IEICE TRANS. INF. & SYST., VOL.E104–D, NO.7 JULY 2021

LETTER

Correlation of Centralities: A Study through Distinct Graph Robustness

Xin-Ling GUO[†], ***Nonmember***, **Zhe-Ming LU**[†a)], ***Member***, ***and*** **Yi-Jia ZHANG**[††b)], ***Nonmember***

SUMMARY Robustness of complex networks is an essential subject for improving their performance when vertices or links are removed due to potential threats. In recent years, significant advancements have been achieved in this field by many researchers. In this paper we show an overview from a novel statistic perspective. We present a brief review about complex networks at first including 2 primary network models, 12 popular attack strategies and the most convincing network robustness metrics. Then, we focus on the correlations of 12 attack strategies with each other, and the difference of the correlations from one network model to the other. We are also curious about the robustness of networks when vertices are removed according to different attack strategies and the difference of robustness from one network model to the other. Our aim is to observe the correlation mechanism of centralities for distinct network models, and compare the network robustness when different centralities are applied as attacking directors to distinct network models. What inspires us is that maybe we can find a paradigm that combines several high-destructive attack strategies to find the optimal strategy based on the deep learning framework.

key words: *scale-free network, random network, robustness, correlation analysis, attack strategies*

1. Introduction

The robustness of a network upon perturbation has significantly gained a lot of interests in recent years. Albert et al. [1] first considered two types of perturbations, one is to remove vertices randomly which is called random attack, and the other is to delete the most connected node which is called intentional attack. They demonstrated that scale-free networks present a surprisingly high tolerance against random errors while a low tolerance against intentional attacks, which inspired abundant studies on the robustness of networks [2]–[4]. Magoni [4] measured a few general intentional attack strategies for complex networks. Newman [5] proposed the flow betweenness centrality. Kermarreca et al. [6] proposed the second-order centrality. Most network security problems can be regarded as the invulnerability of the network, which occurs in many real-world networks. It is significant for us to study the robustness of complex networks. To evaluate the network robustness, the centrality metrics is the vital research topic.

However, few researchers consider the correlations between attack strategies and the difference of the correlations from one network model to another. The contributions of this paper are to provide answers to the following questions. First, if the well-known centralities are correlated with each other? If they are, how well they are correlated for different network models? Second, if these correlations are significant? In this paper, we investigate the correlations of each pair of centralities for different network models from a statistic perspective. Third, if highly correlated centralities give rise to the same level of destruction on the network connection? We also concern about the difference of robustness performance on networks when highly correlated attack strategies are applied to remove vertices.

The remainder of this paper is organized as follows. In Sect. 2, we briefly present two popular network models, and 12 commonly adopted centrality measures. In Sect. 3, we introduce the principal component analysis (PCA) to quantify the correlation between these 12 centralities. In Sect. 4, we present the robustness measures adopted in this paper. In Sect. 5, we present the results of the correlations among centralities and the simulation results of network robustness. Finally, we conclude and discuss the results in Sect. 6.

2. Network Models and Centrality Metrics

Before introducing the network models, we should know that an undirected network is given by a pair $G = \{V, E\}$, where $V = \{1, 2, \ldots, N\}$ is the set of vertices and $E \subseteq V \times V$ is the set of edges. The adjacency matrix $\boldsymbol{A}$, where the element A_{ij} equals 1, if vertices i and j are connected, 0 if they are not. The set of neighbors of vertex i is defined as $V_i = \{j \in V: (i, j) \in E\}$, the degree of vertex i is denoted by $k_i = |V_i|$, and the average network degree is denoted by $\langle k \rangle$.

We consider two network models in this paper. One is the Erdös-Rényi (ER) random graph [7], the other is the Barabasi-Albert (BA) scale-free network [8]. And 12 centralities are present in this paper. In the ER random graph model [7], vertices are randomly connected with each other by the same probability p. The ER random graph does not present any community structure and its degree distribution follows a Poisson distribution. For the BA scale-free network [8]. It starts with k fully connected vertices and keeps adding new vertices with k connections. The probability of a vertex p_i receiving a new connection takes the degree of the vertex divided by the sum of degree over all vertices into consideration. The summary of characteristics of above two networks is shown in Table 1.

Manuscript received December 19, 2020.
Manuscript revised February 17, 2021.
Manuscript publicized April 5, 2021.

[†]The authors are with School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China.
[††]The author is with School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China.
a) E-mail: zheminglu@zju.edu.cn
b) E-mail: waiting@zstu.edu.cn (Corresponding author)

DOI: 10.1587/transinf.2020EDL8163

Table 1 The summary of network models.

Network Model	Degree Distribution	Assortativity	Number of Parameters
ER	Possion	positive	2
BA	Power Law	positive	3

Table 2 The summary of centrality metrics.

Class	Centrality	Formulae
Degree Centrality	DC[10]	$C_D(v) = k_v/(N-1)$ k_v: the degree of vertex v.
Path Centralities	BC[11]	$C_B(v) = \sum_{s=1}^{N}\sum_{t=s+1}^{N}\frac{g_{st}(v)}{g_{st}}$ g_{st}: number of the shortest paths from s to t. $g_{st}(v)$: the number of paths that pass through v.
	SOC[6]	$\sigma(v) = \sqrt{2\sum_{u\in S}M(v,u) - \lvert S\rvert(\lvert S\rvert+1)}$ $M(v,u)$: the expected time starting from state v to u for the first time. S: the finite state space.
	CFBC[12]	$C_{CFB}(v) = \sum_{s\neq t\in V} I_v^{(st)}/\frac{1}{2}N(N-1)$ $I_v^{(st)}$: current flow through vertex v between source s and sink t.
	LoadC[13]	$C_{Load}(v) = \frac{2}{N(N-1)}\sum_{s,t\in V}\theta_{st}(v)$ $\theta_{st}(v)$ = the overall commodity forwarded by vertex v.
Proximity Centralities	CC[6]	$C_C(v) = (N-1)/\left[\sum_{u=1,u\neq v}^{N} d_{vu}\right]$ d_{vu}: the shortest distance between vertices v and u.
	CFCC[12]	$C_{CFC}(v) = N/\left[\sum_{u=1,u\neq v}^{N}(v_{vu}(v) - v_{vu}(u))\right]$ $v_{st}(v)$: the voltage of v when a unit current enters the network at s and leaves it at t.
	HarmC[14]	$C_{Harm}(v) = 1/\sum_{u=1,u\neq v}^{N} d_{vu}$ d_{vu}: the shortest distance between vertices v and u.
	RC[15]	$C_R(v) = 1 + \sum_{x=1}^{k} r_{vx}/x$ r_{vx}: the number of vertices at x hop distance from v.
Spectral Centralities	EVC[16]	$C_{EV}(v) = \alpha(\boldsymbol{A}^T C_{EV})_v + \epsilon$ ϵ: works as a damping factor. α: weighs the relative importance of the peers versus that of the node itself.
	PRC[17]	$C_{PR}(v) = \frac{q}{N} + (1-q)\sum_{j:j\to v} C_{PR}(j)/k_j$ q: weighs the mixture between random walk and random jump.
	SubGC[6]	E_u: an eigenvector of the adjacency matrix $\boldsymbol{A}$ corresponding to the eigenvalue λ_j.

According to the purpose and conceptualization, we can classify the centrality measures into four classes [9]: degree centralities, path centralities, proximity centralities and spectral centralities. Degree centralities evaluate the importance of vertices by the number of neighbors. Different from the degree centralities, path centralities focus on the number of times a vertex acts as a bridge along all paths between any pair of vertices. Proximity centralities are based on distance metrics. Spectral centralities consider the involvement degree of vertices in the substructures of networks. Here we review 12 typical centralities, and the detailed formulae are described in Table 2.

3. Correlation Analysis of Centralities

One of the concerns of this paper is the correlation between centralities in different network models. In this paper, we introduce PCA [18] to quantify the correlation between these 12 centralities. PCA is a technique for reducing the dimensionality of datasets, increasing interpretability but minimizing the information loss.

In this paper, the variables are the 12 centrality indices mentioned in Sect. 2. PCA is applied for dimensionality reduction by projecting the centrality values onto only the first and second principal components to obtain lower-dimensional data while preserving as much of the data's variation as possible. Each centrality could be denoted as the vector of the first and second components. Then the cosine value of the angle between each vector pair could be a statistic estimate of the correlation between the corresponding centrality pair.

4. Robustness Measure

In this paper, the relative size of the largest connected component (LCC) [11] is considered to measure the response of the networks when vertices are removed:

$$LCC = S'/S_0 \tag{1}$$

where S' is the number of vertices in the largest connected component after attacking. S_0 is the number of vertices in the largest connected component of the initial network. Schneider et al. [19] proposed the robustness measure R index:

$$R = \frac{1}{N}\sum_{Q=1}^{N} s(Q) \tag{2}$$

where $s(Q)$ is the fraction of vertices in the largest connected cluster after removing Q vertices. Iyer et al. [20] proposed V index to measure the vulnerability:

$$V = \frac{1}{2} - R \tag{3}$$

5. Simulation Results

The purpose of our experiments is to investigate two statistical issues related to network robustness analysis. One is the correlations between centralities and the difference of the correlations from one network model to the other. The other is the robustness of networks when different attack strategies are applied to remove vertices and the difference of robustness from one network model to the other.

In this paper, all the experiments are based upon Python3.7 environment. 1000 networks are generated for each network model (ER and BA) and there are 1000 nodes in one network. Then we get the correlations of 12 centralities for the two models. Finally, we verify the robustness of the two models.

5.1 Correlation of Centralities for Two Network Models

Based on PCA, the average correlation coefficients of 12 attack strategies for distinct network models are given in Fig. 1 (b) and Fig. 2 (b) (Fig. 1 (a) and Fig. 2 (a) will be explained in Sect. 5.1). It can be understood that light colors represent high correlations. For ER and BA models,

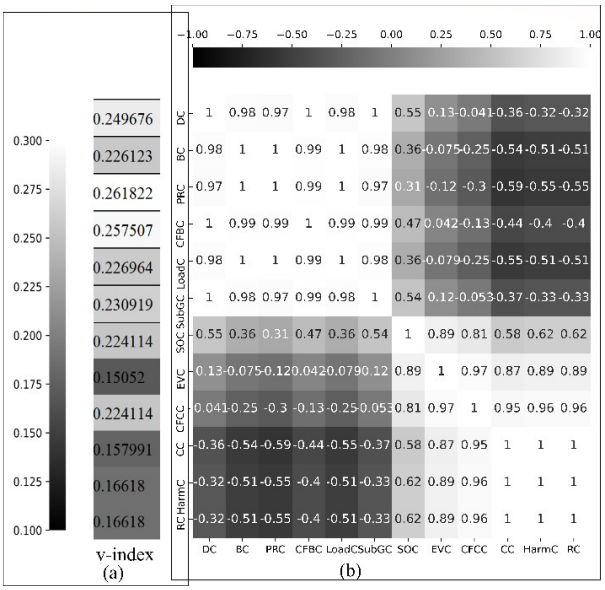

Fig. 1 (a) V-index for ER networks. (b) Correlation of centralities for ER networks.

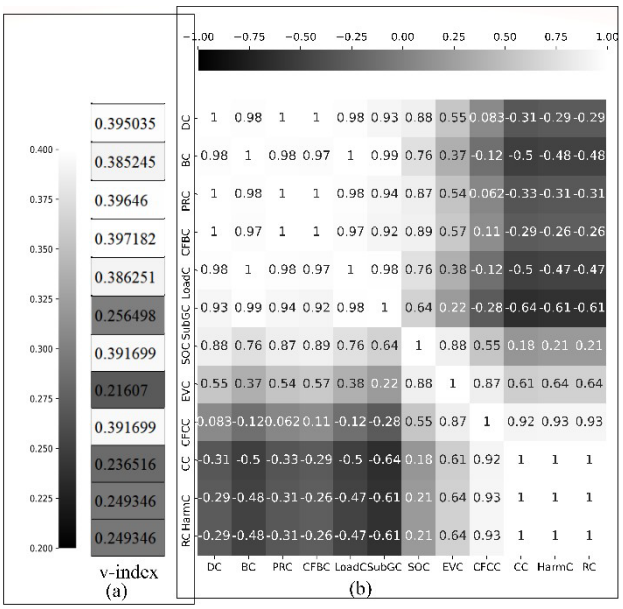

Fig. 2 (a) V-index for BA networks. (b) Correlation of centralities for BA networks.

a quite high positive correlation is observed between any pair of DC, BC, CFBC, LoadC, PRC and SubGC. SOC is highly correlated with EVC and CFCC on ER networks while highly correlated with DC, PRC, CFBC and EVC on BA networks.

The correlation coefficient is exactly 1 for any pair of CC, HarmC and RC. The three indices are all highly correlated with EVC and CFCC on ER models, and only highly correlated with CFCC on BA networks. It can be observed that highly correlated centralities may not belong to the same centrality class according to Table 2.

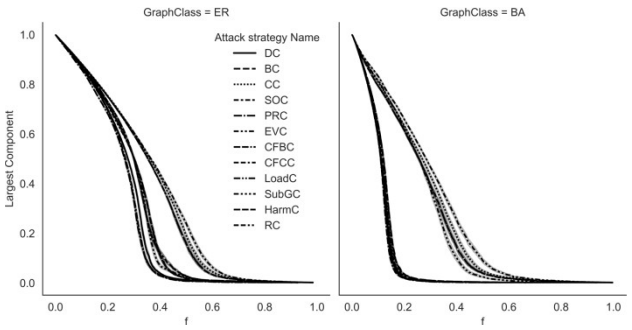

Fig. 3 Robustness of networks on the basis of 12 centralities.

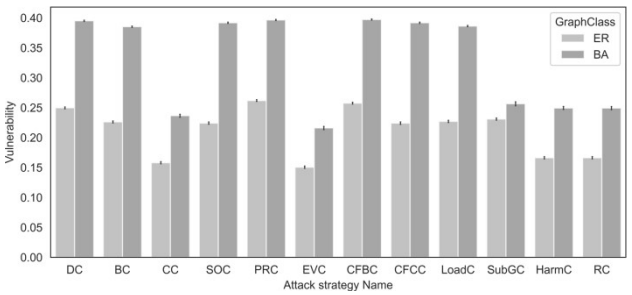

Fig. 4 Vulnerability of three network models with errorbar.

To test the significance of these correlations, significance test is considered in this paper, and $p < 0.05$ is used for the significance criterion. One sample t-test [21] is considered to examine the difference between the mean of correlation values and 0, and the result shows that all the correlations are significantly different from 0.

5.2 Robustness of Two Network Models

For the robustness of two network models, Figure 3 shows how the relative size of the largest connected component and its confidence interval change with the number of removal vertices on 12 attack indices. Shaded areas denote the 95% confidence intervals. And the vulnerability of these networks with errorbar is shown in Fig. 4. From Fig. 3 and Fig. 4, it is apparent that BA networks are more vulnerable than ER networks.

For the sake of convenience in observing the performance of highly correlated attack strategies, we show the average V-index in Fig. 1 (a) and Fig. 2 (a) where the background color of cells ranges from white to dark grey, indicating that the V-index ranges from large to small. From Figs. 1 (b) and 2 (b) it can be observed that the centrality metrics from top to bottom can be divided into two groups according to their correlations, meanwhile, from Figs. 1 (a) and 2 (a), the V-index values of centrality metrics decline from top to bottom (where there are some exceptions). Now, we can say that the highly correlated metrics might cause similar damage to the network connectivity with a high probability.

For the destructiveness, it can be observed that the

removal strategies PRC, CFBC and DC perform better than others. PRC performs best on ER networks, while CFBC performs best on BA networks. One interesting phenomenon is that these three attack strategies are highly correlated with each other on two network models. While for vulnerable BA scale-free networks, BC, SOC, CFCC and LoadC all perform well, and other intentional attack strategies do not perform badly.

6. Conclusion and Discussion

On the basis of above research, we can answer the three questions proposed in the introduction section.

First, for the 12 centrality metrics mentioned in this paper, we find that there are many highly positively correlated metric couples based on the PCA method, and most of the correlations are consistent across ER and BA models where there are some slightly differences between them.

Second, to test the significance of these correlations, one sample t-test is considered and $p < 0.05$ is used for the significance criterion. And the result shows that the means of all correlations are significantly different from 0.

Third, we find that highly correlated metrics might cause similar damage to the network connectivity with a high probability. Especially the three highly correlated metrics: PRC, CFBC and DC, which are the three most destructive ones. PRC performs best on ER networks, while CFBC performs best on BA networks, which indicates that the optimal attack strategy is different for networks with different features, but the optimal attack strategies might be highly correlated with each other.

Deep learning neural networks [22] are an example that natively supports multi-label classification problems. So in the future work, for networks that are labeled with network model classes and other structural characteristics, maybe we can find a paradigm that combines several highly correlated metrics at the same time, such as PRC, CFBC and DC, to generate the optimal attack strategy, by using a regression model that is generated by the neural network.

References

[1] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," Nature, vol.406, pp.378–382, July 2000.
[2] L.K. Gallos, P. Argyrakis, A. Bunde, R. Cohen, and S. Havlin, "Tolerance of scale-free networks: from friendly to intentional attack strategies," Physica A: Statistical Mechanics and its Applications, vol.344, no.3-4, pp.504–509, Dec. 2004.
[3] J. Matta, G. Ercal, and J. Borwey, "The vertex attack tolerance of complex networks," RAIRO-Operations Research, vol.51, no.4 pp.1055–1076, Jan. 2017.
[4] D. Magoni, "Tearing down the Internet," IEEE J. Sel. Areas Commun., vol.21, no.6, pp.949–960, Sept. 2003.
[5] M.E.J. Newman, "A measure of betweenness centrality based on random walks," Social networks, vol.27, no.1, pp.39–54, Jan. 2005.
[6] A.-M. Kermarrec, E.L. Merrer, B. Sericola, and G. Trédan, "Second order centrality: Distributed assessment of nodes criticity in complex networks," Computer Communications, vol.34, no.5, pp.619–628, April 2011.
[7] P. Erdös and A. Rényi, "On random graphs I," Publicationes Mathematicae, vol.6, pp.290–297, 1959.
[8] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," Science, vol.286, no.5439, pp.509–512, Oct. 1999.
[9] F. Grando, L.Z. Granville, and L.C. Lamb, "Machine learning in network centrality measures: tutorial and outlook," ACM Computing Surveys, vol.51, no.5, Article 102, Oct. 2018.
[10] M.E. Shaw, "Group structure and the behavior of individuals in small groups," The Journal of Psychology Interdisciplinary and Applied, vol.38, no.1, pp.139–149, July 1954.
[11] X.-L. Guo, Z.-M. Lu, and H. Li, "The Invulnerability of Traffic Networks under New Attack Strategies," IEICE Trans. Fundamentals, vol.E100-A, no.10, pp.2106–2112, Oct. 2017.
[12] U. Brandes and D. Fleischer, "Centrality measures based on current flow," Annual Symposium on Theoretical Aspects of Computer Science, Lecture Notes in Computer Science, vol.3404, Springer, Berlin, Heidelberg, pp.533–544, 2005.
[13] U. Brandes, "On variants of shortest-path betweenness centrality and their generic computation," Social Networks, vol.30, no.2 pp.136–145, May 2008.
[14] K.-I. Goh, B. Kahng, and D. Kim, "Universal behavior of load distribution in scale-free networks," Physical Review Letters, vol.87, no.27, Article: 278701, Dec. 2001.
[15] O. Skibski and J. Sosnowska, "Axioms for distance-based centralities," Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana USA, pp.1218–1225, Feb. 2018.
[16] B. Ruhnau, "Eigenvector-centrality — a node-centrality?," Social Networks, vol.22, no.4, pp.357–365, Oct. 2000.
[17] V. Grolmusz, "A note on the pagerank of undirected graphs," Information Processing Letters, vol.115, no.6-8, pp.633–634, June 2015.
[18] H. Abdi and L.J. Williams, "Principal component analysis," Wiley Interdisciplinary Reviews: Computational Statistics, vol.2, no.4, pp.433–459, July 2010.
[19] C.M. Schneider, A.A. Moreira, J.S. Andrade Jr., S. Havlin, and H.J. Herrmann, "Mitigation of malicious attacks on networks," Proceedings of the National Academy of Sciences, vol.108, no.10, pp.3838–3841, Feb. 2011.
[20] S. Iyer, T. Killingback, B. Sundaram, Z. Wang, and S. Hayasaka, "Attack robustness and centrality of complex networks," PloS one, vol.8, no.4, Article: e59613, 2013.
[21] N. Cressie, "Relaxing assumptions in the one sample t-test," Australian Journal of Statistics, vol.22, no.2, pp.143–153, June 1980.
[22] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks, vol.61, pp.85–117, Jan. 2015.