SCIENCE CHINA
Information Sciences

• LETTER •

March 2020, Vol. 63 139101:1–139101:3
https://doi.org/10.1007/s11432-018-9855-7

An eigenvalue-based immunization scheme for node attacks in networks with uncertainty

Yizhi REN[1], Mengjin JIANG[2], Ting WU[1], Ye YAO[1], Kim-Kwang Raymond CHOO[3] & Zhen WANG[1*]

[1]*School of Cyberspace, Hangzhou Dianzi University, Hangzhou* 310018*, China;*
[2]*School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou* 310018*, China;*
[3]*Department of Information Systems and Cyber Security, The University of Texas at San Antonio, San Antonio* TX 78249*, USA*

Received 23 December 2018/Accepted 2 April 2019/Published online 10 February 2020

Citation Ren Y Z, Jiang M J, Wu T, et al. An eigenvalue-based immunization scheme for node attacks in networks with uncertainty. Sci China Inf Sci, 2020, 63(3): 139101, https://doi.org/10.1007/s11432-018-9855-7

Dear editor,

Controlling the propagation through immunization has applications in a number of fields [1]. A key facilitator for malware and attacks dissemination is the interconnectivity between networks, systems and devices. Therefore, our essential is to 'break' the interconnectivity structure of networks through immunized nodes, in order to make the remaining networks more resilient to external attacks [2]. An increasingly large number of research in complex networks (e.g., wireless sensor networks, peer-to-peer networks) are focusing on developing efficient and robust security mechanisms to protect them from malicious attacks. While most of them have been dedicated to designing immunization strategies to prevent attacks in deterministic networks, there are other factors led to uncertain networks we need to consider in real-world applications [3]. To deal with those problems, we propose an eigenvalue-based node immunization scheme that is designed to prevent malware attacks or viruses (these two terms will be used interchangeably in this study) from spreading in real-world networks.

System modeling. We abstract an undirected network with uncertainty as a probabilistic graph (or a uncertainty graph) $G = (V, E, P)$, where V is a set of vertices and E is a set of edges. P is a set of real number within $[0, 1]$ and it means the probability that edges exist. If edge $(u, v) \notin E$, we set $p_{u,v} = 0$, and if $p_{u,v} = 1$, we can say that it is a deterministic edge and it exists in any situation. Each edge (u, v) in E is independent and with the existence of probability $p_{u,v}$. Here, a uncertainty graph can be regarded as the basis in many deterministic graphs. For example, let $\tilde{G} = (V, \tilde{E})$ be the sample graph of G, we denote it as $\tilde{G} \subseteq G$. Given $\tilde{G}$ and G, we always have $\tilde{E} \subseteq E$. If G has m uncertainty edges in all (i.e., $|E| = m$), then we have 2^m sample graphs [4]. The probability of observing any sample graph $\tilde{G}_i$ is

$$\Pr(\tilde{G}_i) = \prod_{(u,v)\in\tilde{E}_i} p_{u,v} \prod_{(u,v)\notin\tilde{E}_i} (1 - p_{u,v}). \quad (1)$$

Expected eigenvalue (EE). As we known, the largest eigenvalue λ is closely related to the connectivity of networks, and the larger λ means the tighter connection of network. The malicious attacks or viruses can spread from one node to other neighboring nodes through the connections. In an effort to reduce the impact of node attacks, we will adopt the measure of the largest eigenvalue which reflects the vulnerability of networks [5].

Here, we introduce the concept of expected eigenvalue to be the eigenvalue of uncertainty networks. The expected eigenvalue of the uncertainty

* Corresponding author (email: wangzhen@hdu.edu.cn)

network G can be denoted as

$$\mathrm{EE}(G) = \sum_{j=1}^{N} \Pr(\tilde{G}_j)\lambda(\tilde{G}_j), \tag{2}$$

where $\tilde{G}_j$ is a sample network of G, N is the number of sample networks of G, and $\lambda(\tilde{G}_j)$ is the largest eigenvalue of $\tilde{G}_j$.

Fraction of infected nodes (FI). In SIS (susceptible-infective-susceptible) epidemic model, let β and σ be the infection rate and recovery rate of virus, respectively, and τ be the epidemic threshold, which plays a key role in the design of immune strategies. Since if $\beta/\sigma < \tau$, the infection will die out over time, while $\beta/\sigma > \tau$, the infection will survive and become an epidemic. The largest eigenvalue λ is closely related to the epidemic threshold τ, where $\tau = 1/\lambda$ [5]. Therefore, we can use λ to explore the immune strategy to judge the importance of nodes.

Let $\rho_{i,t}$ be the probability that node i will not receive infection from its neighbors at time t, then we have

$$\rho_{i,t} = \prod_{i \in N(i)} (1 - \beta\psi_{i,t-1}), \tag{3}$$

where $N(i)$ is the neighbor set of i, $\psi_{i,t}$ is the infection probability of i at time t, which can be formulated as

$$\psi_{i,t} = 1 - (1 - \psi_{i,t-1})\rho_{i,t} - \sigma\psi_{i,t-1}\rho_{i,t}. \tag{4}$$

In our model, we first raised the issue of studying the spread of outbreaks in uncertain networks. As we known, the spread of an outbreak is due to the transmission of the epidemic from one node to its neighbors, and it depends entirely on whether there is a connection between the nodes. In uncertain networks, whether the edge exists is determined by its probability. Therefore, we need to combine the network structure of different sample networks to design solutions. For a sample network $\tilde{G}_i$, the fraction of infected nodes can be calculated as follows:

$$\mathrm{FI}(\tilde{G}_i, t) = \frac{\sum_{j=1}^{n} \psi_{i,t}}{n - k}, \tag{5}$$

where n is the number of nodes in $\tilde{G}_i$, $\sum_{j=1}^{n} \psi_{j,t}$ is the number of infected node at time t, and $n - k$ refers to the number of nodes remaining after removing k nodes. For the purposes of discussion, all nodes are initialized to be infection status in simulation section.

Representative instance. Most problems for large-scale uncertain graphs are very expensive, therefore, we propose to extract a representative instance $G^* \sqsubseteq G$ instead of all samples $\tilde{G}_i$ ($i = 1, 2, \ldots, N$). Then, the deterministic algorithm on G^* can effectively deal with the problem on uncertain network G. To ensure accuracy, the representation instance G^* should retain the underlying structure of the uncertain networks. Node degree is one of the most basic properties of the graph structure. By keeping the degree of each node, we can capture the essence of uncertain graph and accurately approximate other properties.

The criterion for extracting a representative instance is to preserve the expected degree of each node. So, our goal is to find a representative instance G^* where the degree of node i in G^* is as close as possible to the expected degree of node i in G. The expected degree of node i in G is the sum of the probability of all links of i. It can be expressed as

$$\deg(i, G) = \sum_{(i,j) \in E, j \in N(i)} p_{i,j}. \tag{6}$$

Let $\mathrm{dis}(i, \tilde{G}_j)$ be the difference between the degree of node i in sample network $\tilde{G}_j$ and the expected degree of i in G, we have

$$\mathrm{dis}(i, \tilde{G}_j) = \deg_{\tilde{G}_j}(i) - \deg_G(i). \tag{7}$$

Then, the total discrepancy of $\tilde{G}_j$ can be expressed as

$$\mathrm{dis}(\tilde{G}_j) = \sum_{i \in V} |\mathrm{dis}(i, \tilde{G}_j)|. \tag{8}$$

Our goal is to get a representative instance G^* which satisfies

$$G^* = \mathrm{argmax}_{\tilde{G}_j \sqsubseteq G} \mathrm{dis}_{\tilde{G}_j}. \tag{9}$$

Simulation. The comparison results of our proposed graphCom and deepGreedy algorithms with degree centrality (Appendix A), betweenness centrality and netshield [6] algorithms are shown in Figure 1. To evaluate the feasibility and effectiveness of our approach, we verified it from simulation network to actual network, from small network to large network. Since the epidemic threshold in a deterministic graph satisfies $\tau = \frac{1}{\lambda}$. Here, let s be the normalized virus strength, we have $s = \lambda\frac{\beta}{\sigma}$ where β and σ are the infection rate and the recovery rate, respectively.

The network density usually determines whether the internodes are closely connected, and it can be represents as the ratio of the number of edges to the maximum number of edges of the network. And that means high degree nodes will be more vulnerable to infection. Here, the factors

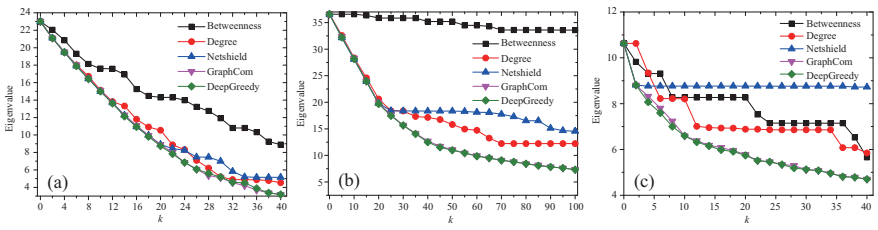

Figure 1 (Color online) Largest eigenvalue vs. the number of immune nodes k in large network. The number of uncertain edges in all networks exceeds 2000. (a) USAir; (b) yeast; (c) router.

of network centrality, eigenvalue and its corresponding eigenvector of network are fully taken into consideration. We introduce representative score (RS) to measure the importance of nodes, and the representative score $\mathrm{RS}(j, G^*)$ of j in G^* can be denoted as

$$\mathrm{RS}(j, G^*) = \begin{cases} 0, & \deg(j, G^*) = 0, \\ |\frac{\theta \deg(j, G^*)}{\mathrm{den}(G^*)}|, & j \in V,\ \mathrm{den}(G^*) \neq 0, \end{cases} \tag{10}$$

where $\deg(j, G^*)$ is the node degree of i, and θ is the corresponding eigenvector of the largest eigenvalue. $\mathrm{den}(G^*)$ is the density of G^*, which can be written as

$$\mathrm{den}(G^*) = \frac{2|E|}{n(n-1)}. \tag{11}$$

Many studies have been done to remove uncertainty by selecting representative instances from uncertain networks. For example, average degree rewiring (ADR) and approximate b-matching (ABM) algorithms are proposed to capture the underlying properties of probabilistic networks well. In this study, we propose a new average degree rewiring (NADR) for eliminating the repeatability and instability caused by random selection, which by evaluating the best selectable edges.

In a probabilistic graph, the probability of each sample network $\Pr(\tilde{G}_i)$ is determined by the number of edges. However, when the number of uncertain edges exceeds 2000, the result will default to 0 based on the requirement of calculation accuracy. Therefore, for larger networks, sampling to compute EE is not feasible. It is sensible to select representative instance to remove uncertainty in place of sampling. The comparison of graphCom and other algorithms is based on the representative instance extracted from the uncertain network. The details are shown in Figure 1. With the increase of the number of immune nodes k, the decline trend of largest eigenvalue is found to suggest that the proposed algorithm has a significant advantage.

Conclusion. This study provides the following key contributions: (1) We formally define and model immunization problem in networks with uncertainty. (2) We adopt the EE and FI security indicators to reflect the strength and the effect of attacks, respectively. (3) To address the complexity problem of sample-based network representation methods when the scale of the uncertainty network is large, we remove the network uncertainty by selecting a representative instance, and the representative instance maximizes the retention of the underlying properties of the probabilistic network. (4) We propose graphCom and deepGreedy algorithms, based on the feature of reducing the largest eigenvalue in the network, to minimize the proposed measure EE and FI. The effectiveness of our method is demonstrated by comparing with related algorithms in both simulation and real-world networks.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 61872120) and Natural Science Foundation of Zhejiang Province (Grant Nos. LY18F020017, LY18F030007).

Supporting information Appendix A. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Roy S, Xue M, Das S K. Security and discoverability of spread dynamics in cyber-physical networks. IEEE Trans Parallel Distrib Syst, 2012, 23: 1694–1707
2 Shang Y. False positive and false negative effects on network attacks. J Stat Phys, 2018, 170: 141–164
3 Dinh T N, Thai M T. Network under joint node and link attacks: vulnerability assessment methods and analysis. IEEE/ACM Trans Netw, 2015, 23: 1001–1011
4 Ren Y, Jiang M, Yao Y, et al. Node immunization in networks with uncertainty. In: Proceedings of IEEE International Conference on Trust, Security And Privacy in Computing and Communications, 2018. 1392–1397
5 Chakrabarti D, Wang Y, Wang C, et al. Epidemic thresholds in real networks. ACM Trans Inf Syst Secur, 2008, 10: 1–26
6 Chen C, Tong H, Prakash B A, et al. Node immunization on large graphs: theory and algorithms. IEEE Trans Knowl Data Eng, 2016, 28: 113–126