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 [a,b,*], Yong Zeng [a,b], Zhi-Hong Liu [a,b], Jian-Feng Ma [a,b]

[a] School of Cyber Engineering, Xidian University, Xi'an, Shaanxi 710071, China
[b] 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 xxxx

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/© 2018 Elsevier B.V. All rights reserved.

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*;