Iterative path attacks on networks

Cunlai Pu[a,*], Siyuan Li[a], Andrew Michaelson[b], Jian Yang[a]

[a] *Department of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China*
[b] *Department of Biomedical Engineering, Stony Brook University, New York 11790, USA*

Abstract

We investigate a path-attack process on model networks and real-world networks. Based on the local topological structure of a path, we propose an attack centrality measure with a control parameter α for quantifying the influence of a path. In the path-attack process, we iteratively remove the path with the largest attack centrality from a network. Results demonstrate that, for a specific network, there is an optimal α which results in maximum attack efficiency. The denser and more homogenous the networks, the more robust the networks are against the iterative path attacks. Our work helps to explain the vulnerability of networks and gives some clues to the protection and design of real complex systems.

Keywords: attack centrality, vulnerability, network efficiency

1. Introduction

Malfunction of many complex systems happens occasionally, such as breakdown of the Internet, blackouts in power grids, distress to the human body, etc. To study the vulnerability of complex systems, researchers simply represent complex systems by networks, in which nodes are individuals and edges are interactions among the nodes. This has led to new insights into the fragility of complex networks [1, 2, 3, 4]. The percolation process is defined as removing some fraction of vertices and incidental edges from networks. There is a percolation threshold at which a giant component forms or vanishes in a network, and this is studied in model networks and real-world

*200 Xiaolingwei, Nanjing 210094, China. Tel: +8613915966537.
Email address: pucunlai@njust.edu.cn (Cunlai Pu)

networks [5, 6, 7]. Albert et al [8] found that based on their subset of network data the Internet as well as the World Wide Web is robust against random failures, while highly vulnerable to intentional attacks on its highest-degree nodes. Similar results were obtained in a much larger mapping of the World Wide Web [9], and other complex networks including metabolic networks [10], email networks [11], and food webs [12] etc. Based on the degree and the betweenness centrality, Holme et al [13] thoroughly studied attacks on nodes as well as edges for model networks and real-world networks. Sometimes, a tiny fraction of attacks or failures will cause a cascade of failures in the network which eventually destroys the function of the system, such as the large-scale blackout of power grids, the collapse of the Internet, etc [14, 15, 16]. Many models of cascading failure [17, 18, 19, 20] assume that overload of a few nodes or edges results in the redistribution of loads among their neighbors, and this causes the overloads to continue. Most recent results [21, 22, 23, 24] show that cascading failures have strong connections with the interaction of two or more networks. For example, failure of a power station in the power grid network cuts off the power supply of many computers in the communication network. The power down of computers in turn affects the communication and control of the power stations. This cycle continues, which leads to the collapse of the power grid network and the communication network [4].

Much progress [2, 3] has been made in understanding the influence of the removal of nodes or edges on the overall network performance. However, in real situations a powerful attack usually affects part of, or even the whole network. For instance, snowstorms often block the transportation of an entire city. Additionally, bombs destroy parts of battlefields. Therefore, a fundamental question is how the removal of a more complex subgraph, such as a path, will affect the overall behavior of the network, and what will happen if we iteratively remove a subgraph from a network. By identifying the key subcomponents, we can better protect networks such as the Internet, power grid, and city transportation networks etc., and effectively degrade those networks including criminal or terrorist networks, contact networks for infectious diseases, etc. In this letter, we measure the vulnerability of model networks and real-world networks under path-based attacks. We propose an attack centrality measure based on only the local topological structure of a path. Then we attack the paths with the largest attack centrality iteratively. We use several ways to quantify the damage of the attacks.

2. Attack centrality of a path

Let $G\langle V, M\rangle$ represent a graph G, where V is the node set of G, and M is the edge set of G. Assuming a path P of length l to be a node sequence $x_1x_2\ldots x_{l+1}$, where $x_ix_{i+1} \in M$, $1 \leq i \leq l$. Let V_P and M_P represent the node set of P and the edge set of P respectively. We define the local topological structure of path P as $G_P\langle V', M'\rangle$, where V' contains all nodes of P and the neighboring nodes directly connected to P. $V' = V_P \cup \{x \in V | \forall y \in V_P, (x, y) \in M\}$. M' is the edge set of G_P, which is composed of all the edges between the nodes in V'. $M' = \{(x, y) \in M | x \in V', y \in V'\}$. Based on the local topological structure of a path, we can obtain the degree of path P which is equal to $|V' - V_P|$, and this is actually the number of nodes connecting P.

In a large-scale network, it is either expensive or impossible to get the whole network topology to quantify the importance of a path [25, 26]. A reasonable way is to evaluate the influence of a path locally. Also, the measure of importance of a path should correspond to specific applications. Here we aim to quantify the importance of a path in network attack dynamics based on the defined local topological structure. When attacking a network, we always expect to attack the critical path, removal of which results in a thorough breakdown of the network. Generally, a path with more neighbor nodes is more important, since more nodes are likely affected by the removal of the path. On the other hand, fewer edges among neighbor nodes of a path are desirable, since the connectivity between these nodes is likely affected by the removal of the path, if the connections among these nodes are sparse. Otherwise, the removal of a path will hardly affect the efficiency of the network. Therefore, to better quantify the importance of a path in an attack scenario, we need to consider both the number of nodes and number of edges in the local structure of a path, which is the primary idea of our attack centrality measure. We adjust the influence of number of nodes ($|V'|$) and the influence of number of edges ($|M'|$) with a control parameter α. The attack centrality of path P is as follows:

$$C_P = \frac{|V'|^\alpha}{|M'|^{1-\alpha}}. \tag{1}$$

Where $0 \leq \alpha \leq 1$. When α equals 1, the attack centrality measure degenerates into the degree centrality measure (DC). When α equals 0, the attack centrality is completely dependent on the number of edges in the local structure, and the fewer the edges, the larger the path centrality is. We call this

the edge-dependent centrality (EDC) measure.

3. Path-attack process

Prior to attacking paths, we need to find out the concerned paths in networks. Considering a network composed of N nodes, the maximal number of paths of length 2 is $N(N-1)(N-2)/2$. Since lots of real-world networks have large topological structures, the number of paths in these networks could be enormous. Furthermore, paths in a network usually have overlapping nodes and edges. Thus, it is hard to enumerate all the paths in a network. For the sake of simplification, we remove only the shortest paths of equal length when attacking networks. For instance, the number of shortest paths of a specific length is not more than N^2 in a network of N nodes, if just one of the shortest paths is considered between any two nodes. Otherwise, it is convenient to find the shortest paths in a network by using the existing algorithms [27] such as the Dijkstra algorithm, the Floyd algorithm, and so on.

To investigate the effects of the shortest paths on the vulnerability of a network, we iteratively remove the shortest paths based on their attack centralities, and measure the behavior of the network subject to this kind of attack. Specifically, each time we remove the shortest path with the largest attack centrality during the path-attack process. The process stops until there are no shortest paths of the given length in the remaining network. Note when attacking a path, we only remove all edges of the path from the network, and keep all nodes of the path in the remaining network for the sake of simplification. Also, we only consider undirected networks. As for the directed real-world networks, we simply ignore the directions of the edges in the experiments.

4. Measures of damages

In the attack process, vulnerability of a network is directly reflected by the change in the largest component's size. Also, degeneration of function of a network is strongly correlated with the decrease of network efficiency. To measure the damage caused by the path attacks, we compute the size of the largest component as well as the efficiency of the remaining network at each

step of the attack process. The network efficiency is defined as follows [28]:

$$E = \frac{1}{N(N-1)} \sum_{i \neq j \in V} \frac{1}{d_{i,j}}. \tag{2}$$

Where V is the node set composed of all the nodes in the network, N is the number of nodes in the network, and $d_{i,j}$ represents length of the shortest path from node i to node j. If node i is not reachable from node j, then $d_{i,j}$ is set to ∞.

Furthermore, we record the total iterative steps of a path-attack process which indicates the robustness of a network to a specific path-attack process. The more attack steps the networks undergo, the more robust the networks are. We also record the time step by which the network efficiency reduces by half. This time step indicates the attack efficiency of a path-attack process, and the smaller the time step, the larger the attack efficiency. Through measuring these properties, we study the effects of the iterative removal of paths on model networks and real-world networks.

5. Simulation results

The model networks we used in the simulation are generated by the ER (Erdös-Rényi) model [29] and the static model [30]. The networks generated by the ER model are called random networks. These random networks have the small-world property and the Poisson distribution of node degrees. The networks generated by the static model are called scale-free networks. The scale-free networks also have the small-world property. The degree distribution of the scale-free networks obeys the power law, which results in the robust yet fragile features of scale-free networks. In addition to model networks, we also employ real-world networks in the simulations. The statistics for real-world networks are shown in Table 1.

5.1. Model networks

First, we investigate the path-attack processes on model networks including random networks and scale-free networks. Network size is $N = 600$. The average node degree is $\langle k \rangle = 7$. The power-law parameter is $\gamma = 2.5$. The length l of the shortest paths for removal is fixed to 4. In the simulation, each time we greedily remove the shortest path with the largest attack centrality, and measure the size of the largest component S and the network efficiency

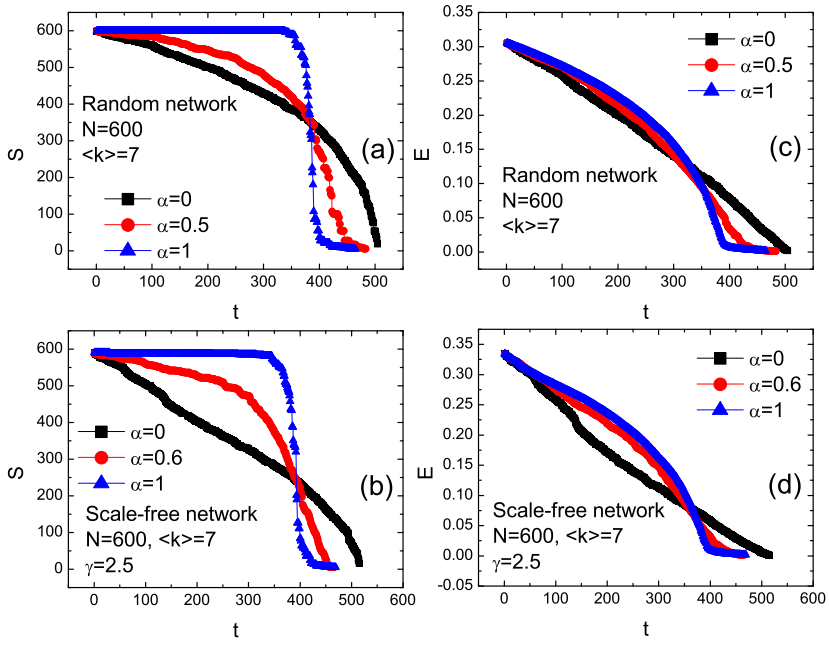

Figure 1: Largest component's size S ((a), (b)) and network efficiency E ((c), (d)) vs. time step t in the path-attack process on model networks.

E. In Fig. 1, S and E decrease with the attack step t for both the random network and the scale-free network. The decrease rate is related to α. Based on the results of S, we can infer that there is phase transition for certain α (for example $\alpha = 1$ in Fig. 1). When $\alpha = 1$, the attack centrality is only dependent on the number of nodes in the local structure of a path according to our definition of path-attack centrality. Then, the paths in the dense part of the network are attacked with priority since they probably have large number of neighboring nodes, and this doesn't affect the giant component's size much at the beginning of the iterative path attacks. However, with the attacks continuing the network becomes more and more sparse, and at a certain time step the network will collapse into pieces shown in Fig. 2. On the other hand, when $\alpha = 0$ the attack centrality of a path is only dependent on the number of edges in the local structure, and fewer edges means larger attack centrality according to Eq. 1. Therefore, when $\alpha = 0$ the path attacks happen in the sparse area of the network with priority, and this causes the decrease of the giant components size directly, shown in Fig. 2. There is no phase transition when $\alpha = 0$, which is opposite to $\alpha = 1$.

Since different parameter α lead to different attack efficiency, we can detect the approximately optimal α (α_{opt}) through enumeration. On the other hand, α_{opt} is determined by the measure of damage caused by the path-attack process. For example, we consider the half efficiency time T_h which is the time step the network efficiency is reduced by 50%. We expect that T_h is as small as possible, since a smaller T_h means a larger attack efficiency. As shown in Fig. 3(a) and (b), T_h generally decreases with α first, then increases with α both for the random networks and the scale-free networks. The minimal T_h corresponds to the approximate α_{opt}. We also calculated the 95% efficiency time $T_{95\%}$, which is the time step that the network efficiency is reduced by 5%, shown in Fig. 3(c) and (d). We can see that for $T_{95\%}$ the optimal parameter α_{opt} is 0. In the calculation of T_h and $T_{95\%}$, network size N is fixed to 600. The average degree $\langle k \rangle$ is 6. The power-law parameter γ for the scale-free networks is 3. The length l of the attacked shortest path is 4. The simulation results are the average of 100 independent runs. The conclusions of the results are consistent for both the random networks and the scale-free networks.

Furthermore, the attack efficiency of the path-attack process, as well as α_{opt}, depends on topological properties of the networks. In Fig. 4, we show the results of α_{opt} and the corresponding T_h as a function of average degree $\langle k \rangle$. N and γ are fixed to be 600 and 3 respectively. $l = 4$. The results are

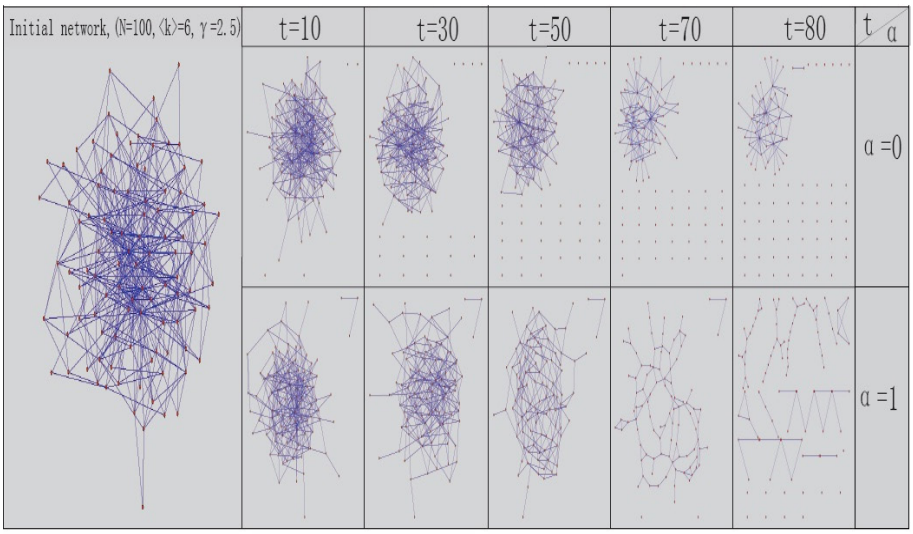

Figure 2: Selected snapshots of the network in the iterative path attacks for $\alpha = 0$ and $\alpha = 1$. The initial network is generated by the static model. The length l of attacked paths is 4. When $\alpha = 0$, paths in the sparse area of the network are attacked with priority. When $\alpha = 1$, paths in the dense area of the network are attacked with priority. We see clearly that there is phase transition for $\alpha = 1$, while there is no phase transition for $\alpha = 0$.

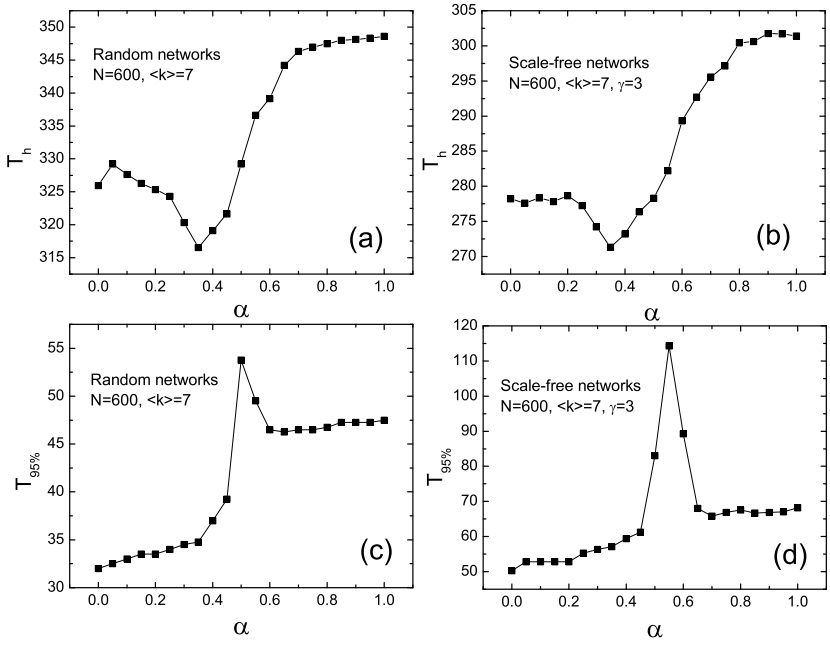

Figure 3: T_h and $T_{95\%}$ vs. α for (a) (c) random networks and (b) (d) scale-free networks. The results are the average of 100 independent runs.

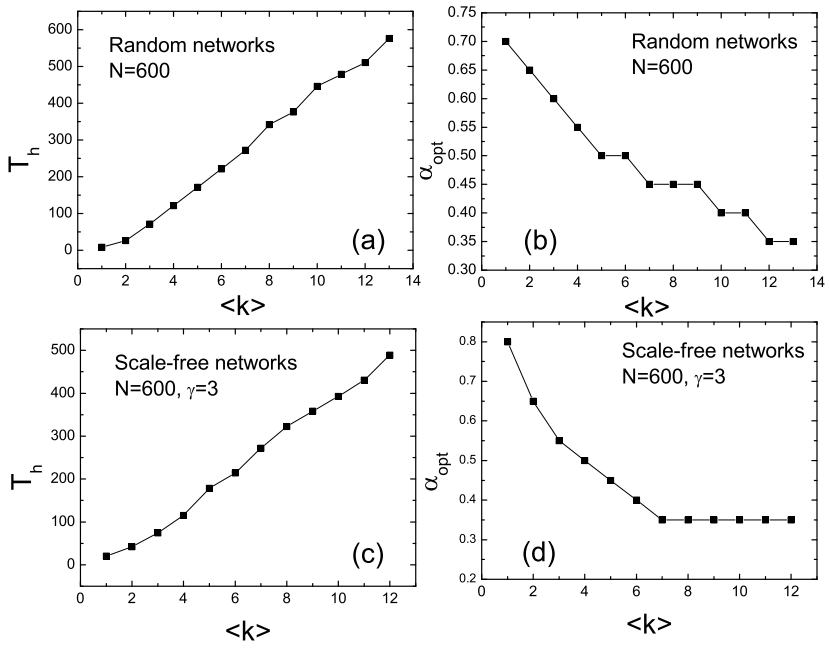

Figure 4: T_h and the corresponding α_{opt} vs. $\langle k \rangle$ for random networks ((a), (b)) and scale-free networks ((c), (d)). The results are the average of 30 independent runs.

the average of 30 independent runs. It can be clearly seen from Fig. 4 that T_h increases with average degree $\langle k \rangle$ almost linearly in both random networks and scale-free networks. This indicates that the more edges the networks have, the more robust the networks are against the path-attack process. When $\langle k \rangle$ is sufficiently large, the probability that the model networks have the shortest paths of length 4 becomes very small, which is predicted from Fig. 5. Therefore, the path-attack process will stop in a few steps, or is not able to start due to lack of the shortest paths of length 4, when $\langle k \rangle$ is large enough. α_{opt} decreases with $\langle k \rangle$ in Fig. 4(b) and (d). This indicates that when the network becomes dense, the number of nodes becomes less important than the number of edges in the local structure of a path, for the process of effectively quantifying the attack centrality of a path. We then study T_h as a function of γ for scale-free networks. N and $\langle k \rangle$ are fixed to be 600 and 7 respectively. $l = 4$. The simulation results are the average of 30 independent runs, shown in Fig. 6. T_h increases abruptly at the beginning, and then saturates with the increase of power-law parameter

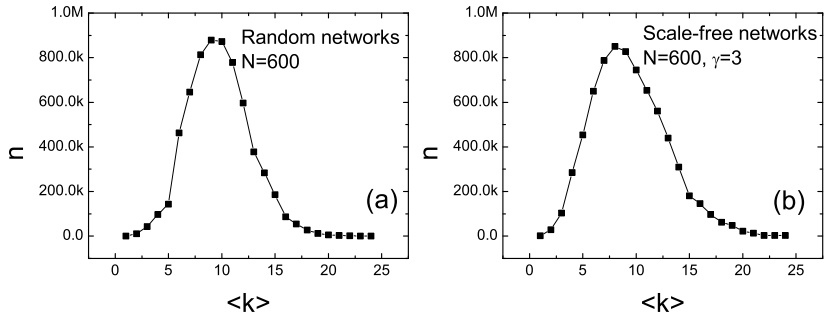

Figure 5: Number n of shortest paths with length $l = 4$ in (a) random networks and (b) scale-free networks. The results are the average of 20 independent runs.

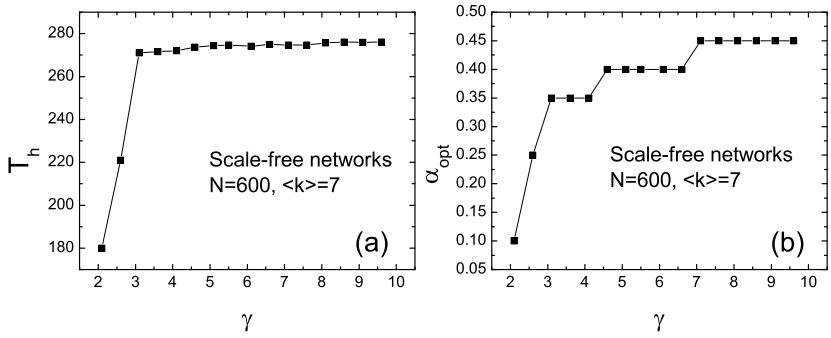

Figure 6: (a) T_h and (b) the corresponding α_{opt} vs. γ for scale-free networks. The results are the average of 30 independent runs.

γ, as shown in Fig. 6(a). This indicates that the more homogeneous the degree distribution, the more robust the networks are, which is consistent with the previous results [8]. In Fig. 6(b), α_{opt} for T_h increases with γ, which means the number of nodes in the local topological structure of a path becomes important to effectively quantify the attack centrality of a path. However, the influence of number of nodes should not surpass the number of edges in the local topological structure of a path, since α_{opt} is less than 0.5 in Fig. 6(b). We also show the results of $T_{95\%}$ as a function of network size N for both the random networks and the scale-free networks in Fig. 7. The simulation results are the average of 100 independent runs, in which $\langle k \rangle = 6$, $\gamma = 2.5$, and $l = 4$. The simulation results demonstrate that $T_{95\%}$ appropriately increases with N linearly for the ER random networks, while

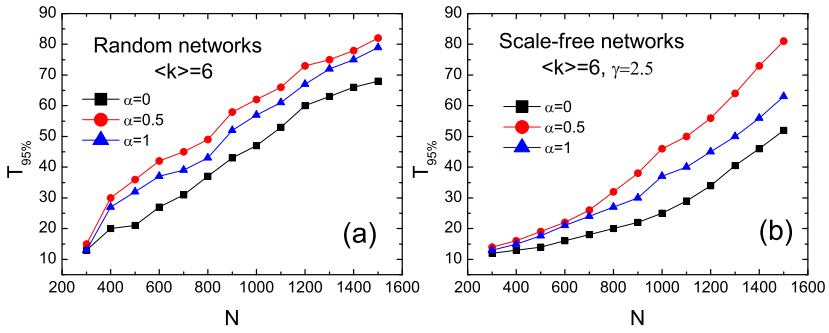

Figure 7: (a) $T_{95\%}$ vs. N for (a) random networks and (b) scale-free networks. The results are the average of 100 independent runs.

exponentially for the scale-free networks.

5.2. Real-world networks

For real-world networks, we focus on the variation of the network efficiency E with attack step t. The shortest paths of length 4 are considered for removal in the simulation. In Fig. 8, we see that E decreases with t, and different α corresponds to different decrease rates, which are similar to the model networks. We also obtain α_{opt} under the measure of T_h. Path length l is set to 3 and 4 in the simulation. These results are shown in Table 2.

6. Conclusions and discussion

In summary, we propose an attack centrality measure characterized with a parameter α for a path based on the local topological structure of a path. Then we study the path-attack process on networks, in which the path with the largest attack centrality is removed from the network each time. Results show that the largest component's size and the network efficiency decrease with the attack steps. The decrease rate depends on the parameter α, which is consistent for both model networks and real-world networks. For certain α (eg $\alpha = 1$), the network will undergo a phase transition during the path-attack process. For a specific network, there is an optimal α that corresponds to the maximum attack efficiency, and the optimal α is generally different for various measures of attack efficiency. We focus on the time step T_h when the network efficiency is reduced by half, and the optimal α corresponding to the minimal T_h. The attack efficiency decreases with average node degree

Table 1: Statistics of some real-world networks.

NAME	NODES	EDGES
PolBooks [31]	105	441
C.elegans [32]	297	2359
Metabolic [33]	453	4596
US Air lines [34]	332	2126
Jazz [35]	198	5484
Political blogs [36]	1222	19021
Word adjacencies [37]	112	425
WorldCities [38]	415	7518
Small world [39]	396	1988
Scotland [40]	244	356
NetScience [41]	1589	2742

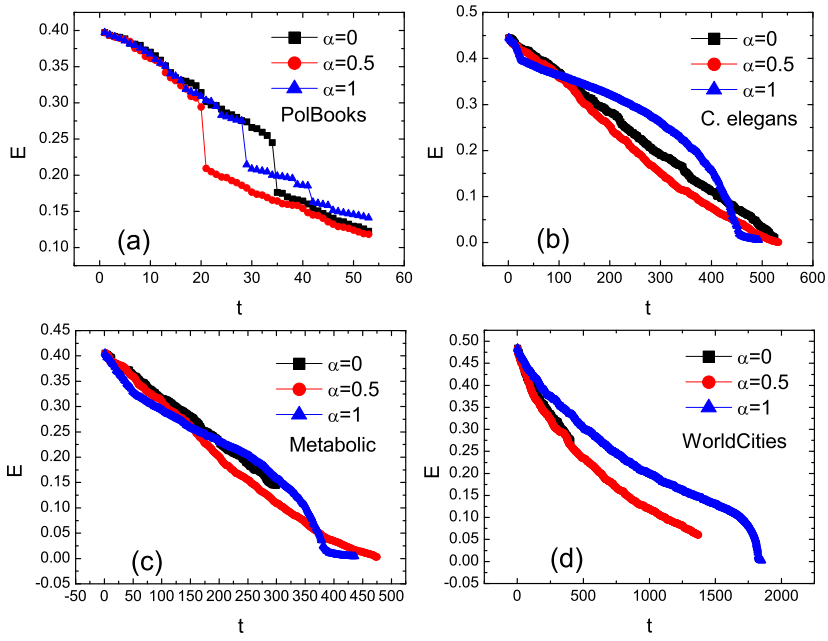

Figure 8: E vs. t for different α on real-world networks.

Table 2: T_h and the corresponding α_{opt} for real-world networks.

NAME	$l=3$		$l=4$	
	T_h	α_{opt}	T_h	α_{opt}
Word adjacencies	65	0.35	43	0.4
C.elegans	313	0.55	225	0.55
Jazz	322	0.55	189	0.5
Metabolic	264	0.55	190	0.6
US Air lines	141	0.4	88	0.4
Political blogs	1183	0.4	806	0.35
WorldCities	765	0.55	465	0.5
Small world	139	0.6	83	0.75
Scotland	24	0.6	21	0.6
NetScience	28	0.7	15	0.6

and network size for both random networks and scale-free networks. The more homogenous the degree distribution, the more robust the networks are against iterative path attacks.

We study the impacts of network structures on the path-attack process based mainly on the T_h measure. The results of optimal α may be different for other measures. Furthermore, for those networks that have special topological structures, their efficiency may not be reduced to half for all parameters of α. In each step of the simulation, we simply use the Dijkstra algorithm to find out all the shortest paths, and remove the one with the largest attack centrality. We can also consider removing other paths, such as general paths, but it is tough to enumerate them in a network. We only focus on the importance of a path on attack dynamics. The influence of a path on the spread [42, 43], control [44, 45], and prediction [46] still needs exploring.

Acknowledgments

This work was supported by the Natural Science Foundation of China (Grant No. 61304154), the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20133219120032), and the Postdoctoral Science Foundation of China (Grant No. 2013M541673).

References

[1] A. Barrat, M. Barthelemy, A. Vespignani, Dynamical processes on complex networks, Cambridge: Cambridge University Press, 2008.

[2] M. Newman, Networks: an introduction, Oxford University Press, 2010.

[3] T. G. Lewis, Network science: Theory and applications, John Wiley & Sons, 2011.

[4] G. D'Agostino, A. Scala, Networks of Networks: The Last Frontier of Complexity, Imprint: Springer, 2014.

[5] D. S. Callaway, M. E.J. Newman, S. H. Strogatz, D. J. Watts, Phys. Rev. Lett. 85 (2000) 5468.

[6] N. Schwartz, R. Cohen, D. Ben-Avraham, A.-L. Barabási, S. Havlin, Phys, Rev. E 66 (2002) 015104.

[7] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, Phys. Rev. Lett. 96 (2006) 040601.

[8] R. Albert, H. Jeong, A.-L. Barabási, Nature 406 (2000) 378.

[9] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener, Computer networks 33 (2000) 309.

[10] H. Jeong, S. Mason, A.-L. Barabási, Z. N. Oltvai, Nature 411 (2001) 41.

[11] M. E. J. Newman, S. Forrest, J. Balthrop, Phys. Rev. E 66 (2002) 035101.

[12] J. A. Dunne, R. J. Williams, N. D. Martinez, Proc. Natl. Acad. Sci. USA 99 (2002) 12917.

[13] P. Holme, B. J. Kim, C. N. Yoon, S. K. Han, Phys. Rev. E 65 (2002) 056109.

[14] A. E. Motter, Y. C. Lai, Phys. Rev. E 66 (2002) 065102.

[15] A. E. Motter, Phys. Rev. Lett. 93 (2004) 098701.

[16] L. Zhao, K. Park, Y. C. Lai, Phys. Rev. E 70 (2004) 035101.

[17] P. Crucitti, V. Latora, M. Marchiori, Phys. Rev. E 69 (2004) 045104.

[18] R. Kinney, P. Crucitti, R. Albert, V. Latora, Eur. Phys. J. B 46 (2005) 101.

[19] W. X. Wang, G. Chen, Phys. Rev. E 77 (2008) 026101.

[20] Z. X. Wu, G. Peng, W. X. Wang W X, S. Chan and E. Wing-Ming. Wong, J. Stat. Mech. 05 (2008) P05013.

[21] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, S. Havlin, Nature 464 (2010) 1025.

[22] J. Gao, S. V. Buldyrev, S. Havlin, H. E. Stanley, Phys. Rev. Lett. 107 (2011) 195701.

[23] C. D. Brummitt, R. M. D'Souza, E. A. Leicht, Proc. Natl. Acad. Sci. USA 109 (2012) E680.

[24] A. Bashan, Y. Berezin, S. V. Buldyrev, S. Havlin, Nature Physics 9 (2013) 667.

[25] W. X. Wang, B. H. Wang, C. Y. Yin, Y. B. Xie, T. Zhou, Phys. Rev. E 73 (2006) 026111.

[26] D. Chen, L. Lü, M. S. Shang, Y. C. Zhang, T. Zhou, Physica A 391 (2012) 1777.

[27] D. B. West, Introduction to graph theory, Upper Saddle River: Prentice hall, 2001.

[28] V. Latora, M. Marchiori, Phys. Rev. Lett. 87 (2001) 198701.

[29] P. Erdös, A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. 5 (1960) 17.

[30] K.-I. Goh, B. Kahng, D. Kim, Phys. Rev. Lett. 87 (2001) 278701.

[31] V. Krebs, unpublished http://www. orgnet. com, 2008.

[32] D. J. Watts, S. H. Strogatz, Nature 393 (1998) 440.

[33] J. Duch, A. Arenas, Phys. Rev. E 72 (2005) 027104.

[34] V. Batagelj, A. Mrvar, Pajek datasets, Web page http://vlado. fmf. uni-lj. si/pub/networks/data, 2006.

[35] P. Gleiser, L. Danon, Adv. Complex Syst. 6 (2003) 565.

[36] L. A. Adamic, N. Glance, in Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem, 2005.

[37] M. E. J. Newman, Phys. Rev. E 74 (2006) 036104.

[38] P. J. Taylor, World city network: a global urban analysis, Psychology Press, 2004.

[39] S. Milgram, Psychology Today 1 (1967) 61.

[40] J. Scott, M. Hughes, J. Mackenzie, The anatomy of Scottish capital: Scottish companies and Scottish capital, 1900-1979, London: Croom Helm, 1980.

[41] M. E. J. Newman, Phys. Rev. E 74 (2006) 036104.

[42] H. X. Yang, W. X. Wang, Z. X. Wu, B. H. Wang, Physica A 387 (2008) 6857.

[43] K. Wang, Y. F. Zhang, S. Y. Zhou, W. J. Pei, S. P. Wang, T. Li, Physica A 390 (2011) 2593.

[44] G. Yan, J. Ren, Y. C. Lai, B. Li, Phys. Rev. Lett. 108 (2012) 218703.

[45] Z. Yuan, C. Zhao, Z. Di, W. X. Wang, Y. C. Lai, Nature communications 4 (2013) 2447.

[46] L. Lü, M. Medo, C. H. Yeung, Y. C. Zhang, Z. K. Zhang, T. Zhou, Phys. Rep. 519 (2012) 1.