International Journal of Innovative Computing, Information and Control
Volume **20**, Number **4**, August **2024**

ICIC International ©2024 ISSN 1349-4198
pp. **979–996**

CORRELATION ANALYSIS OF ATTACK STRATEGIES AND ROBUSTNESS ANALYSIS ON THREE NETWORK MODELS

Xinling Guo[1,2], Yangming Zheng[1,2], Zhe-Ming Lu[1,2], Jialin Cui[1,3] and Hao Luo[1,2,*]

[1]Center for Generic Aerospace Technology
Huanjiang Lab
No. 7, Wenzhong Road, Zhuji 311816, P. R. China
{ nampl; zymsun2002 }@zju.edu.cn

[2]School of Aeronautics and Astronautics
Zhejiang University
No. 38, Zheda Road, Hangzhou 310027, P. R. China
zheminglu@zju.edu.cn; *Corresponding author: luohao@zju.edu.cn

[3]School of Information Science and Engineering
NingboTech University
No. 1, Qianhunan Road, Ningbo 315100, P. R. China
cuijialin@nbt.edu.cn

Received October 2023; revised February 2024

Abstract. *Network centrality metrics based on heuristics which work as a guideline on removing vertices is a vital topic for the research of network robustness. In this paper, we analyze the correlation and difference between 12 popular centrality metrics among three distinct network models, including the correlation analysis and granularity analysis. We also concern about the destructiveness of these centrality metrics when they are used as attack strategies. The results show that most of the centrality metrics are highly correlated with each other across three network models. The granularity analysis on centrality metrics is also considered in this paper. The experiment results also verify the previous conclusion that scale free networks are more vulnerable to intentional attack than other networks. We observe that more correlated centrality metrics would cause more similar damage to network connectivity; at the meantime, we also find that the attack strategies based on few centrality metrics could destroy networks very quickly. All these results inspire us that there is a paradigm that could detect the attack strategy with strong destructiveness for distinct networks by combining the highly correlated and highly destructive metrics.*

Keywords: Complex networks, Correlation analysis, Network models, Attack strategies, Granularity analysis

1. **Introduction.** Complex network theory gives us a series of methods to model and investigate various real-world systems [1], including many large-scale infrastructure networks [2, 3, 4, 5], such as the Internet, power grids, and transportation networks. Indeed, the study of complex networks is inspired by empirical analysis of real networks. Erdös and Rényi [6] introduced random network model in 1959 which is called Erdös-Rényi (ER) random network. And Watts and Strogatz [7] and Barabási and Albert [8] described the collective dynamics of small-world network which is called Watts-Strogatz (WS) small-world network, and the emergence of scaling in random scale-free network which is called Barabási-Albert (BA) scale-free network. A small-world network is structured with a high

DOI: 10.24507/ijicic.20.04.979

clustering coefficient and small average shortest distance, while a scale-free network is a type of network in which the degree distribution of vertices obeys the power-law distribution. Research on the complex networks has been mounting steadily up a number of years; however, there still remains much to be desired for the development complex network. We know that human life is dependent on many critical infrastructure networks, while a slight failure of these infrastructure networks may bring the ordinary activities of human life to a halt. For example, a traffic jam caused by a traffic accident on one expressway can cause the collapse of multiple expressways. The functionality of complex networks has much reliance on their structural robustness [9], i.e., the ability to maintain sufficient connectivity when a subset of their vertices or edges is removed.

The research of the robustness of a network attracts many interests in recent years. Albert et al. [10] first proposed two types of failures, the first one is random attack, and the other is intentional attack. This paper found that the scale-free networks are incredible vulnerable when intentional attack happened. This result inspired many researchers to do an increasing number of studies on the robustness of networks [11, 12, 13]. For intentional vertices attack, most researchers remove the vertices by their importance rank based on heuristics which is called centrality metric. Centrality metrics [14, 15, 16, 17] can evaluate the importance of the vertices based on their influence over the network topological construction and connectivity. Magoni [2] showed a few essential centrality metrics that could be used as the guideline on removing vertices, including degree centrality and betweenness centrality. Newman [18] based on the betweenness centrality presented the flow betweenness centrality by introducing a random walk. To deal with the high computational complexity of global information, Kermarrec et al. [19] proposed the second-order centrality. Most real-world networks security problems can be regarded as the robustness of the networks.

To research the network robustness, some researchers pay attention to the attack strategies on the basis of centrality metrics. Most of the researchers believe that high-complexity metrics can be replaced by high correlated but low-complexity metrics. Valente et al. [20] found strong but varied correlations among nine popular centrality measures on 58 real-world networks. On the research of 15 centrality measures on 68 real-world graphs, Baig and Akoglu [21] also found that these strategies are well correlated across different methodologies including the disruption level that they caused. Besides the correlations of 8 centrality metrics, Grando et al. [22] studied the granularity of these metrics, and found that some of them outmatch the others in granularity. Then they presented a regression model [23] generated by the neural network that could get the approximate values of centrality metrics, but they did not consider the destructive effect when the centrality metrics guiding removing the vertices. Hasson and Hussein [24] achieved the correlation analysis study for centrality metrics on six estimated centrality measures for three different datasets in order to find high correlated alternative measures for high-complexity metrics.

Although some significant contributions have been studied by these excellent researchers, there are still many questions under discussion. Compared with previous researches, this paper tries to answer the following questions. First, can we find a new method to evaluate the correlations between some mainstream centrality metrics from a statistics perspective? Second, if yes for the first question, are their correlations constant among distinct network models? Third, is there any relationship between the destructive level and their correlations among distinct network models when they are used as attacking guidelines? In other words, could highly correlated metrics cause the similar destructiveness? Fourth, is there any relationship between the most destructive centrality guideline

and the network model type? The four questions are approached for our further work about getting the most destructive attacking guideline on distinct networks.

The contributions of this paper are as follows.

- In recent years, researchers have introduced correlation analysis methods in statistics, such as Pearson correlation coefficient and Kendall rank correlation coefficient, into the study of the degree of correlation between central indicators. However, no researchers are concerned about whether these methods can effectively evaluate the degree of correlation between centrality indicators. The main research content of this paper is to propose a more suitable PCA-based method for evaluating the correlation degree between centrality indicators, and compare the effectiveness of this method with traditional correlation analysis methods in statistics used in previous studies in evaluating the correlation degree between centrality indicators.
- We adopt the granularity property to evaluate the differentiate ability of centrality metrics among distinct network models. We observe that more correlated centrality metrics would cause more similar damage to network connectivity, and we find that the attack strategies based on few centrality metrics could destroy networks very quickly. We find that there is a paradigm that could detect the attack strategy with strong destructiveness for distinct networks by combining the highly correlated and highly destructive metrics.

The remainder of this paper is organized as follows. In Section 2, we give the three popular network models in brief: ER random network, WS small-world network and BA scale-free network. In Section 3, we detail the 12 popular centrality metrics that can be used as a guideline on removing vertices. In Section 4, we introduce the principal component analysis (PCA) to quantify the correlations between these centrality metrics from a statistical perspective, and the granularity analysis method which could evaluate the differentiate ability of centrality measures when they are used to rank the importance of vertices. In Section 5, we present the robustness measures: the relative size of the largest connected component (LCC), robustness index R and vulnerability index V. In Section 6, simulation results and discussions are presented. Finally, we conclude the implications of our results in Section 7.

2. Network Models.

Consider that an undirected network is given by a pair $G = \{V, E\}$. $V = \{1, \ldots, N\}$ is the set of vertices and $E \subseteq V \times V$ is the set of edges that exist in this network. Then the adjacency matrix A, in which the element a_{ij} equals 1 represents vertices i and j are connected by an edge, 0 if they are disconnected. One edge $e_{ij} \in E$ indicates that vertices i and j connect to each other. $V_i = \{j \in V : (i, j) \in E\}$ is the set of neighbors of vertex i, and the degree of vertex i is denoted by $k_i = |V_i|$. The average network degree is denoted by $\langle k \rangle$.

Researches during the past several decades on numerous real-world systems, have revealed that many networks share some common features including scale-free degree distribution, community structure and so on [25, 26, 27]. Then the research of complex network models that generate networks of common features has attracted much attention. In this paper, we consider the most well-known 3 complex network models: Erdös-Rényi (ER) random graph [6], Barabási-Albert (BA) scale-free network [8] and Watts-Strogatz (WS) small-world network [7]. The structures of these models are described in detail as follows.

2.1. Erdös-Rényi random network.

The first network model was proposed by Erdös and Rényi, which is called random graphs. For an ER random graph model, a fixed number of vertices N is defined, and the vertices are randomly connected with each other by an identical probability p. In this kind of networks, a higher p indicates a higher

mean degree and density, and a lower network diameter. ER random network model cannot properly implement real-world networks because this model does not show any community structure and their degree distribution follows a Poisson distribution [6].

2.2. **Watts-Strogatz small-world network.** Most vertices are reachable within short paths in small-world networks [7]. Meanwhile, these networks show numerous small cycles, especially of size three. Watts and Strogatz [7] proposed a model to generate networks with the small-world properties. The graph starts with a ring of connected vertices, each one adjacent to its k-nearest neighbors. Then, with probability p, each edge is randomly reassigned to any available position. This relinking method, with an intermediate or small p (typically p should be lower than 0.5), will create paths among distant vertices while keeping a high clustering coefficient among close neighbors. A higher k indicates a higher mean degree $\langle k \rangle$, clustering, and density, although diameter decreases, while a lower clustering coefficient and diameter.

2.3. **Barabási-Albert scale-free network.** The degree distribution of large social networks follows a scale-free power-law distribution, which can be explained that networks expand continuously by the addition of new vertices and that these new vertices are preferentially connected to vertices already well connected. It starts with k number of fully connected vertices and keeps adding new vertices with k connections, defined by a preferential attachment formula [7]. The probability of a vertex p_i receiving a new connection link according to the degree d of the vertex is divided by the sum degree of all vertices. In this way, high degree vertices have a greater chance of receiving new connections than vertices with lower degree. In these networks, a higher k, the mean degree, clustering coefficient, and density get higher, while the diameter sinks.

3. **Attack Strategies Based on Centrality Metrics.** For the study on the robustness of complex infrastructure networks, many researchers concern about how to improve the performance and avoid unexpected damages, either due to random failures or intentional attacks happened on the networks. Intentional vertices attack is considered in this paper. For intentional vertices attack, attacker can attack the vertices according to their centrality values. The method of removing vertices according to their centrality metrics rank is also denoted as attack strategies in the rest of this paper. According to the purpose and conceptualization, centrality measures can be classified into four classes [23, 28, 29] shown in Table 1. The four classes: degree centralities, path centralities, proximity centralities and spectral centralities are respectively denoted as the 1st, 2nd, 3rd and 4th centralities in this paper. Here we briefly present the popular 12 centralities as follows. For the first class, degree centralities (DC) are the simplest and most straightforward centrality measures. These centralities are related with the idea of visibility that a vertex has among its neighbors. For the second class, path centralities evaluate the vertices as being central if they are in between (or at the "crossroads") of many "paths". This fact allows the vertices to control the communication through such paths. Each centrality of this class considers different kinds of paths or consists of a distinct evaluation of these paths. Most

Table 1. Four centrality classes

Centrality class				
Degree (1st)	DC			
Path (2nd)	BC	SOC	CFBC	LoadC
Proximity (3rd)	CC	CFCC	HarmC	RC
Spectral (4th)	EVC	PRC		

of these metrics require the graph to be strongly connected or evaluate each connected component of the graph individually and independently. However, there are more tolerant variations or adaptations that relax these restrictions to any kind of graph structure. For the third class, the basic idea of proximity centralities is that the lower the distance between a vertex to the others, the higher its centrality value and its independence from the network. The main difference among these centralities is that each metric computes the "distance" between vertices in a distinct way. Since these centralities are based on distance metrics, there is an inherent problem with disconnected graphs: depending on the centrality measure, the distance between two disconnected vertices are considered infinite or the largest possible distance for the given network size. For the fourth class, spectral centralities evaluate the vertices centrality by their participation in substructures of the network. They are called spectral measures because of their relation with the set of eigenvalues of the adjacency or Laplacian matrix of the graph representing the network.

3.1. **Degree centralities.** Degree centralities calculate the importance of vertices by the number of neighbors. They were introduced by Shaw [33] and popularized by Freeman [34]. The degree centrality (DC) is described below.

$$C_D(v) = k_v/(N-1) \tag{1}$$

where k_v is the degree of vertex v, and N is the total number of vertices of the network.

If the network is directed (meaning that ties have direction), then two separate measures of degree centralities are defined, namely, indegree and outdegree. They can be thought of as a kind of popularity measure, but crude measures that do not recognize a difference between quantity and quality.

3.2. **Path centralities.** Path centralities evaluate the importance of vertices by the number of times a vertex acts as a bridge among paths existing in the network. The path centralities used in this paper are shown as the following: betweenness centrality (BC) [30], second order centrality (SOC) [19], current-flow betweenness centrality (CFBC) [18] and load centrality (LoadC) [23], while LoadC is slightly different from BC.

For path centralities, the most popular measure is BC considering the shortest paths, called geodesics. This measure was introduced by Shaw [33].

$$C_B(v) = \sum_{s=1}^{N} \sum_{t=s+1}^{N} g_{st}(v) \Big/ g_{st} \tag{2}$$

where g_{st} is the total number of the shortest paths from vertex s to t and $g_{st}(v)$ is the number of those paths that pass through v.

To optimize the imperfection of betweenness centrality as it considers only shortest paths. SOC was introduced by Kermarrec et al. [19] based on a random walk. A distributed algorithm is used to compute the standard deviation of the return times for each vertex based on the Metropolis-Hastings process in which Markov chain is homogeneous and irreducible. A classical discrete time Markov chain is used on the finite state space S to represent the random walk. The standard deviation $\sigma(v)$ is shown as below.

$$\sigma(v) = \sqrt{2\sum_{u \in S} M(v,u) - |S|(|S|+1)} \tag{3}$$

where $M(v,u)$ represents the expected time starting from state v to reach state u for the first time.

CFBC uses an electrical current model for information spreading in contrast to betweenness centrality which uses the shortest paths [18]. CFBC of a vertex v is the average of the current flow over all source-target pairs:

$$C_{CFB}(v) = \sum_{s \neq t \in S} I_v^{(st)} \Big/ \frac{1}{2} N(N-1) \tag{4}$$

where $I_v^{(st)}$ is the current flow through vertex v between source s and sink t.

LoadC is slightly different from BC [32]. Let θ_{st} be a quantity of a commodity that is sent from vertex s to vertex t. We assume the commodity is always passed to the next hop following the minimum weight paths (consider an algorithm to define the minimum weight path), and in case of more than one next hop, traffic is divided equally among them. We call $\theta_{st}(v)$ the overall commodity forwarded by vertex v. The load centrality of v is given by

$$C_{Load}(v) = \frac{2}{N(N-1)} \sum_{s,t \in S} \theta_{st}(v) \tag{5}$$

3.3. **Proximity centralities.** Distance between vertices also is introduced to evaluate their importance which is named as proximity centralities. This centrality believes that the lower the distance between a vertex to the others, the higher its centrality value and its independence from the network. There are four proximity centralities considered in this paper: closeness centrality (CC) [33], current-flow closeness centrality (CFCC) [34], harmonic centrality (HarmC) [33] and reach centrality (RC) [35].

CC of a vertex measures the average farness (inverse distance) to all other vertices.

$$C_C(v) = (N-1) \Big/ \sum_{u=1, u \neq v}^{N} d_{vu} \tag{6}$$

where d_{vu} is the shortest distance between vertices v and u.

CFCC is variant of CC based on effective resistance between vertices in a network. This metric is also known as information centrality. We treat the graph G as a resistor network via replacing every edge e by a resistor with resistance $r_e = 1/w(e)$. Let $v_{st}(v)$ denote the voltage of v when a unit current enters the network at s and leaves it at t. The CFCC of vertex v is defined as

$$C_{CFC}(v) = N \Big/ \sum_{u=1, u \neq v}^{N} (v_{vu}(v) - v_{vu}(u)) \tag{7}$$

HarmC (also known as valued centrality) is a variant of closeness centrality, that was invented to solve the problem the original formula had when dealing with unconnected graphs. As with many of the centrality algorithms, it originates from the field of social network analysis. The HarmC of v is given by

$$C_{Harm}(v) = 1 \Big/ \sum_{u=1, u \neq v}^{N} d_{vu} \tag{8}$$

The use of the HarmC mean avoids cases where an infinite distance outweighs.

RC is related to the number of vertices that can be reached from every other vertex in k hops or less. For $k = 1$, this is equivalent to DC. For directed networks, both in-reach and out-reach are calculated. The routine also calculates weighted distance reach centrality

for each vertex. RC of v is measured using the formula:

$$C_R(v) = 1 + \sum_{x=1}^{k} r_{vx} \Big/ x \tag{9}$$

where r_{vx} is the number of vertices at x hop distance from v. The upper limit in the summation is due to the fact that the maximum hop a node can have is its eccentricity.

3.4. **Spectral centralities.** Metrics in this group consider the involvement of vertices in the substructures of networks. They are called spectral measures because of their relation with the set of eigenvalues of the adjacency or Laplacian matrix of the graph representing the network. While the mostly widely known among these measures is the eigenvector centrality (EVC) [36], PageRank centrality (PRC) [37] and SubGraph centrality (SubGC) [23].

PRC is first used to rank web pages of Google. It is supposed to characterize the behavior of a guest browsing the web pages. This process can be modelled by a simple combination of a random walk with occasional jumps towards randomly selected vertices to evaluate the importance of vertices for a network. This can be described by the simple set of implicit relations:

$$C_{PR}(v) = \frac{q}{N} + (1-q) \sum_{j:j\to v} p(j) \Big/ k_{out}(j) \tag{10}$$

$k_{out}(j)$ the outdegree of vertex j and the sum runs over the vertices pointing towards i. For undirected networks, $k_{out}(j) = k_j$. The damping factor q is a probability, that weighs the mixture between random walk and random jump. On practical applications it is usually set to small values (typically 0.15). For any $q > 0$ the process reaches stationarity, as a walker has a finite (no matter how small) probability to escape from a dangling end, whenever it lands there.

EVC considers not only immediate contacts but also indirect connections with every vertex of the network. Moreover, it weighs contacts of a vertex according to their own centrality. The importance $C_{EV}(v)$ of node v is just proportional to the sum of the importance of the neighboring vertices pointing to it. The EVC of v is given by ϵ, α

$$C_{EV}(v) = \alpha \left(A^T C_{EV}\right)_v + \epsilon \tag{11}$$

The role of the parameter ϵ reminds that of the damping factor q in PRC. The parameter α weighs the relative importance of the contribution of the peers versus that of the node itself.

SubGC of a vertex v is the sum of weighted closed walks of all lengths starting and ending at vertex v. The weights decrease with path length. Each closed walk is associated with a connected subgraph. It can be found using a spectral decomposition of the adjacency matrix:

$$C_{SubG}(v) = \sum_{u=1}^{N} \left(E_u^v\right)^2 e^{\lambda u} \tag{12}$$

where E_u is an eigenvector of the adjacency matrix A of G corresponding to the eigenvalue λ_j.

4. **Centralities Analysis.** One goal of this paper is to analyze the correlation and difference of centrality metrics among distinct network models. In this section, we introduce principal component analysis (PCA) and granularity analysis.

4.1. **Correlation analysis based on PCA.** To evaluate the correlations between centrality metrics, most of the researchers have estimated the correlation of each couple of centrality metrics as a ratio unit less number between the covariance of two variables and the product of their standard deviations [24], or the Pearson correlation coefficient [20]. Correlations are indicators of the strength of the linear relationship between two different variables, x and y. A linear correlation coefficient that is greater than zero indicates a positive relationship.

Many simulation results have shown that traditional correlation analysis methods such as Kendall correlation coefficient, Pearson correlation coefficient, and Spearman correlation coefficient cannot effectively estimate the correlation between central indicators. PCA is the simplest method for analyzing multivariate statistical distributions using feature quantities. The result can be understood as an explanation for the variance in the original data: which direction of data values has the greatest impact on the variance? In other words, PCA provides an effective way to reduce data dimensions; If the analyst removes the component corresponding to the smallest eigenvalue from the original data, the resulting low dimensional data must be optimized (i.e., reducing the dimensionality in this way is the method that loses the least amount of information). Principal component analysis is particularly useful in analyzing complex data, such as facial recognition. As far as we know, there is no research work on correlation analysis among centrality metrics. In this paper, we introduce PCA to evaluate the correlation from a statistical perspective for the first time. PCA is used in exploratory data analysis and for making predictive models. It is commonly used for dimensionality reduction by projecting each data point onto only the first few principal components to obtain lower-dimensional data while preserving as much of the data's variation as possible. The first principal component can equivalently be defined as a direction that maximizes the variance of the projected data. The cosine similarity analysis of the data after PCA dimensionality reduction can be used as an estimate of the correlation coefficient between centrality indicators. The simulation results later will show that the correlation algorithm based on PCA proposed in this paper can effectively quantitatively estimate the correlation between central indicators.

In this paper, the variables are the values of the centrality metrics in Section 3. For each network, the matrix $X \in \mathbb{R}^{N\times 12}$ is composed of 12 centrality values of all vertices where N is the number of vertices, and the flowchart of correlation analysis on the 12 centrality metrics for each network is shown in Figure 1.

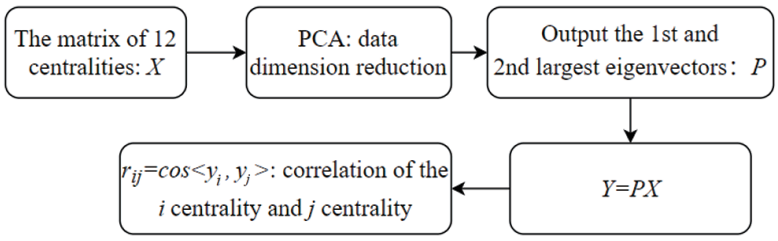

FIGURE 1. Flowchart of PCA method applied to estimate the correlations bentween the 12 centrality metrics

After we get the first and second components for each centrality metrics, the cosine of the angle between each centrality couple could be a statistic estimate of the correlation between them. The correlation value r should range between -1.0 to $+1.0$. A higher coefficient value implies a higher correlation between the centrality indices and vice versa. $r = -1.0$ indicates a perfect negative correlation, while $r = +1.0$ indicates a perfect positive correlation.

4.2. **Granularity analysis.** To evaluate the differentiate ability of centrality measures when they are used to rank the importance of vertices, the granularity property [23, 28, 38] is considered in this paper. We simulate the granularity as the percentage of distinct centrality values of each centrality metric on each network. For example, if there are 50 different values of a certain centrality measure on a network with 100 vertices, we can get that are 50% distinct values for this centrality measure on this network, which means the granularity of this centrality measure on this network is 0.5.

5. **Robustness Measure.** To evaluate the destructiveness of these centralities when they are used for removing vertices, we consider the relative size of the largest connected component (LCC) in this paper. The largest connected component is the connected subgraph with the largest size (number of vertices in it). While the largest connected component will be smaller and smaller with the removal of vertices. The ratio of the largest connected component of the network before and after vertices removal is called LCC [39]:

$$LCC = S'/S_0 \tag{13}$$

where S' is the number of vertices in the largest connected component after the attack. S_0 is the number of vertices in the largest connected component of the initial network. To measure the robustness of network before collapsing, Schneider et al. [32] proposed the robustness measure R index:

$$R = \frac{1}{N}\sum_{Q=1}^{N} s(Q) \tag{14}$$

where $s(Q)$ is the fraction of vertices in the largest connected cluster after removing Q vertices. The normalization factor $1/N$ ensures that the robustness of networks with different sizes can be compared. The range of possible R values is between $1/N$ and 0.5, where these limits correspond, respectively, to a star network and a fully connected graph. Iyer et al. [29] proposed V index to measure the vulnerability of a network to a given scheme of vertex removal, to be the complementary quantity to R:

$$V = \frac{1}{2} - R \tag{15}$$

6. **Simulation and Implementation Results.** Our goal is to study the correlations and difference of 12 popular centrality metrics among three distinct network models. We are also curious about that the robustness of distinct network models against different attack strategies, and the relationship between destructive level of these centrality metrics when they used as attacking guidelines and their correlations among distinct network models. Moreover, we attempt to compare the estimated results with the known ground truth.

All of the simulations are implemented upon Python3.7 workbench. For simulating data, 1000 networks where 1000 vertices exist for each network are generated for each network model (ER, WS and BA) on the basis of well-known python package NetworkX2.5.

6.1. **Correlation analysis between centrality metrics.** It is known that PCA is often used as a dimensionality-reduction technique. After we get the 12 centrality values of all networks, first, we apply dimensionality-reduction technique of PCA to obtaining the first and second components of each centrality metric. Then centrality metrics can be denoted as the vector of their first and second components. Last, the cosine of the angle between the two vectors of each centrality couple could be an estimate of the coefficient of correlation between them. The statistical results of the correlations are shown in Figures

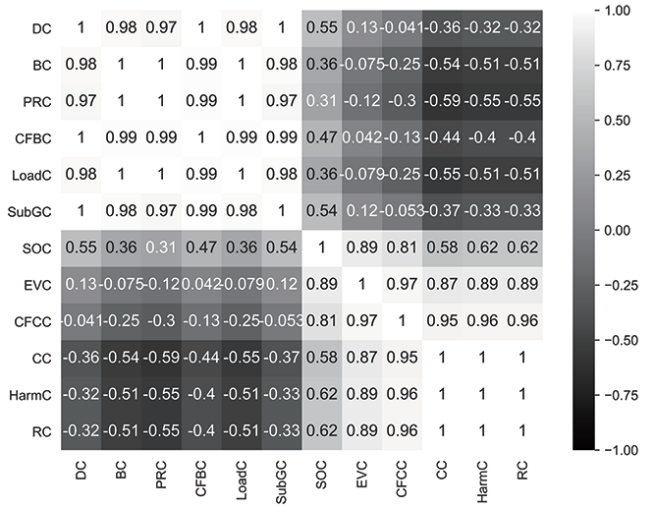

FIGURE 2. Correlations of 12 centralities on ER models

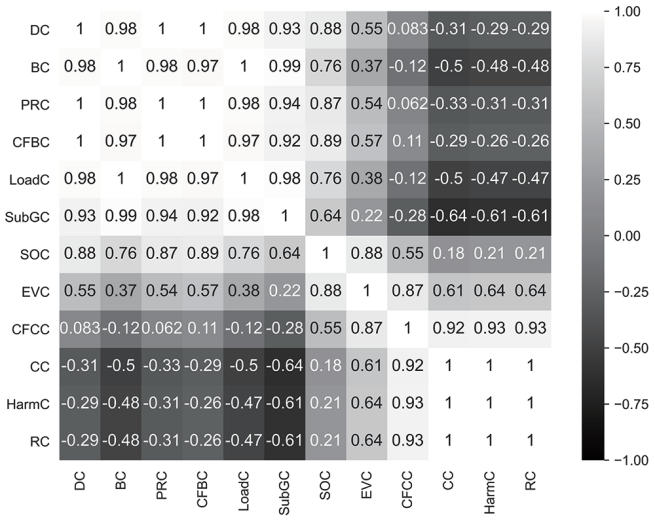

FIGURE 3. Correlations of 12 centralities on BA models

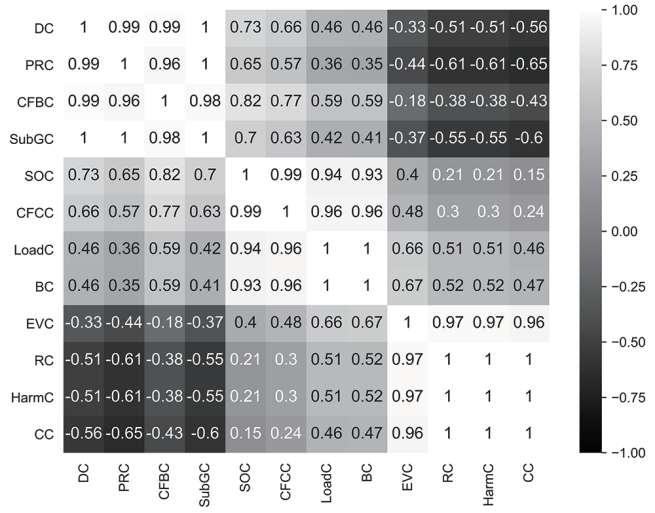

FIGURE 4. Correlations of 12 centralities on WS models

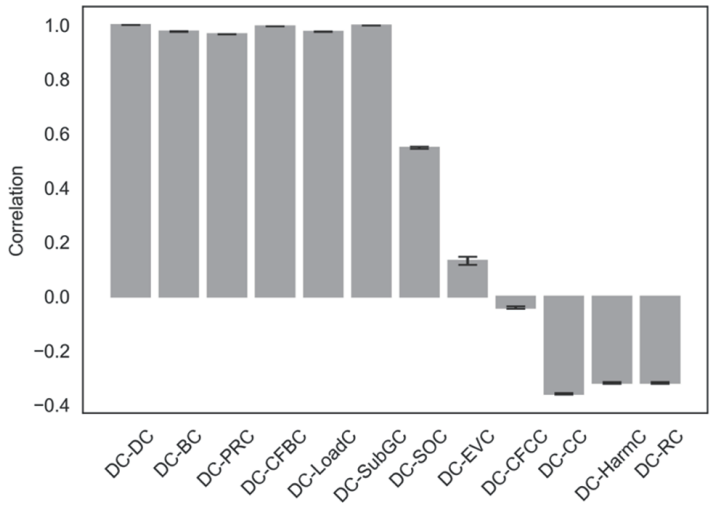

FIGURE 5. Correlations between DC and other centrality metrics based on PCA method for ER models, with errorbar

FIGURE 6. (color online) The distribution of the highly positive correlated ($r > 0.8$) centrality metrics among different networks

2-4. It should be clarified that a lighter color represents a higher positive correlation. One example of the correlations between DC and others with errorbar is given in Figure 5.

As Figures 2-4 show, some centrality metrics are highly correlated with each other, and there is some difference between distinct network models. Figure 6 gives the distribution of the highly positive correlated ($r > 0.8$) centrality metrics among different networks. As Figure 6 shows: DC, PRC, CFBC and SubGC are highly positive correlated with each other among the 3 different network models (ER, BA, WS). However, interestingly, the four metrics are not classified to one same centrality class according to their purpose and conceptualization as shown in Table 1. And there is a strong positive correlation between BC and LoadC across three models, while BC and LoadC are both classified to path centralities. In addition, CC, HarmC and RC are perfect correlated with each other with $r = 1$, because they are equal to each other when the network is undirected and unweighted according to Equations (6), (8) and (9). For the sake of the rest of the discussion, we denote DC, PRC, CFBC, and SubGC as the I correlation centralities, BC and LoadC as the II correlation centralities, CC, HarmC and RC as the III correlation centralities

TABLE 2. The distribution of correlations between centrality metrics across three network models

E+W+B	E+B	E+W	W+B	BA	WS
I∼I	II∼I	SOC∼CFCC	SOC∼CFBC	SOC∼DC, PRC	SOC, CFCC∼III
II∼II	EVC∼SOC, CFCC	EVC∼III			
III∼III	CFCC∼III, EVC				

in this paper according to distribution of their correlations among 3 models. For each correlation centralities group, the metric in each group is highly positive correlated with others of this group across three network models (ER, WS, BA).

The distribution of highly positive correlations ($r > 0.8$) is shown in Table 2. In this table, if $A \sim B$, then each metric of A is highly positive with all of the metrics of B. As Table 2 shows, many of the centralities are highly positive correlated with each other, but vary slightly across distinct network models.

6.2. **Granularity analysis of centrality metrics.** In this paper, the granularity property [38] is considered to evaluate the differentiate ability of centrality metrics among distinct network models. And we are also curious about the relationship between their correlations and their granularity property. The granularity statistic mean values with errorbar of each centrality metrics grouped by the network models ER, WS and BA are shown in Figure 7. It is obvious that the granularity of all centrality metrics presented nearly no difference in the expected values (considering the confidence intervals) on three network models. While interesting, only DC (0.15) and CC (0.75) present a low granularity while others present a great granularity (0.9), which is consistent with the results of [23]. It should be indicated that a higher granularity implies a higher performance of differentiating the vertices of a network, meanwhile, a higher space complexity.

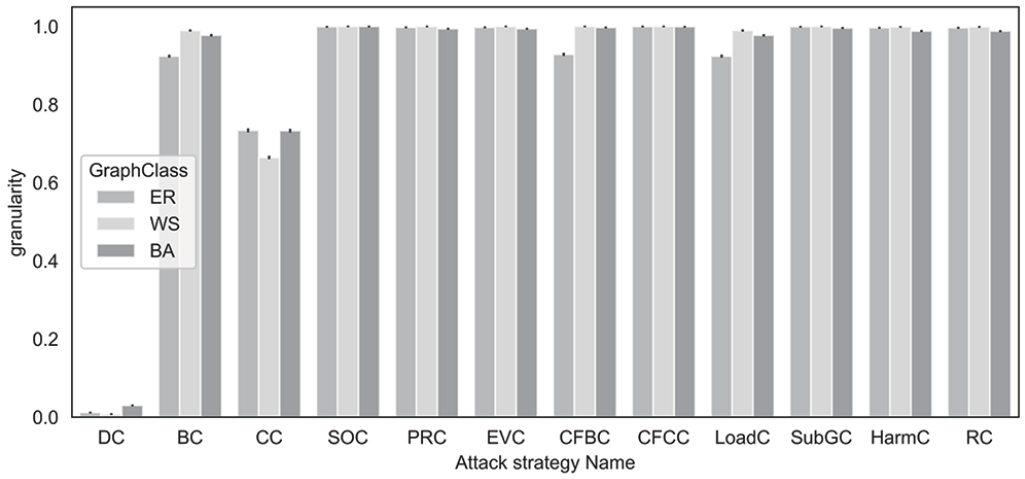

FIGURE 7. Granularity feature of 12 centralities on three network models with errorbar

6.3. **Robustness of networks against centrality metrics attacking guideline.** LCC, robustness index R and vulnerability index V are considered to evaluate the destructive level of these centrality metrics when they are used as guidelines for removing vertices. Figure 8 shows how LCC (shaded areas denote the 95% confidence intervals) decaying with the removal vertices proportions on 12 attack strategies among three network models. It is obvious that network scale shrinks quickly against some attack strategies.

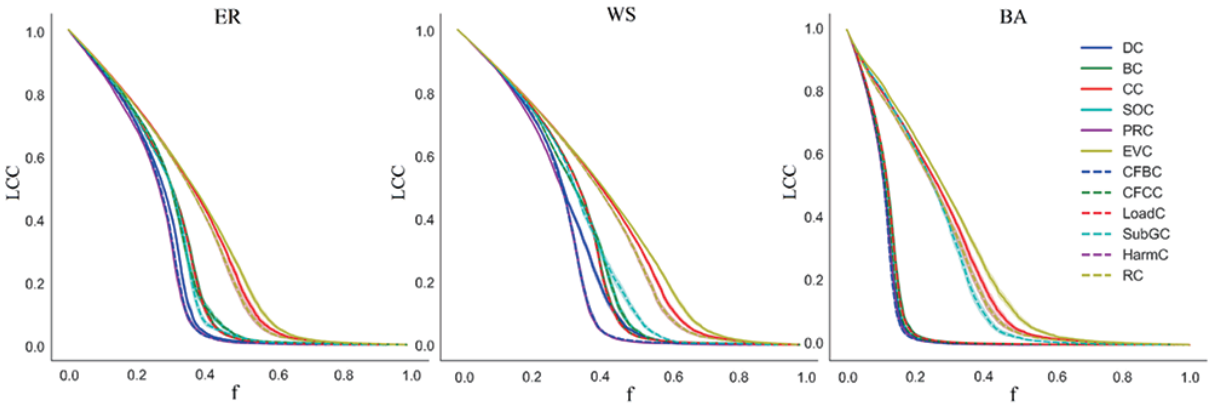

FIGURE 8. (color online) Robustness against simultaneous attack for three network models, where f is the fraction of removed vertices

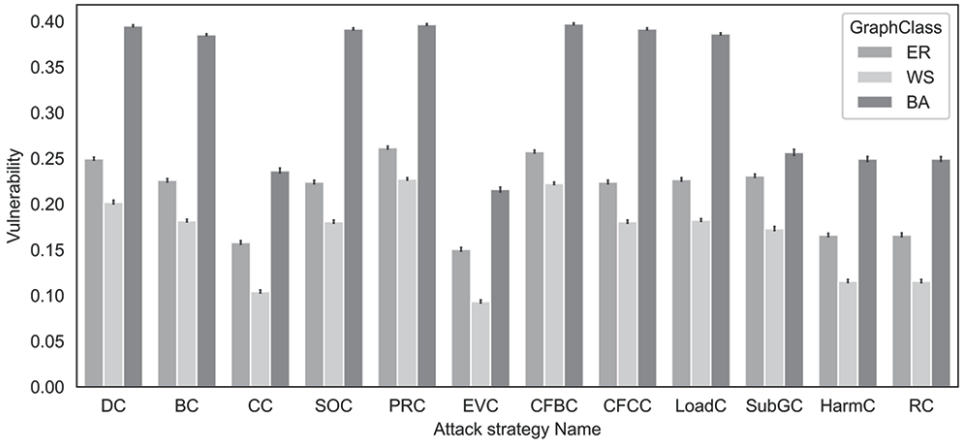

FIGURE 9. Vulnerability of three network models with errorbar

Especially for BA network models, they become completely disconnected (LCC = 1/N) when the removing vertices proportion is nearly to 0.2, which indicates that BA networks are more vulnerable under intentional attack than others.

To observe the global attack performance of 12 attack indices among three network models, the V index of the global destructiveness is given in Figure 9 which is grouped by 12 attack strategies. From Figure 9, we can easily capture that the bar of BA is higher than the other two bars in each attack strategy group which indicates that BA network is more vulnerable than ER and WS networks. It is consistent with the conclusion of many researches [8, 41, 42, 43].

To study the relationship between the correlations of these centrality metrics and their destructive level when they are used as attack strategies among the three network models, more information is listed in Table 3. In Table 3, for each network model, the V-index ranges from large to small. From a global view, it can be observed that the I correlation centralities are more destructive than the II correlation centralities; furthermore, the II correlation centralities are more destructive than the III correlation centralities when they are used as guidelines for removing vertices. Now, we could state that, for two centrality metrics, the higher their correlation, the similar their destructive level when they are used as attack strategies. This evidence confirms that PCA is applicable to quantify the correlations of centrality metrics from another perspective. While there is no

TABLE 3. The V-index of three network models under 12 attack strategies

Correlation group	I				II	
Network model	DC	PRC	CFBC	SubGC	BC	LoadC
ER	0.249676	0.261822	0.257507	0.230919	0.226123	0.226964
WS	0.202025	0.227531	0.222657	0.173125	0.181854	0.182537
BA	0.395035	0.39646	0.397182	0.256498	0.385245	0.386251
Centrality class	1st	4th	2nd	4th	2nd	2nd
Correlation group			III			
Network model	SOC	CFCC	CC	HarmC	RC	EVC
ER	0.224114	0.224114	0.157991	0.16618	0.16618	0.15052
WS	0.180795	0.180795	0.104256	0.115675	0.115675	0.0933616
BA	0.391699	0.391699	0.236516	0.249346	0.249346	0.21607
Centrality class	2nd	3rd	3rd	3rd	3rd	4th

clear relationship between the destructive level and the centrality class which is classified according to their purpose and conceptualization.

To observe the detail of local attack performance to detect the most destructiveness attack strategy of 12 attack indices on three distinct network models, we extract the I correlation centralities for ER and WS networks, I, II, SOC and CFCC for BA networks from Figure 8, as shown in Figure 10. From Figure 10, it can be observed that the attack strategies based on highly positive correlated centralities metrics: PRC and CFBC could rapidly cause serious damage to network connectivity of the three network models, while there is some slightly difference about the damage rapid of the two metrics across three distinct network models. For ER networks, PRC is a little bit more destructive than CFBC from the beginning to the end of the attack program. For WS networks, at the beginning, PRC is the most destructive, and then when the proportion of removing vertices is nearly to 0.3, CFBC and PRC cause the same damage. While for BA networks, CFBC and PRC cause the same damage at the beginning, then when the proportion of removing vertices is nearly to 0.1, CFBC becomes more destructive.

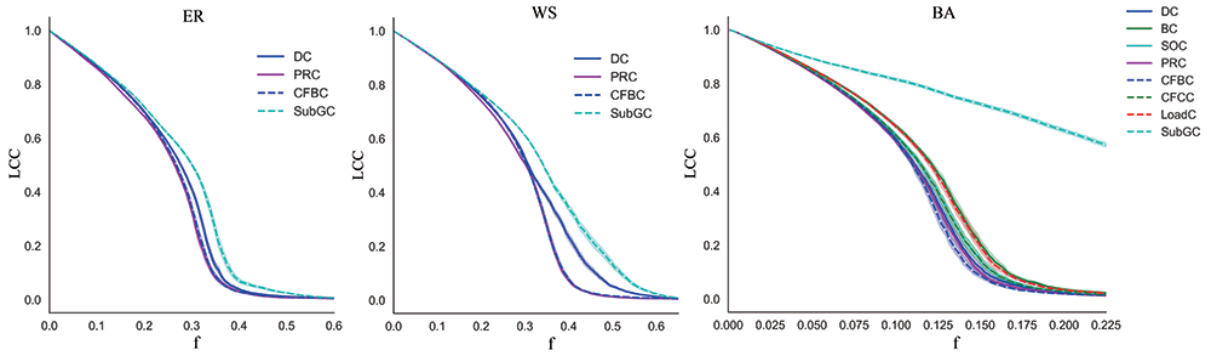

FIGURE 10. (color online) Robustness of ER networks against the high-destructive attack strategies, where f is the fraction of removed vertices

6.4. **Summary.** From above simulation results, we can summarize the following practical insights.

Many of the centralities are highly positive correlated with each other, but vary slightly across distinct network models: DC, PRC, CFBC and SubGC are highly positive correlated with each other among the 3 different network models (ER, BA, WS). However,

the four metrics are not classified to one same centrality class according to their purpose and conceptualization. And there is a strong positive correlation between BC and LoadC across three models, while BC and LoadC are both classified to path centralities. For each correlation centralities group, the metric in each group is highly positive correlated with others of this group across three network models (ER, WS, BA).

The granularity of all centrality metrics presented nearly no difference in the expected values (considering the confidence intervals) on three network models. A higher granularity implies a higher performance of differentiating the vertices of a network, meanwhile, a higher space complexity.

Network scale shrinks quickly against some attack strategies, and BA networks are more vulnerable under intentional attack than others. BA network is more vulnerable than ER and WS networks.

For two centrality metrics, the higher their correlation, the similar their destructive level when they are used as attack strategies. PCA is applicable to quantify the correlations of centrality metrics from another perspective. While there is no clear relationship between the destructive level and the centrality class which is classified according to their purpose and conceptualization.

The attack strategies based on highly positive correlated centralities metrics: PRC and CFBC could rapidly cause serious damage to network connectivity of the three network models, while there is some slightly difference about the damage rapid of the two metrics across three distinct network models.

7. Conclusions.

This paper proposes a more suitable method for evaluating the correlation degree between centrality indicators, and compares the effectiveness of this method with traditional correlation analysis methods in statistics used in previous studies in evaluating the correlation degree between centrality indicators.

We find that BA networks are more vulnerable to intentional attack than others for the global robustness analysis. We also find that the higher correlation of two metrics, the similar their destructive level when they are used as attack strategies. We also observe that PRC and CFBC attack strategies could rapidly cause serious damage to network connectivity of the three network models compared with other metrics, while there is slightly difference across three distinct network models. For ER networks, to cause most serious damage, the attack strategy based on PRC should be applied to attack network. For WS networks, when the proportion of removing vertices is under 0.3, PRC is the best choice. While for BA networks, when the proportion of removing vertices is nearly to 0.1, CFBC is the most destructive one.

In the future work, to detect the most effective centrality metric for network with various features, we will introduce deep learning to the analysis of the relationship of the destructiveness attack strategies on the basis of centrality metrics and network features.

Acknowledgment. This work was partially supported by Ningbo Science and Technology Innovation 2025 major project under grants 2020Z106 and 2023Z040. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] D. J. Watts, The "New" science of networks, *Annu. Rev. Sociol.*, vol.30, no.1, pp.243-270, 2004.
[2] D. Magoni, Tearing down the Internet, *IEEE Journal on Selected Areas in Communications*, vol.21, no.6, pp.949-960, 2003.

[3] X.-L. Guo and Z.-M. Lu, Urban road network and taxi network modeling based on complex network theory, *Journal of Information Hiding and Multimedia Signal Processing*, vol.7, no.3, pp.558-568, 2016.
[4] D. López-Pintado, Diffusion in complex social networks, *Games and Economic Behavior*, vol.62, no.2, pp.573-590, 2008.
[5] Y.-J. Zhang, Z.-J. Kang, X.-L. Guo and Z.-M. Lu, The structural vulnerability analysis of power grids based on overall information centrality, *IEICE Transactions on Information and Systems*, vol.99, no.3, pp.769-772, 2016.
[6] P. Erdös and A. Rényi, On random graphs I, *Publ. Math. Debrecen*, vol.6, pp.290-297, 1959.
[7] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, *Nature*, vol.393, no.6684, pp.440-442, 1998.
[8] A.-L. Barabási and R. Albert, Emerging of scaling in random networks, *Science*, vol.286, pp.509-512, 1999.
[9] B. Min, S. D. Yi, K.-M. Lee and K.-I. Goh, Network robustness of multiplex networks with interlayer degree correlations, *Physical Review E*, vol.89, no.4, 042811, 2014.
[10] R. Albert, H. Jeong and A.-L. Barabási, Error and attack tolerance of complex networks, *Nature* vol.406, no.6794, pp.378-382, 2000.
[11] P. Holme, B. J. Kim, C. N. Yoon and S. K. Han, Attack vulnerability of complex networks, *Physical Review E*, vol.65, no.5, 056109, 2002.
[12] 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, nos.3-4, pp.504-509, 2004.
[13] J. Matta, G. Ercal and J. Borwey, The vertex attack tolerance of complex networks, *RAIRO-Operations Research*, vol.51, pp.1055-1076, 2017.
[14] J. Wang, H. Mo, F. Wang and F. Jin, Exploring the network structure and nodal centrality of China's air transport network: A complex network approach, *Journal of Transport Geography*, vol.19, pp.712-721, 2011.
[15] S. Derrible, Network centrality of metro systems, *PloS One*, vol.7, no.7, e40575, 2012.
[16] H. Kim and R. Anderson, Temporal node centrality in complex networks, *Physical Review E*, vol.85, 026107, 2012.
[17] B. Liu, Z. Li, X. Chen, Y. Huang and X. Liu, Recognition and vulnerability analysis of key nodes in power grid based on complex network centrality, *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol.65, pp.346-350, 2018.
[18] M. E. Newman, A measure of betweenness centrality based on random walks, *Social Networks*, vol.27, pp.39-54, 2005.
[19] 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, 2011.
[20] T. W. Valente, K. Coronges, C. Lakon and E. Costenbader, How correlated are network centrality measures?, *Connections (Toronto, Ont.)*, vol.28, no.1, pp.16-26, 2008.
[21] M. B. Baig and L. Akoglu, Correlation of node importance measures: An empirical study through graph robustness, *Proc. of the 24th International Conference on World Wide Web*, Florence, Italy, pp.275-281, 2015.
[22] F. Grando, D. Noble and L. C. Lamb, An analysis of centrality measures for complex and social networks, *Proc. of IEEE Global Communications Conference (GLOBECOM)*, Washington, D.C., USA, pp.1-6, 2016.
[23] F. Grando, L. Z. Granville and L. C. Lamb, Machine learning in network centrality measures: Tutorial and outlook, *ACM Computing Surveys (CSUR)*, vol.51, no.5, pp.1-12, 2018.
[24] S. T. Hasson and Z. Hussein, Correlation among network centrality metrics in complex networks, *Proc. of the 6th International Engineering Conference on Sustainable Technology and Development*, Erbil, Iraq, pp.54-58, 2020.
[25] J. D. Noh and H. Rieger, Random walks on complex networks, *Physical Review Letters*, vol.92, no.11, 118701, 2004.
[26] A.-L. Barabási, Scale-free networks: A decade and beyond, *Science*, vol.325, no.7, pp.412-413, 2009.
[27] C. Seshadhri, T. G. Kolda and A. Pinar, Community structure and scale-free collections of Erdös-Rényi graphs, *Physical Review E*, vol.85, no.5, 056109, 2012.
[28] S. P. Borgatti and M. G. Everett, A graph-theoretic perspective on centrality, *Social Networks*, vol.28, pp.466-484, 2006.

[29] S. Iyer, T. Killingback, B. Sundaram and Z. Wang, Attack robustness and centrality of complex networks, *PloS One*, vol.8, no.4, e59613, 2013.
[30] N. Meghanathan and X. He, Correlation and regression analysis for node betweenness centrality, *International Journal of Foundations in Computer Science and Technology*, vol.6, no.6, pp.1-20, 2016.
[31] X.-L. Guo, Z.-M. Lu and H. Li, The invulnerability of traffic networks under new attack strategies, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol.100, no.10, pp.2106-2112, 2017.
[32] C. M. Schneider, A. A. Moreira, J. S. Andrade, S. Havlin and H. J. Herrmann, Mitigation of malicious attacks on networks, *Proceedings of the National Academy of Sciences*, vol.108, pp.3838-3841, 2011.
[33] M. E. Shaw, Group structure and the behaviour of individuals in small groups, *The Journal of Psychology*, vol.38, pp.139-149, 1954.
[34] L. C. Freeman, Centrality in social networks: Conceptual clarification, *Social Networks*, vol.1, pp.215-239, 1979.
[35] K.-I. Goh, B. Kahng and D. Kim, Universal behavior of load distribution in scale-free networks, *Physical Review Letters*, vol.87, no.27, 278701, 2001.
[36] Y. Rochat, *Closeness Centrality Extended to Unconnected Graphs: The Harmonic Centrality Index*, Technical Report, 2009.
[37] U. Brandes and D. Fleischer, Centrality measures based on current flow, *Proc. of the 22nd Annual Symposium on Theoretical Aspects of Computer Science*, Stuttgart, Germany, pp.533-544, 2005.
[38] O. Skibski and J. Sosnowska, Axioms for distance-based centralities, *Proc. of the 32nd AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, 2018.
[39] B. Ruhnau, Eigenvector-centrality – A node-centrality, *Social Networks*, vol.22, no.4, pp.357-365, 2000.
[40] V. Grolmusz, A note on the PageRank of undirected graphs, *Information Processing Letters*, vol.115, pp.633-634, 2015.
[41] W. Pedrycz and S. M. Chen, *Information Granularity, Big Data, and Computational Intelligence*, Springer International Publishing, 2014.
[42] J. C. Doyle, D. L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka and W. Willinger, The "robust yet fragile" nature of the Internet, *Proceedings of the National Academy of Sciences*, vol.102, no.41, pp.14497-14502, 2005.
[43] B. Joshi and R. Shrestha, Nepali speech recognition using self-attention networks, *International Journal of Innovative Computing, Information and Control*, vol.19, no.6, pp.1769-1784, 2023.

Author Biography

Xinling Guo received the B.S. degree in Communication Engineering from the Qingdao University of Technology, Qingdao, China, in 2015. She received the Ph.D. degree from Zhejiang University, China, in 2021. She is currently a researcher in Huanjiang Lab, China. Her research interests include complex networks and multimedia information processing, etc.

Yangming Zheng received his Ph.D. degree from Xi'an Electronic Science and Technology University, China, in 2005. Following his doctoral studies, he embarked on a postdoctoral research journey at Zhejiang University from 2005 to 2008, subsequently earning an appointment as an Associate Professor at the same institution. Currently, Prof. Zheng holds a full-time faculty position at Zhejiang University. His prevailing research interests encompass machine intelligence, machine vision, and unmanned factories, wherein he continues to explore and innovate within these technological and industrial domains. His dedication to research and development in these areas aims to further the advancements and applications of intelligent machinery and autonomous systems in various industrial settings.

Zhe-Ming Lu received the B.S. and M.S. degrees in Electrical Engineering and the Ph.D. degree in Measurement Technology and Instrumentation from the Harbin Institute of Technology (HIT), Harbin, China, in 1995, 1997, and 2001, respectively. He became a Lecturer with HIT in 1999. Since 2003, he has been a Professor with the Department of Automatic Test and Control, HIT. He is currently a Full Professor with the School of Aeronautics and Astronautics, Zhejiang University, Hangzhou, China. In the areas of multimedia signal processing and information hiding, he has published more than 350 papers, seven monographs in Chinese, two monographs in English and three book chapters in English. His current research interests include multimedia signal processing, information security, and complex networks.

Jialin Cui received the M.S. degree in Automation from Zhejiang University in 2005. He is currently an Associate Professor at NingboTech University. His current research interests include machine vision, artificial intelligence and application of artificial intelligence technology in information security.

Hao Luo received the B.S., M.S., and Ph.D. degrees from Harbin Institute of Technology, Harbin, China, in 2002, 2004, and 2008, respectively. He is currently an Associate Professor with the School of Aeronautics and Astronautics, Zhejiang University, Hangzhou, China. From 2014 to 2016, he was a Visiting Scholar with the School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, Sydney, Australia. His current research interests include computer vision, embedded artificial intelligence, and signal processing.