Key Nodes Mining Algorithm Based on Complex Network

Deng Ye[1,*], Wu Jun [1**], Tan Yue-Jin [1]

[1]College of Information Systems and Management, National University of Defense Technology, Changsha 410073

Keywords: Key nodes, Mining algorithm, Complex network, Measure

Abstract. The problem of network disintegration, such as suppressing the epidemic spreading and destabilizing terrorist networks, has broad applications and recently has attracted more attentions. The fundamental purpose of this study is to find the critical nodes whose deficiency could lead to the network collapse. By now, most related works of attack strategy in complex networks use certain properties of the nodes, while little is known about optimal attack strategy for complex networks. In this paper, we presented an optimized attack strategy model for complex networks and introduced the tabu search into the problem. The network performance is quantitatively measured by natural connectivity, which can be used to effectively characterize the robustness of complex networks. Numerical experiments suggest that our solution can identify the 'best' choice for node failure attack.

Introduction

Complex networks such as the Internet, metabolic networks, electric power grids, supply chains, urban road networks, and the world trade web etc., are an essential part of modern society[1-3]. Examples include among many others. Most networks are beneficial, as we get power supply from power grids and surf the internet for information. Therefore, many researchers [4-10] have focused on developing methods to enhance the robustness of such networks. In another situation by which this paper is just motivated, however, we want to disintegrate a network by attacks if it is harmful, for example, immunizing a population or a computer network to suppress the epidemic spreading or virus propagation. The immunization problem is mathematically equivalent to asking how to disintegrate a given network with a minimum number of node removals, which is very important since in most cases the number of immunization doses is limited or very expensive. Other examples of network disintegration include destabilizing terrorist networks[11], preventing financial contagion[12], controlling the rumor diffusion[13], and perturbing cancer networks [14].

Although the problem of network disintegration is not obtained more attention than the problem of network protection, some related works have been devoted to the study of the attack strategy. For example, Holme et al.[15] compared the effect of four different targeted disintegration strategies: removals by the descending order of the degree and the betweenness centrality, calculated for either the initial network or the current network during the removal procedure. It is found that the removals by the recalculated degrees and betweenness centralities are often more harmful than the attack strategies based on the initial network. The most traditional way is to select the vertices in the descending order of degrees in the initial network and then to remove vertices one by one starting from the vertex with the highest degree. As Petter summarized, we call it 'ID' attack strategy. It's significantly that 'ID' is just a local attack strategy. Then, researchers used the initial distribution of the betweenness to be global strategy and we call it 'IB'. Chen et al. [16]developed a new immunization strategy, called the "equal graph partitioning" (EGP) strategy. Schneider et al. [17]developed an immunization approach based on optimizing the susceptible size, which outperforms the best known strategy based on immunizing the highest-betweenness links or nodes. Moreover, some works have recently focused on the attack strategies with imperfect information. For example, Dezső et al. proposed a biased treatment strategy against viruses spreading based on uncertain information, in which the likelihood of identifying and administering a cure to an infected node depends on its degrees as k^{α}. Li et al. [18] studied the optimal attack problem based on incomplete information, which means that one can obtain the information of partial nodes, whereas the information is certain.

However, little is known about the optimal attack strategy for complex networks. In this paper, we focus on a global optimal attack strategy. We present an optimized attack strategy model for complex networks, and the network performance is quantitatively measured by natural connectivity. We introduce the tabu search into the network disintegration problem to identify the optimal attack strategy.

Attack Costs model for attack strategy in complex networks

Consider complex networks formalized in terms of a simple undirected graph $G(V,E)$, where V is the set of nodes, and $E \subseteq V \times V$ is the set of edges. Let $N = |V|$ and $W = |E|$ be the number of nodes and the number of edges, respectively. Let $A(G) = (a_{ij})_{N\times N}$ be the adjacency matrix of G, where $a_{ij} = a_{ji} = 1$ if nodes v_i and v_j are adjacent, and $a_{ij} = a_{ji} = 0$ otherwise. It follows that $A(G)$ is a real symmetric matrix with real eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$, which are usually called the eigenvalues of the graph G itself. The set $\{\lambda_1, \lambda_2, ..., \lambda_N\}$ is called the spectrum of G.

We only consider node attack approaches in this study and assume that the attached edges are removed if one node is attacked, i.e., if node v_i is removed, then

$$A(i,:) \leftarrow [\ ] \text{ and } A(:,i) \leftarrow [\ ]. \tag{1}$$

Denote by $\hat{V} \subseteq V$ the set of nodes that are attacked and denote by $n = |\hat{V}|$ the attack strength. Denote by $C = (x_1, x_2, ... x_N)$ an attack strategy, where $x_i = 0$ if $v_i \in \hat{V}$, otherwise $x_i = 1$. Thus we obtain $n = N - \sum_{i=1}^{N} x_i$. Our goal is to find the optimal attack strategy C^*, which maximizes the attack effect $\Phi(C)$.

There are various alternative measures of attack effect. In this paper, we consider the natural connectivity that is required to characterize the network damage due to an attack strategy. The natural connectivity can be defined as an average eigenvalue of the graph as follows.

A walk of length k in a graph G is an alternating sequence of vertices and edges $v_0 e_1 v_1 e_2 ... e_k v_k$, where $v_i \in V$ and $e_i = (v_{i-1}, v_i) \in E$. A walk is closed if $v_0 = v_k$. Closed walks are directly related to the subgraphs of a graph. For instance, a closed walk of length $k = 2$ corresponds to an edge and a closed walk of length $k = 3$ represents a triangle. The number of closed walks is an important index for complex networks. Recently, we have proposed that the number of closed walks of all lengths quantify the redundancy of alternative paths in the graph and can therefore serve as a measure of network robustness.[21, 22] Considering that shorter closed walks have more influence on the redundancy than longer closed walks, we define a weighted sum of numbers of closed walks $S = \sum_{k=0}^{\infty} n_k / k!$, where n_k is the number of closed walks of length $k > 0$ and $n_0 = N$. This scaling ensures that the weighted sum does not diverge. Denote by $n_{ij}^{(k)}$ the number of walks from v_i to v_j of length $k > 0$. Then we have $n_k = \sum_{i=1}^{N} n_{ii}^{(k)}$ .Then we obtain

$$n_k = \sum_{i=1}^{N} n_{ii}^{(k)} = trace(W^k) = \sum_{i=1}^{N} \lambda_i^{\,k}. \tag{2}$$

Therefore, we have

$$S = \sum_{k=0}^{\infty} \frac{n_k}{k!} = \sum_{k=0}^{\infty}\sum_{i=1}^{N} \frac{\lambda_i^{\,k}}{k!} = \sum_{i=1}^{N}\sum_{k=0}^{\infty} \frac{\lambda_i^{\,k}}{k!} = \sum_{i=1}^{N} e^{\lambda_i}. \tag{3}$$

Note that S will be a large number for large N, the natural connectivity is then defined as an average eigenvalue of the graph as follows[19]

$$\bar{\lambda} = \ln\left(\frac{S}{N}\right) = \ln\left(\frac{1}{N}\sum_{i=1}^{N} e^{\lambda_i}\right). \tag{4}$$

In Ref. [19], we have proved that the natural connectivity changes strictly monotonically when edges are added or deleted. Therefore, lower $\bar{\lambda}$ values correspond to more destructive attack strategies. Specifically, the optimization model of attack strategy in complex networks can be described as follows:

$$\begin{aligned} &\max\ \Phi(C = (x_1, x_2,\ \dots x_N)) \\ &s.t.\ \begin{cases} \sum_{i=1}^{N} x_i = n \\ x_i = 0\ or\ 1 \end{cases} \end{aligned} \tag{5}$$

The optimization model above is a large-scale 0-1 integer optimization problem for large N. Because the function Φ has no explicit form, we can only solve it using simulation-based optimization methods, which is very time-consuming. For the convenience of analysis, we consider utilizing an intelligent optimization algorithm for the combinatorial optimization problem.

Solutions to Optimization model for attack strategy in complex networks

In this paper, we utilize the tabu search to solve the optimization problems, which was designed to surmount the limitation of local optimality, and to iteratively move from one potential solution C to an improved solution C' in the neighborhood of C. Procedure continues to sample the unknown search space until finally terminating by the maximum number of iterations without improvement of solution. Local search procedures often become stuck in poor-scoring areas or areas where scores plateau. To avoid these pitfalls and explore regions of the search space that would be left unexplored by other local search procedures, tabu search carefully explores the neighborhood of each solution as the search progresses. Denote by $N^*(C)$ the neighborhood of the current solution. The solutions admitted to the new neighborhood, $N^*(C)$ are determined through the use of memory structures. Using memory structures, the search progresses by iteratively moving from the current solution C to an improved solution C' in $N^*(C)$. These memory structures form what is known as a "tabu list", a set of rules and banned solutions that are used to filter the solutions admitted to the neighborhood $N^*(C)$ to be explored by the search [20, 21]. Each step of the OAS(Optimal Attack Strategy) is described below:

Procedure 1:（initialization parameters） Set length L of the tabu list at 10, define the number $n_{candidate}=100$ of neighbors of the current solution investigated at each iteration. Set the maximum number of successive iterations without any improvement of the objective function value at 1000, and start with the tabu list empty.

Procedure 2: Generate the initial solution C_0: Select a starting solution by generating an attack strategy $C=(\overbrace{1,\ 1,\ 1,\ ......1,\ 1}^{N})$, which means that every node is coded to be an element x_i of the vector and set at a value of 1. Then choose n nodes to randomly remove to obtain the initial solution C_0. Record the current best known solution by setting $C_{best}=C_0$ and the current solution $C_{now}=C_0$. Then calculate the natural connectivity $\bar{\lambda}_{now}$ of the current solution and set $\bar{\lambda}_{best}=\bar{\lambda}_{now}$.

Procedure 3: Determine whether the objective function is reached the maximum number of successive iterations without any improvement of the objective function value. If reached, output the result. If not reached, continue to the next step.

Procedure 4: Generate neighbors (swap operation): Change of a selected variable x_j from 0 to 1, and change of a selected variable x_k from 1 to 0.Utilize the combined swap to obtain $n_{candidate}$ neighbors and record their natural connectivity.

Procedure 5: Select a solution $C_{candidate}$ from neighbors to minimize the objective function $\bar{\lambda}$, then check whether the corresponding swap of $C_{candidate}$ is on the tabu list. If not, record the current solution $C_{now}=C_{candidate}$. Besides, though the corresponding swap is on the tabu list, it satisfies the aspiration criterion that tabu can be disregarded. In concrete terms, it means that of a swap has been tabooed, however, the corresponding objective function is better than the current value of $\bar{\lambda}_{now}$. Then the tabooed swap would be released and record the current solution $C_{now}=C_{candidate}$. Otherwise, select a solution with second optimal value of objective function which has not been tabooed to be C_{now}. Add the serial numbers of two nodes which are belonged to 'best' swap to be an element of tabu list. Then update the tabu list. Most notably, all the elements in the tabu list are abandoned in a limited number of iterations and the taboo number of each tabu "swap" eliminates once after each iteration. Next, if the element has been tabooed for L iterations, such element will be allowed to expire from the tabu list. At next moment, if $\bar{\lambda}_{now}<\bar{\lambda}_{C_{best}}$, then let $C_{best}=C_{now}$.

Procedure 6: Turn to step 3.

To prove the feasibility and efficiency of the OAS, numerical experimentswere carried out by using the OAS, degree-based attack strategyand betweenness-based attack strategyboth in model and real-world complex networks. The nodes were removed according to our optimal attack strategy and according to the criterion defined by other attack strategies,We use the natural connectivity, which is the fraction of nodes into the giant component, as the measure function Γ of network performance. The technique for computing the $\bar{\lambda}$ is just as Equation 4. We here consider two model networks: ER random networks and CM random scale-free networks.

The ER random network model[22]generates a graph by connecting nodes randomly. Each edge is included in the graph with a probability p independent from every other edge. The parameter p in this model can be thought of as a weighting function; as p increases from 0 to 1, the model becomes more likely to include graphs with more edges and less likely to include graphs with fewer edges.

The CM random scale-free network[2, 23]is defined as follows. Consider a large network with a degree distribution $P(k) \sim k^{-\gamma}$, such that $P(k)$ is the proportion of nodes in the network having a degree of k. We choose a degree sequence, which is a set of N values of the degrees k_i of vertices $i = 1...N$, from this distribution. We generated the degree sequence of the nodes by randomly drawing N values $k_1,...,k_N$ from the degree distribution. Then, for each node i, we assigned a link with node j with probability $P(k_i)P(k_j)$.

In fig.2, results of this experiment are displayed for two model networks, respectively. As indicated in the following figure, the OAS can be much more destructive than the degree-based attack strategyand betweenness-based attack strategy.

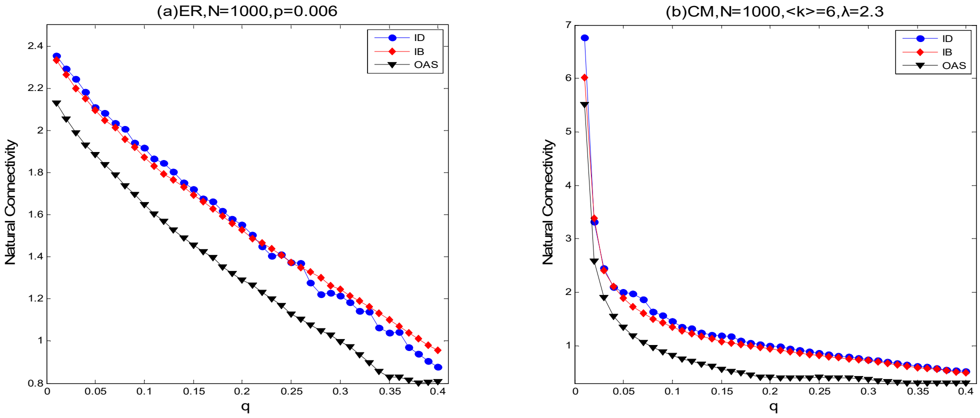

Fig. 2.Value of natural connectivity of the OAS attack strategy and other attack strategies versus the fraction q of nodes removed in the ER random graph (a), CM random scale-free network (b).

The study of disintegration is important for many real-world systems such as rumor spreading in online social networks, disease transmission through airlines and foodweb. To evaluate the performance of our method, we investigate a real-world network: the network of the US air transportation system (USAir)(http://toreopsahl.com/datasets/usairports)，basic statistics of these networks are shown in Table 1.

Table 1.Basic statistics of three real networks.

N and W are the number of nodes and links. $\langle k \rangle$ is the average degree. C is the clustering coefficient. r is the assortativity. $\langle l \rangle$ is the average shortest distance.

Networks	N	W	$\langle k \rangle$	C	r	$\langle l \rangle$
USair	1574	28236	35.8	0.384	-0.113	3.14

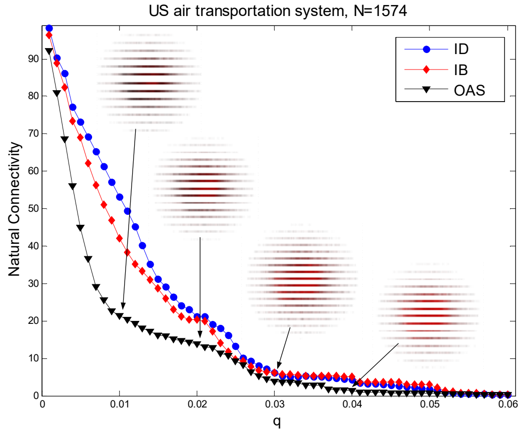

Fig. 3.Value of natural connectivity of the OAS attack strategy and other attack strategies versus the fraction q of nodes removed from the US air transportation system, the four sub-pictures are corresponding to the network states after removal of nodes in which attacked nodes and edges are in red.

Fig.3 presents the destructive effect among OAS, betweenness-based attack strategy and degree-based attack strategy for the US air transportation system. Significantly, based on the simulation results, US-air crashed dramatically with the OAS when about 10% of all the network nodes were taken down. As shown in fig.3, although the destructive effect of IB and ID attack almost approached the OAS, the efficiencies was still more destructive than other attack strategy.

Summary

Based on the simulation results, it has been proved that the OAS could efficiently find the near-optimal solution for the network disintegration problem, which allows the optimal attack combination that yields the strongest destructive effect to be obtained. In this paper, we present an optimized attack strategy model for complex networks and introduce the tabu search into the problem. The network performance is quantitatively measured by natural connectivity, which can be used to efficiently characterize the robustness of complex networks. Numerical experiments suggest that our solution can identify the 'best' choice for node failure attack. As presented here, we can approximate the 'best' choice for nodes failure attacks through a global search.

In addition, this study can help improve the robustness of various networks, as well as shed some light on protection strategies of modern infrastructure networks due to the importance of nodes chosen by OAS which could make the network crash rapidly.

Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant Nos 71031007 and 71171195, and the Program for New Century Excellent Talents in University under Grant No NCET-12-0141.

References

[1]. Albert R, Barabási AL. Statistical mechanics of complex networks [J]. Rev Mod Phys, 2002, 74 (1):47-51

[2]. Newman MEJ. The structure and function of complex networks [J]. SIAM Rev, 2003, 45 (2):167-256

[3]. Wang XF. Complex networks: Topology, dynamics and synchronization [J]. Int J Bifurcation &Chaos, 2002, 12 (5):885-916

[4]. Shargel B, Sayama H, Epstein IR, Bar-Yam Y. Optimization of robustness and connectivity in complex networks [J]. Phys Rev Lett, 2003, 90 (6):068701

[5]. Paul G, Tanizawa T, Havlin S, Stanley HE. Optimization of robustness of complex networks [J]. Eur Phys J B, 2004, 38 (2):187-91

[6]. Valente AXCN, Sarkar A, Stone HA. Two-peak and three-peak optimal complex networks [J]. Phys Rev Lett, 2004, 92 (11):118702

[7]. Liu JG, Wang ZT, Dang YZ. Optimization of robustness of scale-free network to random and targeted attacks [J]. Mod Phys Lett, 2005, 19 (16):785-92

[8]. Tanizawa T, Paul G, Cohen R, Havlin S, Stanley HE. Optimization of network robustness to waves of targeted and random attacks [J]. Phys Rev E, 2005, 71 (4):047101

[9]. Beygelzimer A, Grinstein GE, Linsker R, Rish I. Improving network robustness by edge modification [J]. Physica A, 2005, 357 (3-4):593-612

[10]. Schneider CM, Moreira AA, Andrade JS, Havlin S, Herrmann HJ. Mitigation of malicious attacks on networks [J]. Proc Natl Acad Sci U S A, 2011, 108 (10):3838-41

[11]. Raab J, Milward HB. Dark networks as problems [J]. J Public Adm Res Theory, 2003, 13 (4):413-39

[12]. Kobayashi T, Hasui K. Efficient immunization strategies to prevent financial contagion [J]. Scientific reports, 2014, 4

[13]. Tripathy RM, Bagchi A, Mehta S. Proceedings of the 19th ACM international conference on Information and knowledge management [C]. Toronto, Canada: ACM, 2010.

[14]. Quayle AP, Siddiqui AS, Jones SJM. Preferential network perturbation [J]. Physica A, 2006, 371 823-40

[15]. Holme P, Kim BJ, Yoon CN, Han SK. Attack vulnerability of complex networks [J]. Phys Rev E, 2002, 65 (5):056109

[16]. Chen Y, Paul G, Havlin S, Liljeros F, Stanley HE. Finding a better immunization strategy [J]. Physical review letters, 2008, 101 (5):058701

[17]. Schneider CM, Mihaljev T, Havlin S, Herrmann HJ. Suppressing epidemics with a limited amount of immunization units [J]. Physical Review E, 2011, 84 (6):061911

[18]. Li J, Wu J, Li Y, Deng H-Z, Tan Y-J. Optimal Attack Strategy in Random Scale-Free Networks Based on Incomplete Information [J]. Chinese Physics Letters, 2011, 28 (6):068902

[19]. Wu J, Barahona M, Tan YJ, Deng HZ. Natural Connectivity of Complex Networks [J]. Chinese Physics Letters, 2010, 27 (7):078902

[20]. Glover F. Tabu search-part I [J]. ORSA Journal on computing, 1989, 1 (3):190-206

[21]. Glover F. Tabu search—part II [J]. ORSA Journal on computing, 1990, 2 (1):4-32

[22]. Erdos P, Rényi A. On the evolution of random graphs [J]. Bull Inst Internat Statist, 1961, 38 (4):343-7

[23]. Albert R, Barabási A-L. Statistical mechanics of complex networks [J]. Reviews of modern physics, 2002, 74 (1):47