of Networking and Switching Technology (No. SKLNST-2010-1-04) and the SCUEC foundation (No. YZZ06025).

REFERENCES

[1] Albert R., Jeong H. and Barabási A.-L., Nature, 406 (2000) 378.
[2] Holme P., Kim B. J., Yoon C. N. and Han S. K., Phys. Rev. E, 65 (2002) 056109.
[3] Tan Y., Wu J., Deng H. and Zhu D., Syst. Eng., 24 (2006) 1.
[4] Stanley W. and Katherine F., Social Network Analysis: Methods and Applications (Cambridge University Press) 1994.
[5] Vázquez A., Pastor-Satorras R. and Vespignani A., Phys. Rev. E, 65 (2002) 066130.
[6] Barabási A. L. and Albert R., Science, 286 (1999) 509.
[7] Olaf S., Complexity, 8 (2002) 56.
[8] Broder A., Kumar R., Maghoul F., Raghavan P., Rajagopalan S., Stata R., Tomkins A. and Wiener J., Comput. Netw., 33 (2000) 399.
[9] Jeong H., Mason S. P., Barabási A. L. and Oltvai Z. N., Nature, 411 (2001) 41.
[10] Jennifer A. D., Richard J. W. and Neo D. M., Ecol. Lett., 5 (2002) 558.
[11] Newman M. E. J., Forrest S. and Balthrop J., Phys. Rev. E, 66 (2002) 035101.
[12] Magoni D., J. Sel. Areas Commun. IEEE, 21 (2002) 399.
[13] Zheng B., Technical Report, Tsinghua University, PostDoc116-2008 (2009).
[14] Zheng B., Wang J., Chen G., Jiang J. and Shen X., Chin. Phys. Lett., 28 (2011) 018901.
[15] Freeman L. C., Soc. Netw., 1 (1978) 215.
[16] Brin S. and Page L., Comput. Netw. ISDN Syst., 30 (1998) 107.
[17] Page L., Brin S., Motwani R. and Winograd T., Technical Report, Computer Science Department, Stanford University (1998).
[18] Jon M. K., J. Assoc. Comput. Mach., 46 (1999) 604.
[19] Newman M. E. J., SIAM Rev., 45 (2003) 167.
[20] Clauset A., Shalizi C. and Newman M. E. J., SIAM Rev., 51 (2009) 661.