Physica A 514 (2019) 121–132

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Identifying critical nodes' group in complex networks

Zhong-Yuan Jiang *, Yong Zeng, Zhi-Hong Liu, Jian-Feng Ma

School of Cyber Engineering, Xidian University, Xi'an, Shaanxi 710071, China
Shaanxi Key Laboratory of Network and System Security, Xidian University, Xi'an, Shaanxi 710071, China

HIGHLIGHTS

- Network robustness is evaluated under various attack strategies.
- Critical nodes' group mining problem is proposed and extensively discussed.
- Potential applications are possible in network security area.

ARTICLE INFO

Article history:
Received 30 March 2018
Received in revised form 25 June 2018
Available online 19 September 2018

Keywords:
Key node
Target attack
Network vulnerability
Complex network

ABSTRACT

Recently, network vulnerability or security has attracted much attention in various networked systems, and especially in security related attacks or protections, there are a set of influential nodes that can remarkably break the network connectivity. In this work, we firstly present eight attack mechanisms including target attack, random failure, betweenness based attack, closeness based attack, PageRank based attack, k-shell based attack, greedy algorithm, and low-degree attack. Secondly, inspired by the dynamic node removal process, we propose to recalculate the metrics for every node removal strategy, and evaluate the network robustness against all these heuristic attack strategies with and without recalculations in scale-free networks, random networks, and many real network models. The simulations indicate that most of the attack strategies with recalculations appear to imperil the network structure security more. Furthermore, considering that key node set mining is very critical for network structure protections, we employ minimum number of key nodes (MNKN) metric to further discuss the network vulnerability against all the attack strategies with or without recalculations. The results show that the critical nodes' group can be more efficiently found under the PageRank based attack with recalculations than under other attack disciplines with or without recalculations in most of the classic and real network models. This work investigates network structure vulnerability and security from a new perspective, and has potential applications into network structure protection or planning.

1. Introduction

Nowadays, almost all things are connected via real or virtual links, constructed various complex network systems [1–3] such as communication networks, World Wide Web (WWW), social networks (*e.g.* WeChat [4], Facebook [5]), neural networks, ecosystem, food web, power grid, highway networks, Internet of things, and so on. The network structure plays a critical role for every networked system to realize its functions or values. Moreover, many empirical studies [6]

* Corresponding author.
E-mail address: zyjiang@xidian.edu.cn (Z.-Y. Jiang).

https://doi.org/10.1016/j.physa.2018.09.069
0378-4371/

have proved that a fraction of critical nodes in a network are very influential in vulnerability evaluating [7,8], cascade spreading [9], controlling [10], synchronizing [11] and virus marketing [12], and the node importance ranking has attracted lots of attention in recent decades. The well-known PageRank [13] algorithm can be efficiently employed into large-scale networked systems for critical node mining. Inspired by the characterized networks features such as node degree centrality, betweenness centrality [14,15], closeness centrality [16], k-shell index [17], and so on, many heuristic influential node mining strategies have been extensively evaluated by the Susceptible–Infected–Recovered (SIR) model [18]. Furthermore, the vital link identifying [19] and path based attack [20] has been concerned with other robustness analysis [21–26] or improvement [27,28].

However, the influence of one key node is very limited, and the most important is the key-node-set problem [8] which mainly focuses on finding a set of nodes whose simultaneous failure will lead to the whole collapse of a network. Such a set of nodes are very critical for many real complex systems [29]. In transport networks, large scale traffic congestion is often caused by the original jams on several vital road sections. In airline systems, for the convenience of resource locations or passengers, *e.g.* maintenance crews, it is very vital to control the hub nodes of an airline [30]. In IT infrastructure, service providers often control the Internet traffic on many critical nodes in the search for viruses [10]. In interdependent networks (*e.g.* power grids and communication networks), a portion of vital nodes may lead to the collapse of whole interdependent network, such as the largest blackout of the power grid and the outages of the Internet [31,32]. In social science, for security purpose, a fraction of inside agents are located to intercept all communications in a network of terrorists [33]. In a food web, the predation relation of all kinds of species are strongly dependent, and due to the disappearance of several species, a large scale of other species will suffer species' extinction [34]. Our previous work [9] aims to discover critical nodes group which is the threshold of the network structure security. As discussed in Ref. [7], the network vulnerability is a fundamental security character. When a fraction of nodes with adjacent links are removed, the network broke into many sub-network pieces or even whole collapse. In this work, we aim to first evaluate the effect of different key node mining methods on network robustness, and then discuss the comparisons of the minimum number of key nodes which can lead to total network collapse under all employed heuristic mechanisms.

2. Methods & models

2.1. Attack strategies

Inspired by our previous work [8], in network robustness evaluation, the selection sequence of attacked nodes can remarkably influence final results. For example, under the target attack [28] mechanism, the nodes are sorted by degree in original network from high to low, and the attacked nodes are selected one by one from the sequence. However, the attack process is dynamic. When a fraction of nodes of high degrees are removed, a node of high degree in original network might have very small degree or even be isolated in the survived network. From the attacker perspective, he might sufficiently sense the dynamic characters and change attack targets intensively. In other words, in our opinion, the recalculation of the used heuristic characters might lead to larger destruction of network structure. Therefore, from comparison perspective, here we employ several heuristic influential node mining methods with recalculations and without recalculations.

It is widely observed that a node of the highest degree is often considered as an important one in a network structure [35], so under the target attack mechanism, the nodes of the highest degrees are removed subsequently to disconnect the network connections. Given a network F, it can be described as follows:

Target attack (TA) without recalculation:
Step 1: Calculate the degree of all nodes, and sort all nodes in descend order, denoted as *seq*;
Step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
Step 3: The step 2 is repeated until all nodes are removed from the *seq*.

Target attack (TA) with recalculation:
step 1: Calculate the degree of all nodes, and sort all nodes in descend order, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: If the *seq* is empty, exit; else recalculate the degree of the all nodes in *seq*, and sort the *seq* in descend order, denoted as *seq* again, then go on the step 2.

Target attack can significantly imperil the structure safety of Barabási–Albert (BA) [36] network which has high robustness to random failure which can be described as follows:

Random failure (RF) without recalculation:
step 1: Calculate the degree of all nodes, and sort all nodes in descend order, denoted as *seq*;
step 2: Randomly choose a node in *seq*, remove the selected node and all links adjacent to this node, and remove this node from *seq*;
step 3: The step 2 is repeated until all nodes are removed from the *seq*.

Random failure (RF) with recalculation:
step 1: Calculate the degree of all nodes, and sort all nodes in descend order, denoted as *seq*;

step 2: Randomly remove a non-isolated node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: If the *seq* is empty, exit; else recalculate the degree of the all nodes in *seq*, and sort the *seq* in descend order, denoted as *seq* again, then go on the step 2.

In random failure mechanism, every node is selected randomly without calculating any metric of the networks, so at first glance, the RF with recalculation might has similar results compared to the RF without recalculation.

As discussed in Ref. [37], a node of high betweenness which is defined as the number of the shortest paths passing through the node, often plays an important role in spreading epidemics. In a communication or transportation network, the betweenness represents the traffic volume on the node, and directly implies the influence of the node.

Betweenness based attack (BBA) without recalculation:
step 1: Calculate the betweenness of all nodes, and sort all nodes in descend order by betweenness, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: The step 2 is repeated until all nodes are removed from the *seq*.

With a fraction of nodes removed, the betweenness of every survived node might be remarkably changed [38]. If at each node removal step, we recalculate the betweenness of all survived nodes, then selection of key nodes will be much more accurate than that without recalculation.

Betweenness based attack (BBA) with recalculation:
step 1: Calculate the betweenness of all nodes, and sort all nodes in descend order by betweenness, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: If the *seq* is empty, exit; else recalculate the betweenness of the all nodes in *seq*, and sort the *seq* in descend order, denoted as *seq*, then go on the step 2.

Closeness [16] of a node v is often used for the distance centrality and defined as the reciprocal of the sum of geodesic distances to all other nodes in the network.

$$CL(i) = \frac{1}{N-1} \sum_{j \in V \setminus i} \frac{1}{d_{ij}}, \tag{1}$$

where V is the set of all nodes in the network, and d_{ij} represents the shortest path length from node i to node j.

Closeness based attack (CBA) without recalculation:
step 1: Calculate the closeness of all nodes, and sort all nodes in descend order by closeness, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: The step 2 is repeated until all nodes are removed from the *seq*.

Closeness based attack (CBA) with recalculation:
step 1: Calculate the closeness of all nodes, and sort all nodes in descend order by closeness, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: If the *seq* is empty, exit; else recalculate the closeness of the all nodes in *seq*, and sort the *seq* in descend order, denoted as *seq*, then go on the step 2.

The PageRank algorithm [13] can efficiently identify the most influential nodes and be used to discover key nodes. Here we also employ the PageRank method to find key nodes' group.

PageRank based attack (PBA) without recalculation:
step 1: Calculate the PageRank index of all nodes, and sort all nodes in descend order by index, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: The step 2 is repeated until all nodes are removed from the *seq*.

PageRank based attack (PBA) with recalculation:
step 1: Calculate the PageRank index of all nodes, and sort all nodes in descend order by index, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: If the *seq* is empty, exit; else recalculate the PageRank index of the all nodes in *seq*, and sort the *seq* in descend order, denoted as *seq*, then go on the step 2.

The classic k-shell decomposition [17] method can also be used to classify the importance level of all nodes. A node of higher k-shell metric is considered to be more influential in influence spreading.

k-shell based attack (KBA) without recalculation:
step 1: Calculate the k-shell decomposition metric of all nodes, and sort all nodes in descend order by the metric, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: The step 2 is repeated until all nodes are removed from the *seq*.

k-shell based attack (KBA) with recalculation:

step 1: Calculate the k-shell decomposition metric of all nodes, and sort all nodes in descend order by the metric, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: If the *seq* is empty, exit; else recalculate the k-shell decomposition metric of the all nodes in *seq*, and sort the *seq* in descend order, denoted as *seq*, then go on the step 2.

As discussed in our previous work [8], in the vulnerability evaluation, at each step, the selection of every node removal is very critical for network robustness. When a node or a fraction of nodes (denoted as g) fails, the network robustness can be denoted as the relative size of the giant component

$$G(g) = \frac{N'}{N}, \tag{2}$$

where N' is the giant component size, and N is the network size. The robustness G induced by a single node can be gained by simulations. If the removed nodes are chosen according to the G, can we have a better results? Here this method is called greedy algorithm.

Greedy algorithm (GA) without recalculation:
step 1: Calculate the robustness metric G of all nodes, and sort all nodes in descend order by the metric, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: The step 2 is repeated until all nodes are removed from the *seq*.

Greedy algorithm (GA) with recalculation:
step 1: Calculate the robustness metric G of all nodes, and sort all nodes in descend order by the metric, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: If the *seq* is empty, exit; else recalculate the robustness metric G of the all nodes in *seq*, and sort the *seq* in descend order and denoted as *seq*, then go on the step 2.

As discussed in Ref. [39], if hubs suffer from heavy load, they may filter or validate the vast information, and became less susceptible to receive any information from the outside of the trust network. Then the low-degree seeding might have better effects. Here we also evaluate the effects of low-degree attack on network robustness.

Low-degree attack (LA) without recalculation:
step 1: Calculate the degree of all nodes, and sort all nodes in ascend order, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq*, and remove this node from *seq*;
step 3: The step 2 is repeated until all nodes are removed from the *seq*.

Low-degree attack (LA) with recalculation:
step 1: Calculate the degree of all nodes, and sort all nodes in ascend order, denoted as *seq*;
step 2: Remove the first node and all links adjacent to this node in *seq* and labeled as non-survived, and remove this node from *seq*;
step 3: If the *seq* is empty, exit; else recalculate the degree of the all nodes in *seq*, and sort the *seq* in ascend order denoted as *seq*, then go on the step 2.

Besides the random failure method, all other attack mechanisms used relative heuristic information such as the node degree, betweenness, closeness, k-shell metric, and so on.

2.2. Network models

In this work, we employ two widely used network models including the Barabási–Albert (BA) [36] scale-free network model and Erdös–Rényi (ER) [40] random network model. Many empirical research results have demonstrated that lots of real world networks are associated with scale-free and small-world [41] properties, and we adopt BA [36] network model to represent the infrastructure of complex networks such as the communication networks. The degree distribution of a BA [36] network is $P(k) \sim k^{-3}$. The construction of a BA [36] network is as follows. Starting from m_0 fully connected nodes, a new node with m ($m \leq m_0$) edges is added to the existing network, and the other end of every new edge is selected preferentially according to the probability

$$\Pi_i = \frac{k_i}{\sum_j k_j}, \tag{3}$$

where k_i and k_j are the degrees of node i and j respectively.

The generation of a ER [40] random network is simple and efficient. Beginning with N isolated nodes, a link is connected between every pair of nodes with probability p. Finally, a random network of about $pN(N-1)/2$ undirected links is composed.

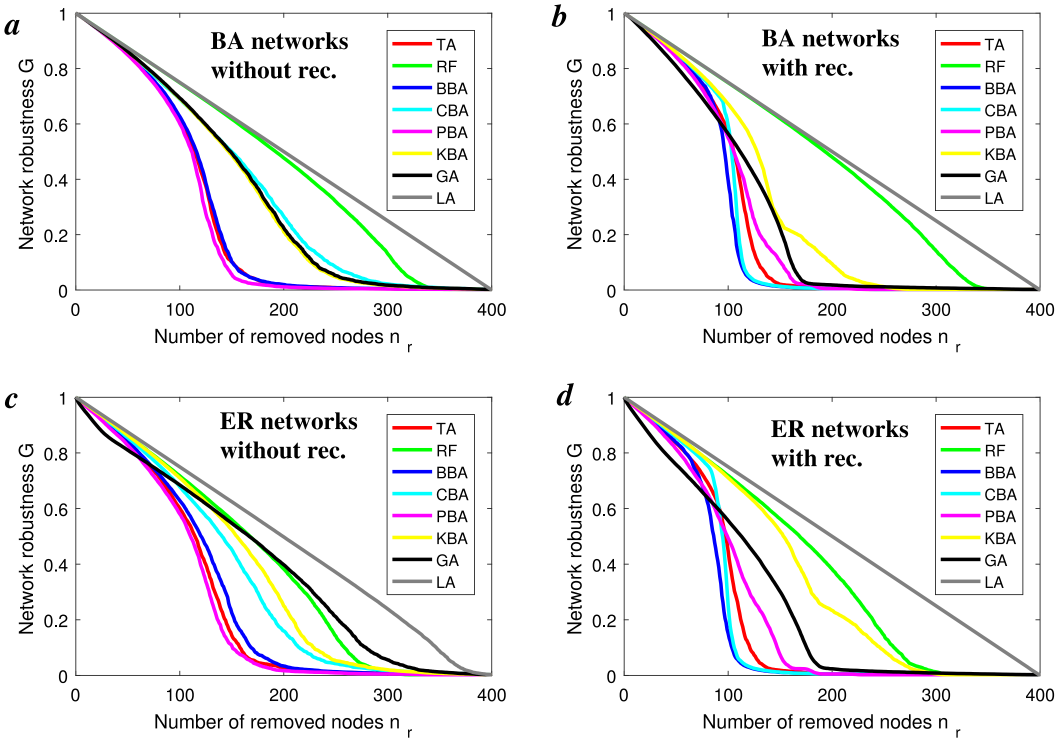

Fig. 1. (Color online). The evolutions of network robustness G under all attack methods in BA and ER networks. (a) methods without recalculations in BA networks; (b) methods with recalculations in BA networks; (c) methods without recalculations in ER networks; (d) methods with recalculations in ER networks.

3. Results

Extensive simulations are designed to test the effectiveness of the above mentioned methods. Here we assume the network size $N = 400$, and average degree $\langle k \rangle = 4$ for both BA scale-free networks and ER random networks. For every network model, the results are the average of the realizations at least 50 maps of networks. In this part, we first evaluate the network robustness, and then employ the minimum number of key nodes as another metric for critical nodes' group mining.

3.1. Robustness evaluation

As shown in Fig. 1(a), in BA scale-free networks, among all attack strategies without recalculations, the PBA method has the most destructiveness, followed by the TA and BBA methods. The BA network is very resilient to the LA mechanism. Because the LA chooses the node with the lowest degree at every step, the connectivity can be well remained. In Fig. 1(c), in ER random networks, similar results are gained. With recalculations for all attack mechanisms, in Fig. 1(b) and Fig. 1(d) the network robustness of BA and ER networks are simulated respectively. It can be found that both BA and ER networks are vulnerable to BBA method, and the closeness based attack also can remarkably imperil the network robustness.

In Fig. 2, we investigate the comparisons of all methods with or without recalculations in BA scale-free networks. It is interesting that under the random failure, PageRank based attack and low-degree attack, the network robustness seems to be the same with and without recalculations. As discussed in the above part, the random failure process with or without is almost the same, so the results are the same. In low-degree attack, due to the preferential attachment of links of BA networks, in low degree removal process with or without recalculations the selected nodes' sequences are the same. The PageRank appears to be very steady. Except these three, under the other methods, the removal process with recalculations seems to reduce the network robustness significantly, especially under the closeness based attack and greedy algorithm in Fig. 2(d) and Fig. 2(g) respectively.

In Fig. 3 we investigate the comparisons of all methods with or without recalculations in ER random networks. Different to the results of PBA in Fig. 2(e), in Fig. 3(e) the ER networks are a bit more vulnerable to PBA with recalculations than without recalculations. Still under the RF and LA, the results with or without recalculations are almost the same. Due to the homogeneous structure of ER networks, the k-shell decomposition appears to be not efficient enough. The other attack methods with recalculations deduce high vulnerability risks in ER networks too.

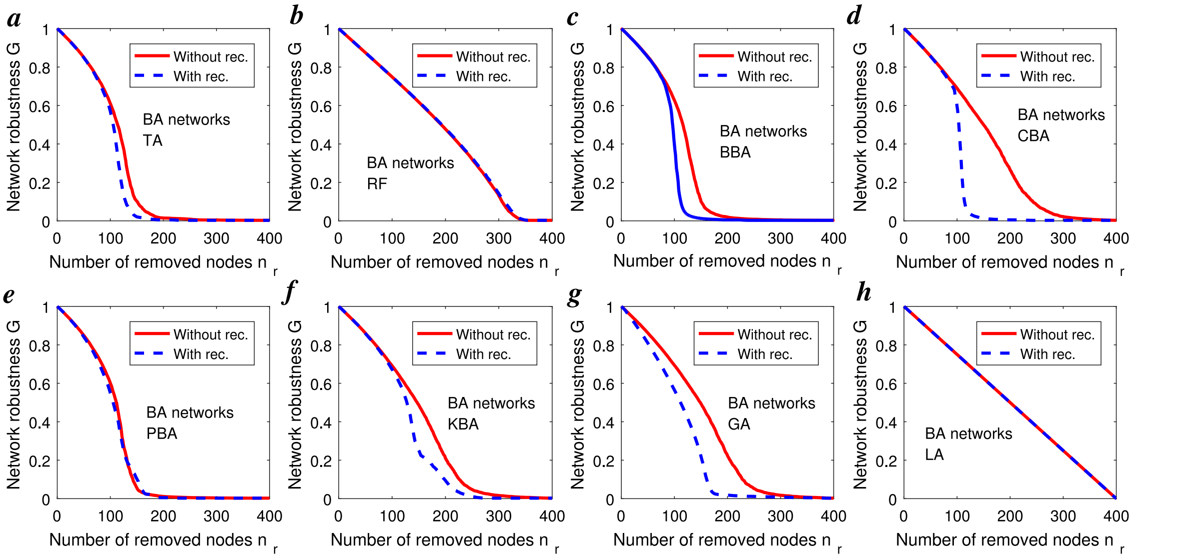

Fig. 2. (Color online). Comparisons of all methods with or without recalculations in BA networks. (a) target attack (TA); (b) random failure (RF); (c) betweenness based attack (BBA); (d) closeness based attack (CBA); (e) PageRank based attack (PBA); (f) k-shell based attack (KBA); (g) greedy algorithm (GA); (h) low-degree attack (LA).

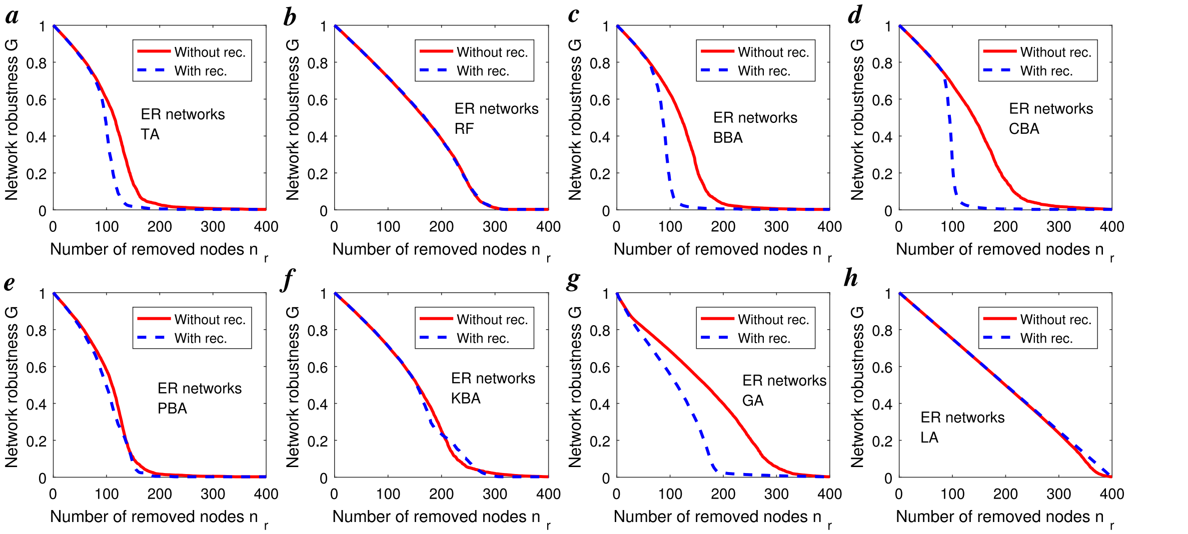

Fig. 3. (Color online). Comparisons of all methods with or without recalculations in ER networks. (a) target attack (TA); (b) random failure (RF); (c) betweenness based attack (BBA); (d) closeness based attack (CBA); (e) PageRank based attack (PBA); (f) k-shell based attack (KBA); (g) greedy algorithm (GA); (h) low-degree attack (LA).

In Fig. 4, we investigate results in different classic network models. Fig. 4(a_1)–(a_8) shows the comparison results of all methods without recalculations, while Fig. 4(b_1)–(b_8) shows the ones with recalculations. ER networks seem to be more vulnerable under RF and CBA methods with and without recalculations. Meanwhile, BA networks are vulnerable to the GA and KBA methods with or without recalculations. It is interesting to see that the BBA without recalculations has larger effect in BA networks, while ER networks are a bit vulnerable under BBA with recalculations. Under the other attack methods, the effects on network robustness in both BA and ER networks are similar.

By the comparisons, one can see that attack methods with or without recalculations might have different effects in different network models.

In Fig. 5, we investigate the evolutions of network robustness G under all attack methods in many real network models [10] without and with recalculations respectively. We find that the results in most of the real networks are very similar, and without loss of generality here we select four (*i.e grassland, yeast2, politicalblogs, airports*) of them as examples. All used real networks in this work has the high robustness against the LA method, while under most of other attack methods

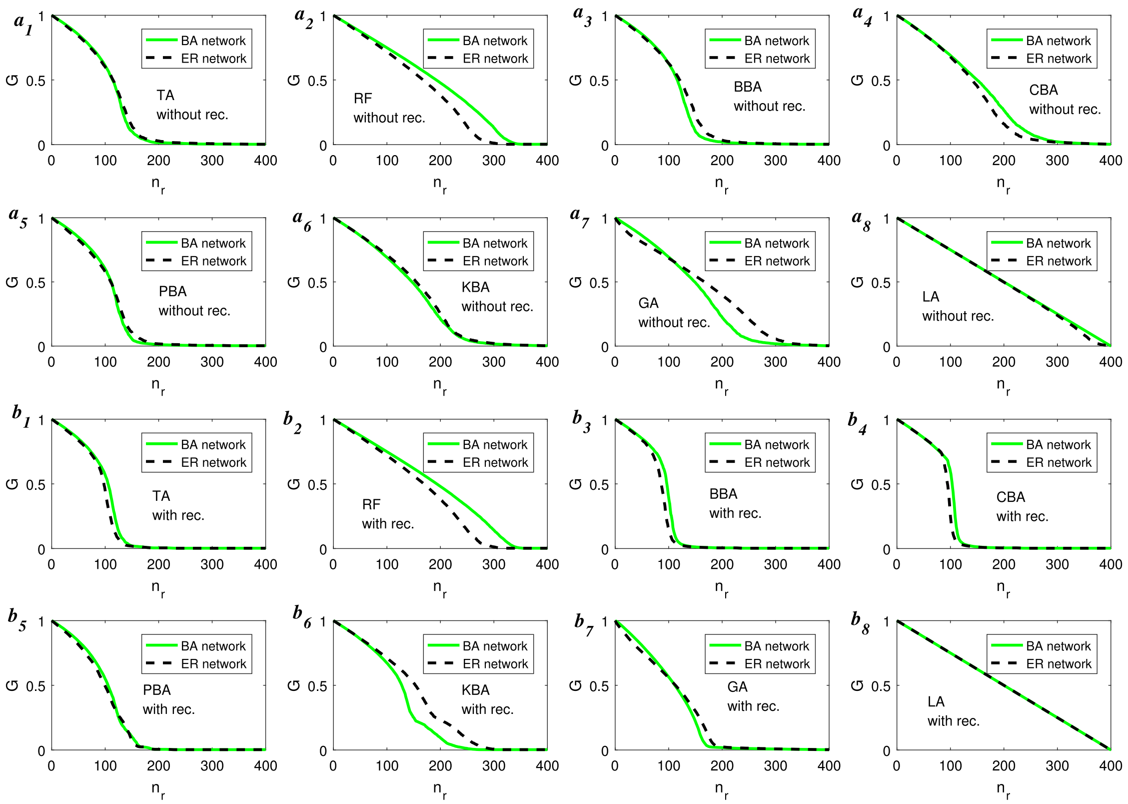

Fig. 4. (Color online). Comparisons of all methods in BA and ER networks. (a_1) TA without recalculations; (a_2) RF without recalculations; (a_3) BBA without recalculations; (a_4) CBA without recalculations; (a_5) PBA without recalculations; (a_6) KBA without recalculations; (a_7) GA without recalculations; (a_8) LA without recalculations. (b_1) TA with recalculations; (b_2) RF with recalculations; (b_3) BBA with recalculations; (b_4) CBA with recalculations; (b_5) PBA with recalculations; (b_6) KBA with recalculations; (b_7) GA with recalculations; (b_8) LA with recalculations.

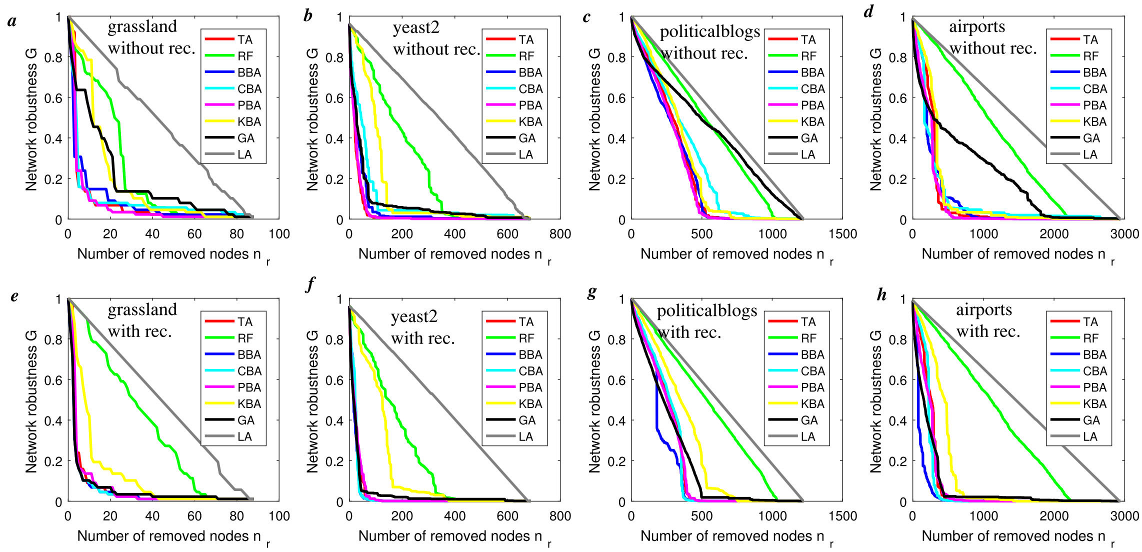

Fig. 5. (Color online). Evolutions of network robustness G under all attack methods without recalculations in many real network models.

the results are not smooth. On the whole, in Fig. 5 without recalculations the PBA seems to imperil the network structure security most, and in Fig. 5, with recalculations the BBA mechanism appears to be reduce the robustness metric G quickly especially in real network '*politicalblogs*'.

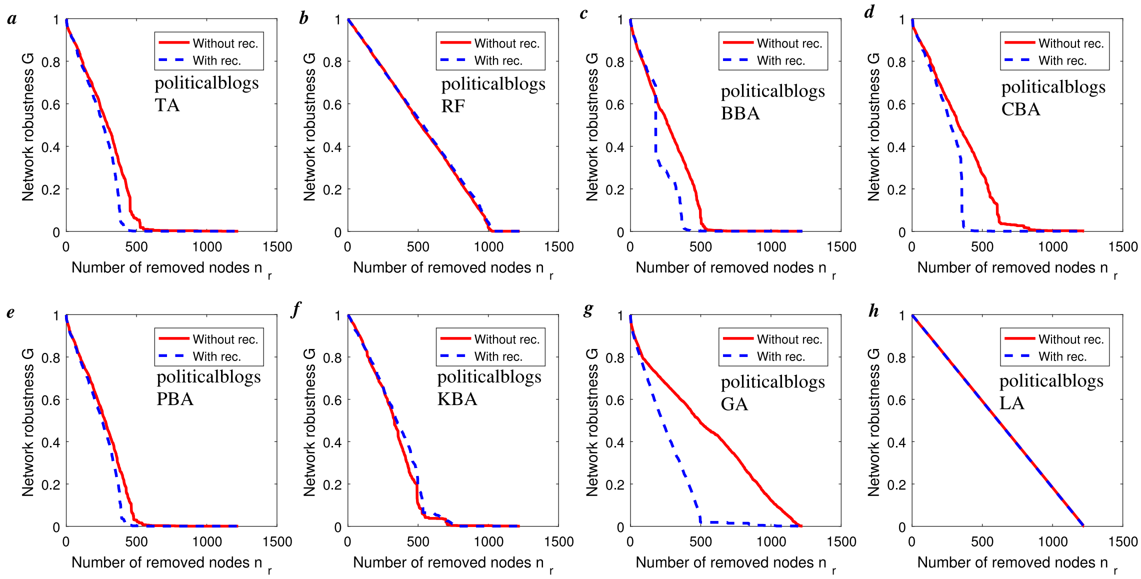

Fig. 6. (Color online). Comparisons of network robustness G under all attack methods with or without recalculations in the *politicalblogs* network. (a) TA; (b) RF; (c) BBA; (d) CBA; (e) PBA; (f) KBA; (g) GA; (h) LA.

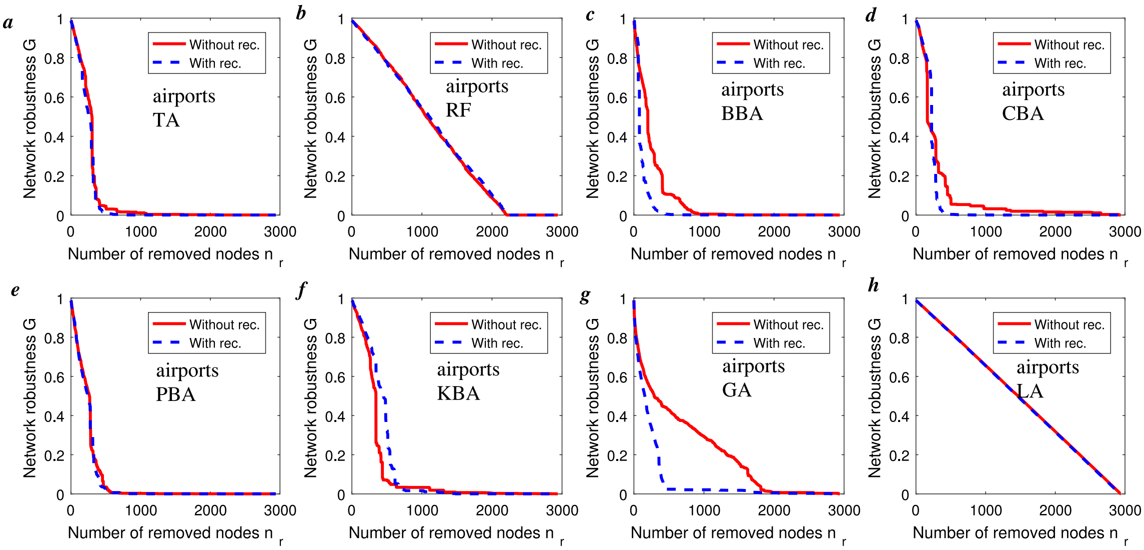

Fig. 7. (Color online). Comparisons of network robustness G under all attack methods with or without recalculations in the *airports* network. (a) TA; (b) RF; (c) BBA; (d) CBA; (e) PBA; (f) KBA; (g) GA; (h) LA.

In Figs. 6 and 7, we compare all the attack methods without and with recalculations in real network '*politicalblogs*' and '*airports*' respectively. Without recalculations, in network '*politicalblogs*' and '*airports*', these two networks are more vulnerable to the attack strategies with recalculations than without recalculations.

Therefore, the above extensive simulations in classic network models and real networks have well confirmed that most of the attack methods with recalculations have larger effects on network robustness than that without recalculations. That is to way, with recalculations a set of critical nodes can be mined more accurately. In the following part, we will discuss the key node set problem.

3.2. Minimum number of key nodes

In the above section, we mainly evaluate the evolutions of network robustness G as a function of the number of removed nodes. Here one critical problem is still open. With the removal of a fraction of nodes and their adjacent links, the network

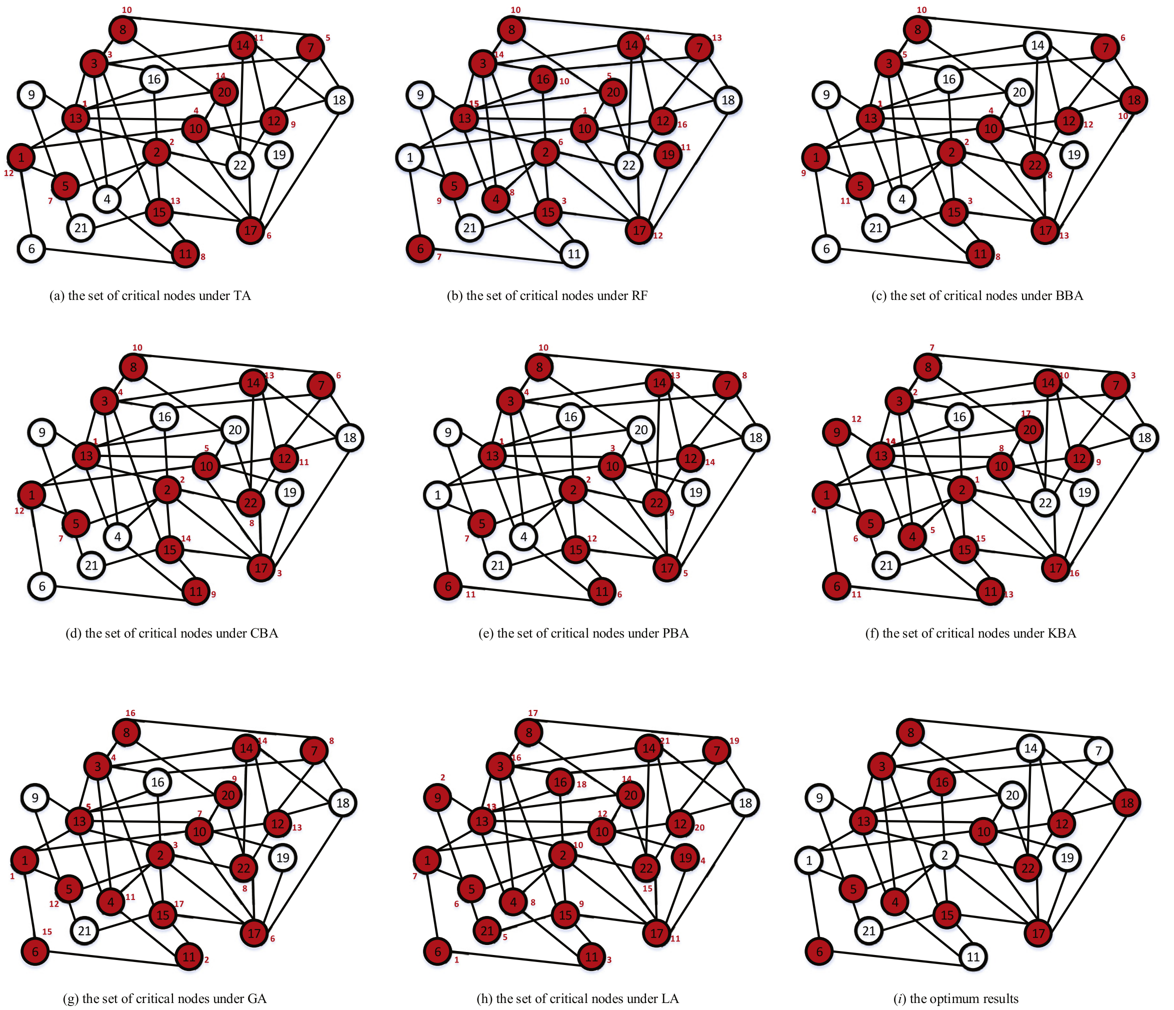

Fig. 8. An example of key node set discovery methods. (a) TA; (b) RF; (c) BBA; (d) CBA; (e) PBA; (f) KBA; (g) GA; (h) LA; (i) the optimum results . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

structure or connectivity is broken. As we known, links between nodes are the key roles in network function operations. If no links are survived, then the nodes in the network are all isolated, namely total collapse. Then under an attack strategy, at least how many nodes have to be attacked in order to destroy all the links? Here we use *minimum number of key nodes* (MNKN) to evaluate the network vulnerability from the key node set perspective.

To begin with, here we give a simple example. As shown in Fig. 8, in a simple network with 22 nodes, all 8 attack methods with recalculations and the optimum results are compared. The red nodes means the attacked nodes, and the number beside each removed node represents the selected order number under the attack method. In Fig. 8(i), the result shows that the optimal minimum number of key nodes is 13, and the results under the TA, BBA, CBA, PBA are all 14, only a bit larger than the optimal value. The node selection sequence under all methods are very different. For example, under TA the sequence is {13, 2, 3, 10, 7, 17, 5, 11, 12, 8, 14, 1, 15, 20}, under BBA {13, 2, 15, 10, 3, 7, 22, 1, 18, 5, 8, 11, 12, 17}, and under PBA {13, 2, 10, 3, 17, 11, 5, 7, 22, 8, 6, 15, 14, 12}.

The optimal MNKN is obtained by emulating all possible node sets of size from small to big. For each case, we evaluate the network connectivity. Once all nodes are isolated, the emulating process stops. The nodes in the optimal result are selected simultaneously. The optimal result for this example is {3, 4, 5, 6, 8, 10, 12, 13, 15, 16, 17, 18, 22}, in which some of the nodes have low degrees such as node 6 and 8. That is to say, key nodes in a group are not necessary all the high-degree ones.

In Fig. 9, we investigate the minimum number of key nodes under all attack methods in both BA scale-free and ER random networks. The average MNKN is the average of many networks with the same network size and average degree. In Fig. 9(a), in BA networks, among all attack mechanisms with or without recalculations, the PBA gains the lowest MNKN. Because the PageRank ranks the node importance according to the connections between nodes, it can efficiently find out the key node

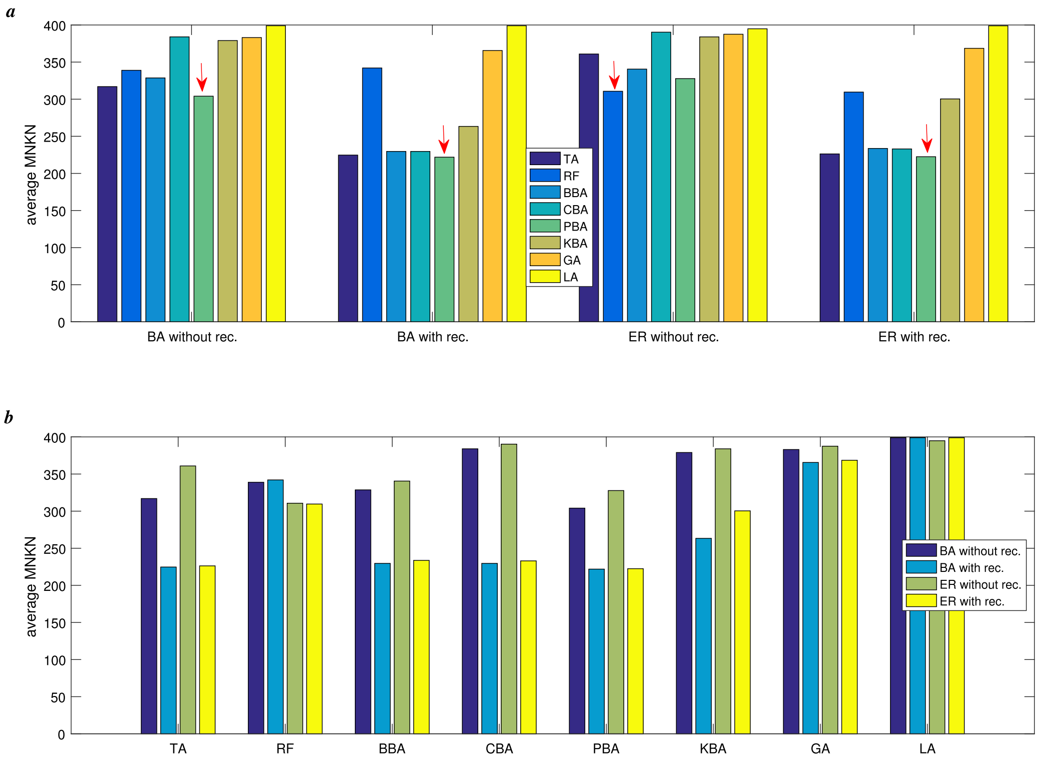

Fig. 9. (Color online). Comparisons of MNKN under all attack methods in BA and ER networks.

set which deduces the total collapse of the networks. In ER networks, among all attack methods without recalculations, it is interesting to see that the RF method can get the smallest MNKN, and this can further confirm the conclusion that the ER networks are vulnerable to random failures. With recalculations, the PBA strategy appears to have better result.

In Fig. 9(b), for every attack strategy, we compare the results in different networks models. On the whole, besides the RF and LA, other strategies all appear to have smaller MNKN with recalculation than without recalculations. Under the k-shell based attack, it is much easier to achieve small MNKN in BA network than in ER networks. Except these three, all other methods can more efficiently find out the MNKN in both BA networks and ER networks.

In the following part, we further employ all the attack methods used in this work into many real network models [10] , and analyze the results and have discussions.

In Table 1, the number in red color for each network represents the smallest MNKN under all attack strategies with and without recalculations. One can see that the smallest MNKN is gained mainly by PBA, and only two real network models (*i.e*. '*Neural*' network and '*E. coli-2*' network) gained via TA, and some networks (*e.g.* '*E. coli-2*', '*Littlerock*', and '*Ownership*') gained by both TA and PBA. It is interesting to see that without recalculations the PBA still can get the smallest MNKN in most real network models, and many networks including '*Leadership*', '*Prison*', '*St.Marks*', '*E. coli-1*', '*Ppi*' and '*Netscience*' have the minimum MNKN under RF. The BBA without recalculations also can find out better results for network '*St.Martin*', '*Ythan*', and '*Wtn61*'. On the whole, the PBA appears to be a better choice for MNKN discovery in most of real network models and classic network models.

4. Conclusions

To summarize, the network vulnerability estimation is strongly related to the selection of removed nodes. Considering different network characters, we discussed eight node attack strategies including target attack which preferentially removed the high-degree nodes, random failure randomly removing nodes, betweenness based attack which chose high-betweenness nodes, closeness based attack, PageRank based attack, k-shell based removal, greedy algorithm, and low-degree attack. The network robustness which was defined as the relative size of giant component was evaluated via extensive simulations in BA scale-free, ER random and many real networks. Influential nodes often play an important role in network connections. If an attacker or a protector want to destroy or preserve the network structure, the most critical thing is to find a set of key nodes whose removal will deduce the whole collapse of the network. We used the minimum number of key nodes as a metric to compare the network robustness against all proposed attacks with and without calculations. The results indicated that the PageRank method with recalculations could quickly break the network into all isolated pieces and find out the critical nodes' group very efficiently for most of classic and real network models.

Table 1
Comparisons of the number of critical nodes under each strategy (the number labeled by red and blue represents the best value under no recalculations (no for short) and recalculations (rec. for short) respectively)

Type	Name	N	L	TA		RF		BBA		CBA		PBA		KBA		GA		LA	
				no	rec.	no	rec.	no	rec.	no	rec.	no	rec.	no	rec.	no	rec.	no	rec.
Regulatory	TRN-Yeast-2	688	1079	521	123	416	409	517	127	684	124	234	122	683	377	679	622	678	677
Trust	Prison-inmate	67	182	61	42	54	51	60	43	63	42	58	41	63	51	62	62	65	65
	Netscience	1461	5484	1458	899	1033	1039	1457	900	1460	899	1385	899	1459	928	1458	1456	1460	1182
	Leadership	32	96	26	20	24	29	28	19	30	21	26	18	26	24	29	24	30	31
Food Web	Grassland	88	137	54	33	54	66	84	34	85	33	45	33	65	43	79	71	87	86
	Seagrass	49	226	41	34	44	43	41	34	45	33	40	33	44	37	43	37	48	48
	Littlerock	183	2476	145	81	176	168	130	83	147	82	169	81	146	113	120	96	182	182
	St.Marks	49	223	41	33	40	42	41	34	45	33	40	33	44	37	43	37	48	48
	St.Martin	45	224	39	28	38	39	36	29	39	31	39	28	39	35	39	40	44	44
	Ythan	135	597	81	57	115	104	69	59	129	59	74	57	90	114	131	106	134	134
Biologic Network	E.coli-1	99	212	92	63	76	76	86	65	95	64	87	64	93	74	95	96	97	95
	E.coli-2	418	519	416	103	257	236	416	105	416	103	186	103	415	248	416	413	414	388
	S.cerevisiae	688	1209	521	145	433	414	517	150	685	150	249	145	640	385	680	624	678	679
	Ppi	990	9374	961	590	831	847	981	604	987	597	928	588	970	630	986	980	989	955
	Neural	297	2345	259	192	263	259	260	195	284	193	256	193	259	227	278	270	295	296
Electronic Circuits	S208	122	189	112	66	87	86	108	71	117	75	84	63	116	88	115	115	120	120
	S420	252	399	234	133	188	192	226	148	244	145	175	126	242	182	239	240	250	248
	S838	512	819	478	267	372	379	449	298	495	291	355	252	494	370	487	492	510	504
World Wide Wibe	Politicalblogs	1224	19022	1106	564	1027	1042	1049	567	1223	566	863	563	1203	757	1223	1188	1221	1222
Transposition	Airports	2939	30501	2828	1102	2214	2244	2898	1110	2935	1108	2142	1092	2912	1482	2935	2875	2938	2928
Lauguage	Japanese	2704	8300	2639	602	1784	1770	2614	613	2702	609	1118	598	2698	708	2702	2642	2702	2700
BA Model Network	SF2-1	400	797	201	166	297	287	309	167	399	171	191	164	363	198	363	355	399	399
	SF2-2	400	797	201	170	292	284	276	173	397	174	196	167	384	201	384	360	398	399
	SF3-1	400	1194	230	195	316	309	280	201	372	198	226	192	377	227	377	374	399	399
	SF3-2	400	1194	247	215	322	330	307	215	384	216	246	209	373	243	373	353	399	399
	SF4-1	400	1590	272	231	351	348	337	233	384	231	269	225	398	267	398	365	399	399
	SF4-2	400	1590	263	217	340	347	289	221	381	222	254	213	379	258	379	373	398	399
Business	Ownership	141	189	133	51	90	94	132	52	139	52	80	51	133	77	137	122	140	127
	Wtn61	218	5851	159	114	199	198	135	115	159	116	157	114	166	142	217	181	217	217

As we know, finding out the optimal MNKN is a computation consuming process, and the employed eight attack methods tried to find a near optimal result. Furthermore, is there any more efficient key node mining method? Or can the result be optimized further? This is still an open problem, and we will further focus on this subject and share the research results.

This work can be applied into many real networks such as the communication network. From network security perspective, the attackers might aim to attack a set of servers, and they have to control a set of agents to extensively use up the resource of servers, resulting in security events such as the denial of service (DoS). Network service providers have to deploy their severs in suitable locations to reduce the risk of being attacked, and find out vulnerable ones which should be protected properly. For both of the attackers and protectors, finding out the set of important nodes is definitely critical. In other words, this work studied the network vulnerability or security from a very novel perspective, and the results are helpful for network structure protections and planning in the future.

Acknowledgments

The authors were grateful to the anonymous reviewers for their valuable comments and suggestions. This work was partly supported by the National Natural Science Foundation of China (No. 61502375), the Natural Science Basis Research Plan in Shaanxi Province of China (No. 2016JQ6046), the Fundamental Research Funds for the Central Universities, China (No. JB171502), the Key Program of NSFC-Guangdong Union Foundation, China (No. U1405255), the National High Technology Research and Development Program (863 Program), China (No. 2015AA016007 and 2015AA017203), the China 111 Project, China (No. B16037), and the National Key Research and Development Program of China (2016YFB0800601).

References

[1] A.E. Motter, Y. Yang, The unfolding and control of network cascades, Phys. Today 70 (1) (2017) 32–39.
[2] M.E. Newman, The structure and function of complex networks, SIAM Rev. 45 (2) (2003) 167–256.
[3] R. Albert, A.L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (1) (2002) 47.
[4] www.wechat.com.
[5] www.facebook.com.
[6] D. Chen, L. Lv, M.S. Shang, Y.C. Zhang, T. Zhou, Identifying influential nodes in complex networks, Physica A 391 (2012) 1777–1787.
[7] R. Albert, H. Jeong, A.L. Barabási, Error and attack tolerance of complex networks, Nature 406 (2000) 378.
[8] Z.Y. Jiang, Z.Q. Liu, X. He, J.F. Ma, Cascade phenomenon against subsequent failures in complex networks, Physica A 499 (2018) 472–480.
[9] J. Liu, Z.Y. Jiang, N. Kato, O. Akashi, A. Takahara, Reliability evaluation for NFV deployment of future mobile broadband networks, IEEE Wirel. Commun. 23 (3) (2016) 90–96.
[10] Y.Y. Liu, J.-J. Slotine, A.-L. Barabási, Controllability of complex networks, Nature 473 (7346) (2011) 167.
[11] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, C. Zhou, Synchronization in complex networks, Phys. Rep. 469 (3) (2008) 93–153.
[12] O. Hinz, B. Skiera, C. Barrot, J.U. Becker, Seeding strategies for viral marketing: An empirical comparison, J. Market. 75 (6) (2011) 55–71.
[13] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Comput. Netw. ISDN Syst. 30 (1–7) (1998) 107–117.
[14] L.C. Freeman, A set of measures of centrality based on betweenness, Sociometry 40 (1977) 35.
[15] L.C. Freeman, Centrality in social networks conceptual clarification, Soc. Netw. 1 (1979) 215.
[16] G. Sabidussi, The centrality index of agraph, Psychometrika 31 (1966) 581.
[17] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, E. Shir, A model of internet topology using k-shell decomposition, Proc. Natl. Acad. Sci. 104 (27) (2007) 11150–11154.
[18] R.M. Anderson, R.M. May, B. Anderson, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, USA, 1992.

[19] W. Du, B. Liang, G. Yan, O. Lordan, X. Cao, Identifying vital edges in Chinese air route network via memetic algorithm, Chin. J. Aeronaut. 30 (1) (2017) 330–336.
[20] C. Pu, S. Li, A. Michaelson, J. Yang, Iterative path attacks on networks, Phys. Lett. A 379 (2015) 1633–1638.
[21] Z. Chen, J. Wu, Y. Xia, X. Zhang, Robustness of interdependent power grids and communication networks: a complex network perspective, IEEE Trans. Circuits Syst. II: Express Briefs 99 (2017) 1.
[22] J. Wu, J. Zeng, Z. Chen, K.T. Chi, B. Chen, Effects of traffic generation patterns on the robustness of complex networks, Physica A 492 (2018) 871–877.
[23] Z.Y. Jiang, J.F. Ma, Y.L. Shen, Y. Zeng, Effects of link-orientation methods on robustness against cascading failures in complex networks, Physica A 457 (2016) 1–7.
[24] W.B. Du, X.L. Zhou, O. Lordan, Z. Wang, C. Zhao, Y.B. Zhu, Analysis of the Chinese Airline Network as multi-layer networks, Transp. Res. Part E: Logist. Transp. Rev. 89 (2016) 108–116.
[25] R.R. Liu, M. Li, C.X. Jia, Cascading failures in coupled networks: The critical role of node-coupling strength across networks, Sci. Rep. 6 (2016) 35352.
[26] R.R. Liu, D.A. Eisenberg, T.P. Seager, Y.C. Lai, The weak interdependence of infrastructure systems produces mixed percolation transitions in multilayer networks, Sci. Rep. 8 (1) (2018) 2111.
[27] X.-B. Cao, C. Hong, W.-B. Du, J. Zhang, Improving the network robustness against cascading failures by adding links, Chaos Solitons Fractals 57 (2013) 35–40.
[28] Z. Jiang, M. Liang, D. Guo, Enhancing network performance by edge addition, Int. J. Mod. Phys. C. 22 (11) (2011) 1211–1226.
[29] Z.Y. Jiang, J.F. Ma, Deployment of check-in nodes in complex networks, Sci. Rep. 7 (2017) 40428.
[30] P. Jaillet, G. Song, G. Yu, Airline network design and hub location problems, Locat. Sci. 4 (1996) 195–212.
[31] S.V. Buldyrev, R. Parshani, G. Paul, H.E. Stanley, S. Havlin, Catastrophic cascade of failures in interdependent networks, Nature 464 (2010) 1025–1028.
[32] H. Tu, Y. Xia, H.C. Iu, et al., Optimal robustness in power grids from a network science perspective, IEEE Trans. Circuits Syst. Express Briefs 1 (2018) 99.
[33] S. Kamijo, Y. Matsushita, K. Ikeuchi, M. Sakauchi, Traffic monitoring and accident detection at intersections, IEEE Trans. Intell. Transp. Syst. 1 (2000) 108–118.
[34] J.A. Dunne, R.J. Williams, N.D. Martinez, Food-web structure and network theory: the role of connectance and size, Proc. Natl. Acad. Sci. 99 (2002) 12917–12922.
[35] N. Hanaki, A. Peterhansl, P.S. Dodds, D.J. Watts, Cooperation in evolving social networks, Manage. Sci. 53 (7) (2007) 1036–1050.
[36] A.L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[37] D.J. Watts, The new science of networks, Annu. Rev. Sociol. 30 (2004) 243–270.
[38] A.E. Motter, Y.C. Lai, Cascade-based attacks on complex networks, Phys. Rev. E 66 (6) (2002) 065102(R).
[39] A. Galeotti, G. Sanjeev, Influencing the influencers: a theory of strategic diffusion, Rand J. Econ. 40 (3) (2009) 509–532.
[40] P. Erdös, A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. 5 (1960) 17.
[41] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440.