Research Article

Key Node Discovery Algorithm Based on Multiple Relationships and Multiple Features in Social Networks

Xianyong Li [ID]**, Ying Tang, Yajun Du, and Yanjie Li**

School of Computer and Software Engineering, Xihua University, Chengdu 610039, China

Correspondence should be addressed to Xianyong Li; lixy@mail.xhu.edu.cn

The key nodes play important roles in the processes of information propagation and opinion evolution in social networks. Previous work rarely considered multiple relationships and features into key node discovery algorithms at the same time. Based on the relational networks including the forwarding network, replying network, and mentioning network in a social network, this paper first proposes an algorithm of the overlapping user relational network to extract different relational networks with same nodes. Integrated with these relational networks, a multirelationship network is established. Subsequently, a key node discovery (KND) algorithm is presented on the basis of the shortest path, degree centrality, and random walk features in the multirelationship network. The advantages of the proposed KND algorithm are proved by the SIR propagation model and the normalized discounted cumulative gain on the multirelationship networks and single-relation networks. The experiment's results show that the proposed KND method for finding the key nodes is superior to other baseline methods on different networks.

1. Introduction

With the rapid development of social networks (e.g., Facebook, Twitter, and Sina Weibo), they have become main platforms for people to obtain, spread, and exchange information. In social networks, how do we quickly spread information? How do we effectively control the speed of virus diffusion? How do we efficiently suppress the width of rumor propagation? How do we correctly control and guide the evolution trend of public opinion? For these practical application scenarios, key nodes are able to play important roles in the structures and functions of networks [1–3]. In recent decades, scholars have mainly focused on single-relation networks [4, 5]. Single-relation networks mean that the networks consist of the same type of nodes and only one type of relationship between nodes. Traditional single feature mining contained degree centrality and related variants of degree centrality [6]. Chen et al. [7] proposed a degree discount centrality algorithm to effectively make influence maximization. Sheikhahmadi et al. [8] proposed degree distance centrality algorithm. Their experiments showed that the performance of this algorithm was better than other

measures including high degree and betweenness in eight large-scale networks. Wang et al. [9] proposed degree punishment method to select spreaders. They adopted SIR (Susceptible-Infected-Recovered) model to assess the performance of the method. Combined with the neighbors' numbers, neighbors' influences, and clustering coefficient, Chen et al. [10] proposed a cluster rank algorithm. Their experiments showed that the performance of this algorithm is significantly superior to the degree centrality and k-core decomposition. Besides, there are many methods to find the key nodes in networks, such as feature vector method [11], shortest path increment [12], spreading influence related centrality [13], PageRank [14], LeaderRank [15], HITS [16], k-shell centrality [17], and k-shell improved algorithm [18–21]. From these research results, it can be seen that the key node identification is very successful in single-relation networks.

However, in real social networks, users often participate in different social activities in various ways and form various connections with different users [4, 22–24]. The different interactive relationships form multirelationship networks [25]. For the multirelationship networks, there are many

types of relationships between nodes; and the types of nodes may be different. Obviously, the multirelationship networks possess more information than single-relation networks to effectively express the diversity of relationships for social networks. Battiston et al. [25] first proposed the relevant symbolic representation of the multirelation network. Boccaletti et al. [26] further enhanced the understanding of multirelationship networks. Al-Garadi et al. [27] applied the node identification method of single-relation networks to multirelation networks. They only mined the single feature of networks, failing to accurately identify the key nodes. Chen et al. [13, 28] integrated multiple centrality measures into the hierarchical ranking method to identify the key nodes in complex social networks. Wang et al. [29] proposed a new metric for measuring key nodes in multilayer networks. They further verified the metric in single-layer networks, multirelationship networks, and aggregated networks, respectively. Li et al. [30] proposed a key node identification method for multilayer networks based on evidence theory. Their method had high computational complexity and was not applicable to large social networks. Pedroche et al. [31, 32] extended some algorithms for single-layer networks to two-layer networks. Fu et al. [33] used the representation learning to learn the global structural features and local structural features of networks. Their method well represented the characteristics of nodes. Singh et al. [34] proposed a new multirelationship network aggregation method to identify key nodes. Their experiments showed that their method has obviously advantage for influence maximization across multiple social networks. Huang et al. [35] treated the community as a node through extended neighbor strategy. They transformed the multilayer network into a single-layer network and then proposed an algorithm to find the key nodes in monolayer or multilayer networks. Thus, the key node discovery is beneficial to understand the structure properties of multirelation networks.

At present, there is no unified concept of the key node discovery in multirelation networks [36–40]. The existing methods of key node discovery for multirelationship networks have not fully used the importance of the relationships between different layers and the importance of the edges between layers. Therefore, the key node discovery algorithms have yet to study for multirelation networks.

In this paper, we will design a novel algorithm for identifying top-K key nodes in social networks. Firstly, we propose an algorithm of overlapping user relational networks to establish different relational networks in social networks. Subsequently, we reconstruct a multirelationship network based on these relational networks. Then, we build a node influence measure based on the multiple relationships and multiple features to rank influential nodes. On the multirelationship network, we propose a key node discovery (KND) algorithm to obtain key node. Some comparison experiments are made to verify the effectiveness of the KND method on social networks. The main contributions of this paper are as follows:

(1) An algorithm of overlapping user relational network is proposed. This algorithm is able to extract different relational networks with the same nodes.

(2) An algorithm of establishing multirelationship network is proposed based on the different relational networks. The multirelationship network fully integrates the multiple relationships and multiple features in social networks.

(3) A key node discovery (KND) algorithm with multiple relationships and multiple features is proposed. In this method, a novel node influence measure is built on the multirelationship network. Based on the node influence method and the algorithms of overlapping user relational network and multirelationship network, the key node discovery algorithm is designed to find the key nodes on multirelationship network. Some comparison experiments on different datasets verify that the performance of the proposed KND algorithm is better than baseline methods. By the evaluation of the normalized discounted cumulative gain (NDCG), the proposed KND algorithm gets the best NDCG score in different networks. These results show that the proposed KND algorithm can accurately find the top-K nodes in the social networks.

The rest of the paper is organized as follows. Section 2 presents related definitions and proposed algorithms. Section 3 explains the experimental setup and result analyses of comparison experiments on three datasets. Section 4 draws conclusions and future directions.

2. Key Node Discovery Algorithm Based on Multiple Relationships and Multiple Features of Social Networks

2.1. Related Definitions. In a social network, there exist many relational networks. The ith relational network can be represented as an undirected graph $G_i = (V_i, E_i, R_i, W_i)$ for $1 \leq i \leq l$, where $V_i = \{v_1, v_2, v_3, \ldots, v_n\}$ represents the set of users in the ith relational social network; $E_i = \{e_1, e_2, e_3, \ldots, e_n\}$ represents the set of edges between nodes in the ith relational network; $R_i = (r_1, r_2, r_3, \ldots, r_l)$ represents the set of relationships among users at each layer in the social network, and their values belong to [0, 1]; and $W_i = \{W_{1i}, W_{2i}, W_{3i}, \ldots, W_{ni}\}$ represents the set of weights in the ith relational social network.

Integrated with the l relational networks, the *multirelationship network* is defined, where $V = \{v_j | v_j \in V(G_i), 1 \leq i \leq l\}$, $I = \{E_{ab} \subseteq V_a * V_b; a, b = 1, 2, \ldots, l, a \neq b\}$, $W = \{W_i | 1 \leq i \leq |V|\}$, and $W_i = \{W_{1i}, W_{2i}, W_{3i}, \ldots, W_{li}\}$. Specifically, E represents the set of interlayer edges connected between the ath and bth relational networks; V_a and V_b, respectively, represent the nodes of the ath and bth relational networks; and W is the set of the weights of all user pairs in different relationships. In order to better measure the weights of the relationships in the social network, their values will be normalized in the following. Figure 1 shows an example
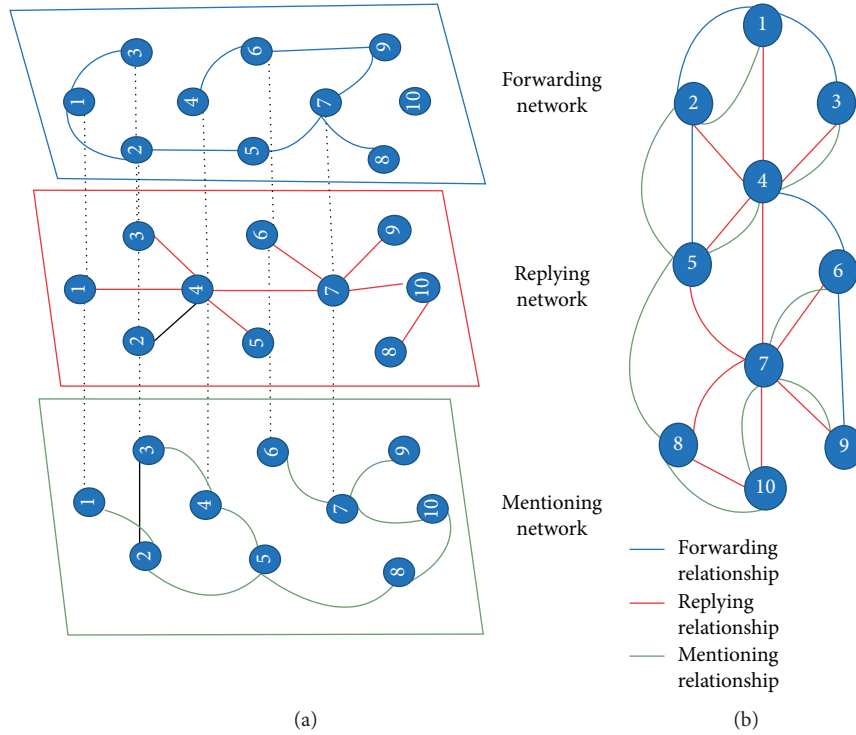
FIGURE 1: A social network and the multirelationship network.

of a social network and the multirelationship network. In Figure 1(a), a social network has 3-layer relational networks, where each layer network has 10 nodes, and the dotted lines represent the edges between layers.

Each layer represents a type of relationship; the first layer network is a forwarding network, where the network edges with the blue lines represent the relationships between users; the second layer is a replying network, with relationships between users within the network represented by red lines; and the third layer is a mentioning network, with relationships between users of the network represented by green lines. Assume that users in different relational networks are connected by a user participating in two or more relationships. Then all the relational networks aggregate the multirelationship network, as shown in Figure 1(b).

Definition 1 (Neighbors) [34]. The neighbors for a node u are denoted by $N(u) = \{v|uv \in E(G)\}$.

Definition 2 (Degree centrality) [34]. For a node $u \in G$, the degree centrality is the number of links incident upon node u; that is, $D(u) = |N(u)|$.

Definition 3 (Shortest path of a node) [12]. Let d_{iv} represent the distance from node i to node v. Then the sum of the distances from node i to the other nodes in a graph G, denoted by $C(i)$, is called the shortest path of node i; that is,

$$C(i) = \sum_{i \in V} d_{iv}. \tag{1}$$

Definition 4 (Shortest path in the graph G) [12]. Shortest path in graph G represents the sum of the shortest paths of all nodes of G. That is,

$$C(G) = \sum_{i \in V} C(i) = \sum_{v \in V} \sum_{i \in V} d_{iv}. \tag{2}$$

2.2. Overlapping User Relational Networks. In social networks, users actively participate in different relational networks. These users are called overlapping users. The subgraph induced by the overlapping users in the ith relational network is called the ith overlapping user relational network for $1 \leq i \leq l$. The network that is aggregated by all overlapping user relational networks is called multirelationship network. In the following, for convenience, the ith overlapping user relational network for $1 \leq i \leq l$ is briefly called the ith relational network.

In the kth relationship network $(1 \leq k \leq l)$, the weight of the edge between users i and j is defined as

$$w_{ij}^k = \begin{cases} 1, & ij \in E_k(G_k), \\ 0, & \text{others.} \end{cases} \tag{3}$$

Then, the algorithm of overlapping user relational networks that are abstracted from a social network with relational networks $G_i' = (V_i', E_i', R_i', W_i')$ $(1 \leq i \leq l)$ is represented by Algorithm 1.

2.3. Establishing Multirelationship Network. In a social network containing l relational networks, the weight of the edge between users i and j is calculated as follows:

> **Input:** Relational networks in the social network $G'_i(V'_i, E'_i, R'_i, W'_i)$, $1 \le i \le l$
> **Output:** Each overlapping relational network: $G_i(V_i, E_i, R_i, W_i)$, $1 \le i \le l$
> (1) $V_i \longleftarrow \varnothing$, $E_i \longleftarrow \varnothing$, $W_i \longleftarrow \varnothing$, $R_i \longleftarrow \varnothing$
> (2) **for** each graph G_i, $1 \le i \le l$ **do**
> (3) $V_{\mathrm{common}} = V(G_1) \cap V(G_2) \cap \ldots \cap V(G_l)$
> (4)   **for** $v \in V_{\mathrm{common}}$ **do**
> (5)   **for** $v \in N(u)$ in each graph G_i, $1 \le i \le l$ **do**
> (6) $E_i \longleftarrow E_i \cup (u, v)$
> (7) $V_i \longleftarrow V_i \cup \{u, v\}$
> (8) **for** each edge $uv \in E_i$ **do**
> (9)   the weight of the edge (u, v) in the ith overlapping relational network: $w^i_{uv} \longleftarrow 1$
> (10) $R_i \longleftarrow 1$
> (11) Return $G_i(V_i, E_i, R_i, W_i)$

ALGORITHM 1: Overlapping user relational network.

$$W_{ij} = \sum_{k=1}^{l} w^k_{ij}. \tag{4}$$

Based on the definition of the multirelationship network, we can further establish a multirelationship network with weights of edges through aggregating the l relational networks. For example, Figure 2 gives four relational networks: following, forwarding, replying, and mentioning networks. Figure 3 illustrates the process of aggregating the four relational networks and the multirelationship network. The multirelationship network can be established by Algorithm 2.

2.4. Key Node Discovery of the Multirelationship Network. Let $G \backslash i$ be the graph removing node i. Based on the definition of the shortest path for a node, the shortest path increment for a node i, denoted by $SPIS_i$, is the following:

$$SPIS_i = C(G \backslash i) - (C(G) - C(i)) = C(G \backslash i) + C(i) - C(G). \tag{5}$$

Let DC_i and PR_i denote the degree centrality [6] and PageRank value [14] of a node i in graph G, respectively; and let $Sd(\cdot)$ denote 0-1 normalization method. Then, we have

$$Sd(SPIS_i) = \frac{SPIS_i - \min_{i \in V(G)} SPIS_i}{\max_{i \in V(G)} SPIS_i - \min_{i \in V(G)} SPIS_i}, \tag{6}$$

$$Sd(DC_i) = \frac{D_i}{|V(G)| - 1}, \tag{7}$$

$$Sd(PR_i) = \frac{1 - a}{|V(G)|} + a \sum_{j \in N(i)} \frac{PR(j)}{|N(i)|}, \tag{8}$$

where D_i stands for the degree of node i; and a is the damping coefficient, generally 0.85.

Since a multirelationship network is aggregated by different relational social networks through the overlapping users, the node influence score of a user i in the network can be calculated by combining with the degree centrality, PageRank value, and shortest path increment as follows:

$$INF_i = \alpha * Sd(SPIS_i) + \beta * Sd(DC_i) + \gamma * Sd(PR_i), \tag{9}$$

where α, β, and γ are parameters and $\alpha + \beta + \gamma = 1$.

The shortest path increment is based on global features; the degree centrality is based on local features; and PageRank is a random walk feature. Thus, this formula combines both local and global features.

In multirelationship network, based on the weight of edge ij in the kth relationship network ($1 \le k \le l$), the node weight of node i, denoted by $NW(i)$, can be defined as follows:

$$NW(i) = \sum_{j \in N(i)} w^k_{ij}. \tag{10}$$

Considering the edges of each relational network and the characteristics of the network structure of different relational networks, the final node influence score of a node i, denoted by $INF_{final}(i)$, is aggregated with the node weight and the node influence score in the multirelationship network.

$$INF_{final}(i) = NW(i) * INF_i = \sum_{j \in N(i)} w^k_{ij} (\alpha * Sd(SPIS_i) + \beta * Sd(DC_i) + \gamma * Sd(PR_i)), \tag{11}$$

where $\alpha + \beta + \gamma = 1$.

By the algorithms on the overlapping user relational network and multirelationship network, the key node discovery algorithm based on multiple relationships and multiple features of social networks (KND algorithm) can be stated by Algorithm 3.

3. Experiments

In this section, we will give the experimental dataset, baseline methods, and evaluation experiments of KND algorithm.

3.1. Datasets. Four datasets are employed to verify the performance of the proposed KND algorithm. They are classified into two groups: single-relationship networks and multirelationship networks. The single-relationship
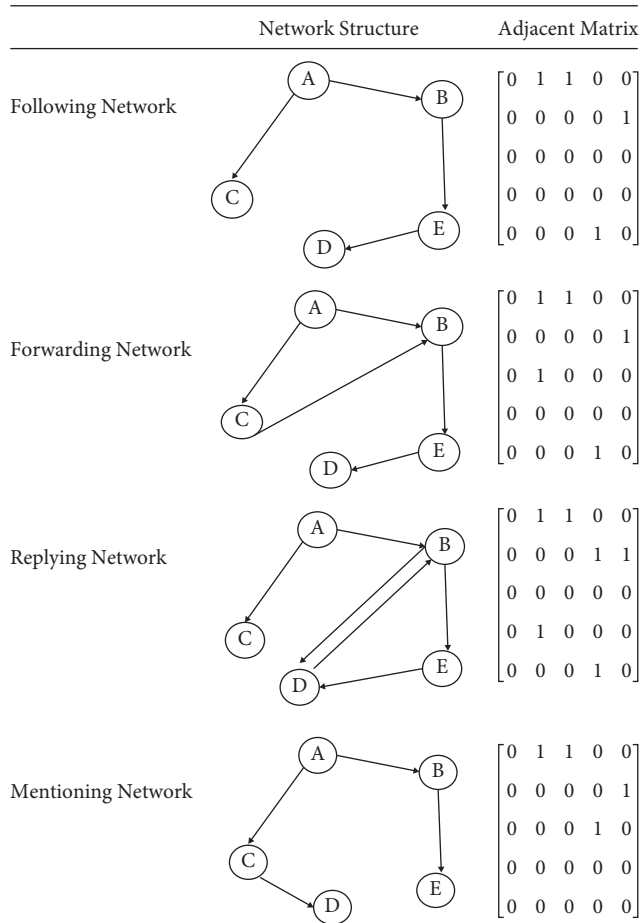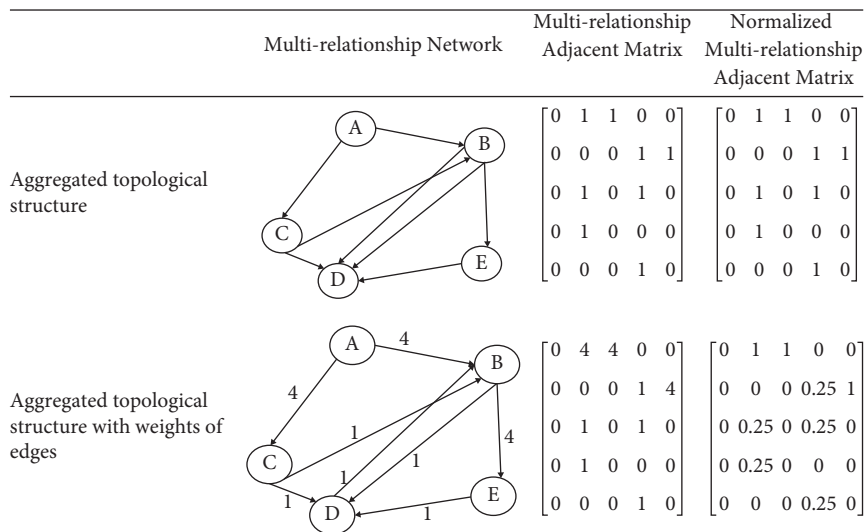
FIGURE 2: Four relational networks.

FIGURE 3: Multirelationship network.

networks include two LFR artificial synthetic networks and a karate club network. The multirelationship network is Higgs Twitter network, including following, forwarding, replying, and mentioning relational networks. The statistics of the two groups of datasets are shown in Tables 1 and 2.

Input: $G_i(V_i, E_i, R_i, W_i)$ $(1 \leq i \leq l)$ obtained from Algorithm 1
Output: Multirelationship network $G(V, E, R, W)$
(1) $V \longleftarrow \varnothing, E \longleftarrow \varnothing, W \longleftarrow \varnothing, R \longleftarrow \varnothing$
(2) **for** each graph G_i, $1 \leq i \leq l$ **do**
(3)   **for** each edge $(u, v) \in E_i$ **do**
(4) W_{uv} is updated by formula (4)
(5) $E \longleftarrow E \cup (u, v)$
(6) $V \longleftarrow V \cup \{u, v\}$
(7)   **end for**
(8) **end for**
(9) $\text{Max}_W = \max_{uv \in E(G)} (W_{uv})$
(10) **for** each edge $(u, v) \in E$ **do**
(11) $W_{uv} \longleftarrow (W_{uv}/\text{Max}_W)$
(12) $R \longleftarrow 1$
(13) **end for**
(14) Return $G(V, E, R, W)$

ALGORITHM 2: Establishing multirelationship network.

Input: $G(V_i, E_i, R_i, W_i)$, $1 \leq i \leq l$: which obtained from Algorithm 1
Output: top-K nodes: S
(1) $V \longleftarrow \varnothing, E \longleftarrow \varnothing, W \longleftarrow \varnothing, R \longleftarrow \longleftarrow$
(2) **for** each graph $G(V_i, E_i, R_i, W_i), 1 \leq k \leq l$ **do**
(3)   **for** each edge $ij \in E_k$ **do**
(4) $\text{NW}(i) \longleftarrow \sum_{j \in N(i)} w_{ij}^k$ by formula (10)
(5)   **end for**
(6) **end for**
(7) Obtain the multirelationship network $G(V, E, R, W)$ by Algorithm 2
(8) Compute SPIS_i by formula (6)
(9) Compute DC_i by formula (7)
(10) Compute PR_i by formula (8)
(11)   Compute INF_i by formula (9)
(12) **for** $i \in V(G)$ **do**
(13) Compute $\text{INF}_{\text{final}}(i) = \text{NW}(i) * \text{INF}_i$ by formula (11)
(14) **end for**
(15) Sort $(\{\text{INF}_{\text{final}}(i)|i \in V(G)\})$
(16) Return top-K nodes S

ALGORITHM 3: Key node discovery algorithm.

TABLE 1: Statistics of single networks.

Dataset	Number of nodes	Number of edges
Karate club network	34	77
LFR-500 network	500	1657
LFR-1000 network	1000	3200

TABLE 2: Statistics of Higgs Twitter network.

Relational networks	Following	Forwarding	Replying	Mentioning
Number of nodes	456626	256491	38918	116408
Number of edges	14855842	328132	32523	150818

3.1.1. Karate Club Network (https://blog.csdn.net/weixin_41857995/article/details/105454517).

It was a friendship social network between 34 members of a college karate club in a college of America.

3.1.2. LFR Artificial Synthesis Networks (LFR Networks) [41].

Since the networks follow power law distribution, they can be used to simulate real networks. If the LFR network has M nodes, then the network is denoted by LFR-M network.

3.1.3. Higgs Twitter Network (http://snap.stanford.edu/data/higgs-twitter.html).

The twitters contained news about the discovery of new particles with Higgs boson features before, on, and after July 4, 2012. This dataset has four types of

networks: following, forwarding, replying, and mentioning networks.

3.2. Baseline Methods.

To evaluate the accuracy of the KND algorithm, six baseline methods are selected as follows:

(1) PageRank (PR) [14]: The initial values of nodes are given the same score. Then each value is constantly updated by an iteration formula of PR. When the iteration results tend to converge in a stable state, the top-K nodes with higher scores are selected as the key node.

(2) Degree centrality (DC) [11]: It computes the centrality of a network. Then, the top-K nodes with high centrality are selected as the key nodes.

(3) SPIS [12]: It computes the shortest path increments of all nodes in networks. Then, the nodes that have the highest top-K shortest path increments are selected as the key nodes.

(4) MCIM [42]: This method considers four metrics to evaluate a node. The overlapping influence and the influence between nodes and their neighbors are both considered to evaluate a node. Then, the nodes with top-K evaluation values are selected as the key nodes.

(5) Eigenvector [17]: The importance of a node depends on both the number of neighbor nodes and the importance of its neighbor nodes. Both the topology and the properties of the node are considered. Then the nodes with top-K eigenvector scores are selected as the key nodes.

(6) Random [43]: It randomly selects K nodes as the key nodes.

3.3. Evaluation Experiments of KND Algorithm.

The SIR (Susceptible-Infected-Removed) model [9] is adopted to compare transmission ability on the KND algorithm and baseline methods. The susceptible population S is converted to the infected population with the probability η, and the infected population I recovers to be immune to the information with probability ξ. When there are no infected nodes in the network or a preset number of iterations is reached, the propagation process stops in networks. Based on different sizes of dataset, we adopt different values of η and ξ according to [44]: $\eta_{min} \approx \langle k \rangle / \langle k^2 \rangle$, where $\langle k \rangle$ indicates average degree in the whole network and $\langle k^2 \rangle$ indicates mean of the squared degrees in all nodes of the network. For convenient comparison, we set $\xi = 0.01$ and $\eta := \eta + 0.02$. The values of η for each network are shown in Table 3.

Moreover, parameters α, β, and γ are set as 0.4, 0.2, and 0.4, respectively. Considering the fact that all experiments are probabilistic experiments, we get 50 times of calculations each time and take their average result as the final result to avoid the effect of random error. For the Higgs Twitter network, to reduce the complexity of the experiments, we only consider three important relationships: replying,

TABLE 3: Value of parameter η in different networks.

Networks	η	$\eta + 0.02$
Karate club network	0.303748	0.323748
LFR-500	0.352644	0.372644
LFR-1000	0.376703	0.396703
Replying network	1.105682	1.125682
Mentioning network	0.678706	0.698706
Forwarding network	0.726671	0.746671
Higgs multirelationship network	0.007877	0.027877

mentioning, and forwarding. The three relational networks are aggregated as the Higgs multirelationship network by Algorithm 2. Under the above parameters, we make some comparison experiments for the KND, DC, PR, Eigenvector, MCIM, SPIS, and Random methods on the karate club, LFR-500, LFR-1000, replying, mentioning, forwarding, and Higgs multirelationship networks. The more infected nodes are, the stronger transmission ability of the initial infected nodes is. When the initial infected nodes are the same, the more infected nodes are, the better the performances of the methods are.

From Figure 4, on the karate club network, whatever the initial numbers of infected nodes are, the total number of infected nodes stays the same, close to 34. The reason is that the size of the karate club network is too small; and the initial infected nodes easily infect other nodes of the network. The performances of the KND, DC, PR, Eigenvector, MCIM, and SPIS methods are almost the same and better than the performance of the Random method on the karate club network.

On the LFR-500 and LFR-1000 networks, with the increase of initial numbers of infected nodes, the total numbers of infected nodes for the KND and baseline methods are going up; the performance of the KND method is the best; and the performance of the Random method is the worst. Since the sizes of the two networks are different, at the same initial number of infected nodes, the total number of infected nodes on LFR-500 is more than that on LFR-1000. It indicates that the more the number of nodes in the network is, the weaker the transmission ability is under the same initial number of infected nodes.

On the replying, mentioning, and forwarding networks, the performance of the KND method is better compared to all baseline methods; the Random method gets the worst performance. On the replying and mentioning networks, as the initial numbers of infected nodes increase, the total numbers of infected nodes steadily rise. However, on the forwarding network, the total numbers of infected nodes rise from 5 to 10; from 10 to 20, they keep about 280; from 20 to 25, they go through a descent; and then they increase slowly. Therefore, the KND method on the three single networks has the best performance overall.

On the Higgs multirelationship network, from 5 to 10, the performance of the MCIM method is the best, while the performance of the KND method is the second; from beginning to end, the total numbers of infected nodes continuously rise; from about 10 to end, the performance of the KND method is the best; and, from beginning to end, the Eigenvector method is the worst.

(a)

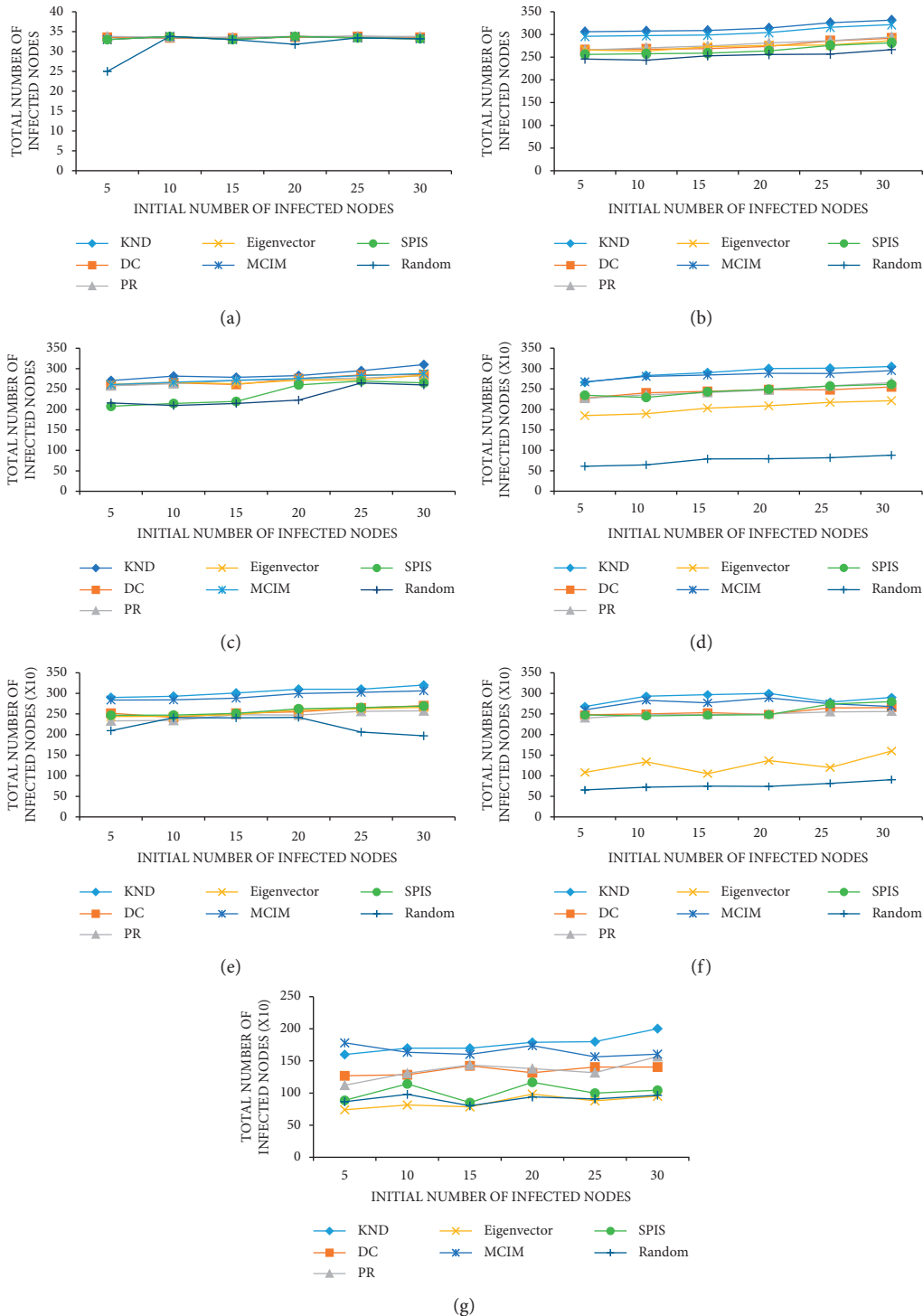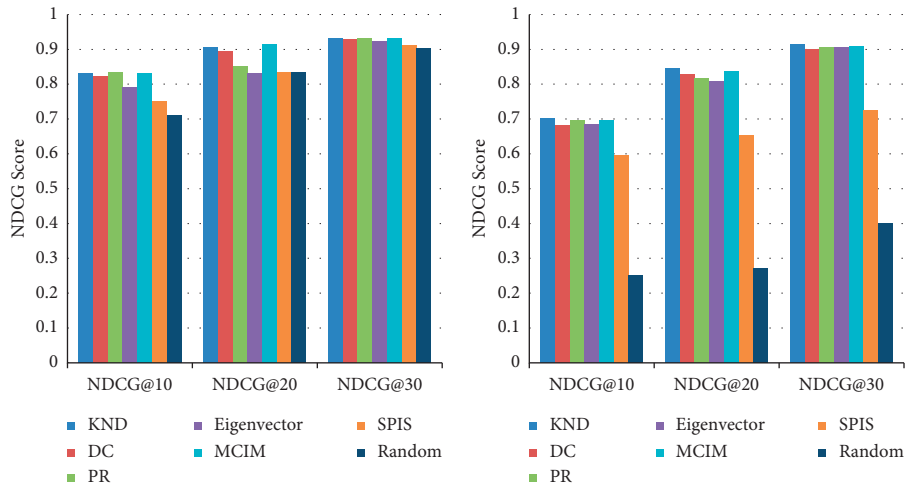
(b)

(c)

(d)

(e)

(f)

(g)

FIGURE 4: Total numbers of infected nodes for all methods on different networks. (a) Karate club network. (b) LFR-500 network. (c) LFR-1000 network. (d) Replying network. (e) Mentioning network. (f) Forwarding network. (g) Higgs multirelationship network.

On the replying, mentioning, forwarding, and Higgs multirelationship networks, when the initial numbers of infected nodes are the same, the total number of infected nodes on the Higgs multirelationship network is lower than those on the replying, mentioning, and forwarding networks. This implies that the transmission abilities on single networks are stronger than those on multirelationship

network. This is because real social networks always contain multiple relationships and features, which reduces the transmission capacity to some extent. Therefore, the whole performance of the KND method is superior to the baseline methods on single-relationship and multirelationship networks; and the multirelationship network can indicate social networks better.
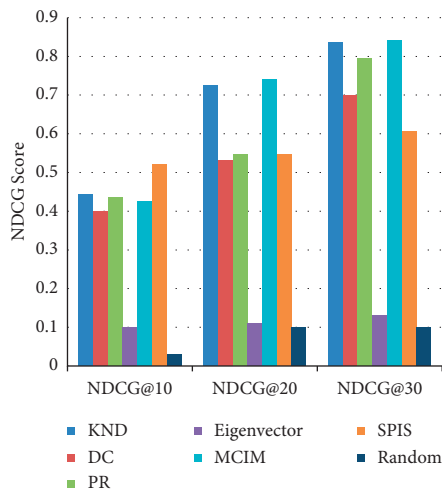
(a)

(b)
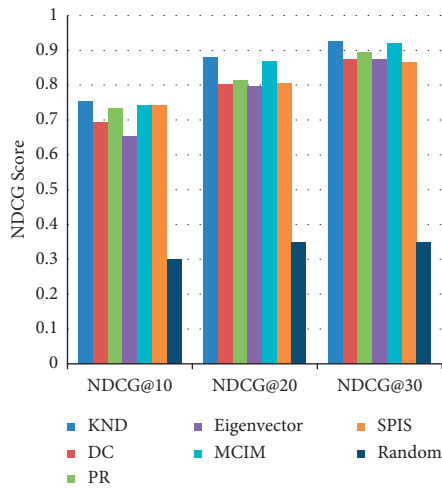
(c)

(d)

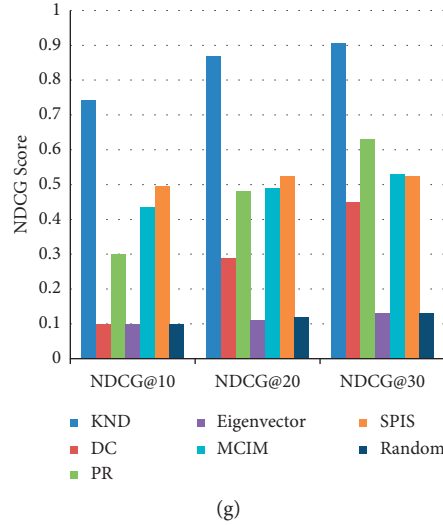(e)

(f)

FIGURE 5: Continued.

(g)

FIGURE 5: NDCG scores on different networks. (a) Karate club network. (b) LFR-500 network. (c) LFR-1000 network. (d) Replying network. (e) Mentioning network. (f) Forwarding network. (g) Higgs multirelationship network.

Summing up these discussions, compared with all baseline methods, the proposed KND method can effectively discover the key nodes in social networks.

3.4. Evaluation Indicator and Its Analysis.
To evaluate the sorting qualities of key nodes (top-K nodes) for all methods, the normalized discounted cumulative gain (NDCG) [45] is adopted. Suppose that NDCG@n denotes the normalized discounted cumulative gain of the first n nodes; DCG_n represents the cumulative loss gain of the nodes; and $IDCG_n$ represents the maximum DCG_n in the ideal case. Then

$$NDCG@n = \frac{DCG_n}{IDCG_n}, \qquad (12)$$

where $DCG_n = \sum_{i=1}^{n} Rl(i)/\log(i+1)$; and $Rl(i)$ represents the correlation between node i and the final result.

Figure 5 shows the values of NDCG@n on different networks. From Figure 5(a), on the karate club network, when $n = 10, 30$, the KND, PR, and MCIM methods get the largest NDCG scores. When $n = 10, 20, 30$, the Random method is the worst. When $n = 20$, the NDCG score of the MCIM method is the highest, whereas the NDCG score of the KND method is very close to the highest value. When $n = 30$, the NDCG scores of all methods are slightly different. The reason is that the size of the network is very small. It can be concluded that the advantage of the KND method is not obvious in small size networks.

From Figure 5(b), on the LFR-500 network, when $n = 10, 20, 30$, the KND method is the best; the Random method is the worst; and the NDCG scores of the KND, DC, PR, Eigenvector, and MCIM methods are subtly different. They show that the KND method is the best.

From Figure 5(c), on the LFR-1000 network, when $n = 10, 20, 30$, the NDCG scores of the KND are the largest; and the NDCG score of the Random method is the smallest. When $n = 10$, the NDCG scores of the PR, MCIM, and SPIS

methods are slightly different. When $n = 30$, the DC, Eigenvector, and SPIS methods almost obtain the same NDCG@30. They show that the KND method can take the best performance on the LFR-1000 network.

From Figure 5(d), on the replying network, when $n = 10$, the NDCG score of the KND method is the same as that of the MCIM, the highest. When $n = 20$, the DC method gets the largest NDCG@10, but the NDCG score of the KND method is slightly lower than that of the DC method. When $n = 30$, the KND and MCIM methods obtain the highest NDCG@30. When $n = 10, 20, 30$, the NDCG scores for Random method are the lowest. These imply that the overall performance of the KND method is superior to those of the other methods on the replying network.

From Figure 5(e), on the mentioning network, when $n = 10$, the SPIS method gets the largest NDCG@10; and the KND method obtains the second NDCG@10. When $n = 20$, the NDCG score of the MCIM method is the largest and slightly more than KND method. When $N = 30$, the NDCG scores of the KND and MCIM methods are almost the same and the largest. When $n = 10, 20, 30$, the Random method gets the worst performance. These imply that the more the value of n is, the better the KND method obtains the performance on mentioning network.

From Figure 5(f), on the forwarding network, when $n = 10, 20, 30$, the NDCG scores of the KND, DC, PR, MCIM, and SPIS methods are slightly different; the KND method gets the best performance; and the Random method obtains the worst performance. Thus, the performance of the KND method is the best on the forwarding network.

From Figure 5(g), on the Higgs multirelationship network, when $n = 10, 20, 30$, the NDCG score for the KND method far exceeds all baseline methods. Thus, the KND method can get the best performance on the Higgs multirelationship network.

Therefore, the overall sorting quality of the KND method is superior compared to all baseline methods on the single-

relationship network and multirelationship network. In particular, the sorting quality of the KND method is far more than all baseline methods on the multirelationship network. It is concluded that the KND method is very helpful to find the key nodes in social networks.

4. Conclusion

In social networks, a user always has multiple relationships and features. Considering the fact that the different relationships constitute different relational networks, this paper proposed the algorithm of overlapping user relational network to find different relational network consisting of the same nodes in social networks. Based on these relational networks, we first proposed the algorithms of the overlapping user relational network and multirelationship network. Then, we proposed the key node discovery algorithm with multiple relationships and multiple features to find the top-K nodes in social networks. The experiments of the KND algorithm and six baseline methods on karate club, LFR-500, LFR-1000, replying, mentioning, forwarding, and multirelationship networks show that the KND algorithm can obtain the best performance on the multirelationship network. The key node discovery methods with multiple relationships, multiple features, and users' attributions are worth research directions in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] M. A. Al-Garadi, M. S. Khan, K. D. Varathan, G. Mujtaba, and A. M. Al-Kabsi, "Using online social networks to track a pandemic: a systematic review," *Journal of Biomedical Informatics*, vol. 62, pp. 1–11, 2016.

[2] M. A. Al-Garadi, K. D. Varathan, and S. D. Ravana, "Cybercrime detection in online communications: the experimental case of cyberbullying detection in the twitter network," *Computers in Human Behavior*, vol. 63, pp. 433–443, 2016.

[3] V. Chang, "A cybernetics social cloud," *Journal of Systems and Software*, vol. 124, pp. 195–211, 2017.

[4] W. Liu, X. Wu, W. Yang, X. Zhu, and S. Zhong, "Modeling cyber rumor spreading over mobile social networks: a compartment approach," *Applied Mathematics and Computation*, vol. 343, pp. 214–229, 2019.

[5] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, no. 20, Article ID 208701, 2002.

[6] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.

[7] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 199–208, Paris, France, June 2009.

[8] A. Sheikhahmadi, M. A. Nematbakhsh, and A. Shokrollahi, "Improving detection of influential nodes in complex networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 436, pp. 833–845, 2015.

[9] X. Wang, Y. Su, C. Zhao, and D. Yi, "Effective identification of multiple influential spreaders by degreepunishment," *Physica A: Statistical Mechanics and Its Applications*, vol. 461, pp. 238–247, 2016.

[10] D.-B. Chen, H. Gao, L. Lü, and T. Zhou, "Identifying influential nodes in large-scale directed networks: the role of clustering," *PLoS One*, vol. 8, no. 10, Article ID e77455, 2013.

[11] P. Bonacich, "Some unique properties of eigenvector centrality," *Social Networks*, vol. 29, no. 4, pp. 555–564, 2007.

[12] J. Zhu, L. Zhu, C. Bao, L. Zhou, C. Wang, and B. Kong, "Top-k structure holes detection Algorithm in Social network," in *Proceedings of the 2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 1064–1071, Hangzhou, China, October 2018.

[13] X. Chen, M. Tan, J. Zhao, T. Yang, D. Wu, and R. Zhao, "Identifying influential nodes in complex networks based on a spreading influence related centrality," *Physica A: Statistical Mechanics and Its Applications*, vol. 536, Article ID 122481, 2019.

[14] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1–7, pp. 107–117, 1998.

[15] L. Lü, Y. C. Lü, C. H. Yeung, and T. Zhou, "Leaders in social networks, the delicious case," *PLoS One*, vol. 6, no. 6, Article ID e21202, 2011.

[16] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pp. 668–677, 1998.

[17] M. Kitsak, L. K. Gallos, S. Havlin et al., "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.

[18] J. Bae and S. Kim, "Identifying and ranking influential spreaders in complex networks by neighborhood coreness," *Physica A: Statistical Mechanics and Its Applications*, vol. 395, pp. 549–559, 2014.

[19] J.-G. Liu, Z.-M. Ren, and Q. Guo, "Ranking the spreading influence in complex networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 392, no. 18, pp. 4154–4159, 2013.

[20] Z. Ye, X. Zhan, Y. Zhou, C. Liu, and Z.-K. Zhang, "Identifying vital nodes on temporal networks: an edge-based k-shell decomposition," in *Proceedings of the 2017 36th Chinese Control Conference (CCC)*, pp. 1402–1407, Dalian, China, July 2017.

[21] A. Zeng and C.-J. Zhang, "Ranking spreaders by decomposing complex networks," *Physics Letters A*, vol. 377, no. 14, pp. 1031–1035, 2013.

[22] M. Chen, Y. Huang, S.-H. Yu, and C.-H. Wu, "Lucky reply effect," *Journal of Organizational and End User Computing*, vol. 32, no. 4, pp. 162–174, 2020.

[23] W. Liu and S. Zhong, "Web malware spread modelling and optimal control strategies," *Scientific Reports*, vol. 7, no. 1, pp. 42308–42319, 2017.

[24] Y.-Q. Zhu and B. Hsiao, "What attracts followers?" *Journal of Organizational and End User Computing*, vol. 33, no. 1, pp. 71–91, 2021.

[25] F. Battiston, V. Nicosia, and V. Latora, "Structural measures for multiplex networks," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 89, no. 3, Article ID 032804, 2014.

[26] S. Boccaletti, G. Bianconi, R. Criado et al., "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, no. 1, pp. 1–122, 2014.

[27] M. A. Al-Garadi, K. D. Varathan, S. D. Ravana, E. Ahmed, and V. Chang, "Identifying the influential spreaders in multilayer interactions of online social networks," *Journal of Intelligent and Fuzzy Systems*, vol. 31, no. 5, pp. 2721–2735, 2016.

[28] A. Zareie and A. Sheikhahmadi, "A hierarchical approach for influential node ranking in complex social networks," *Expert Systems with Applications*, vol. 93, pp. 200–211, 2018.

[29] D. Wang, H. Wang, and X. Zou, "Identifying key nodes in multilayer networks based on tensor decomposition," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 27, no. 6, Article ID 063108, 2017.

[30] M. Li, Q. Zhang, and Y. Deng, "Evidential identification of influential nodes in network of networks," *Chaos, Solitons & Fractals*, vol. 117, pp. 283–296, 2018.

[31] F. Pedroche, L. Tortosa, and J. F. Vicent, "An eigenvector centrality for multiplex networks with data," *Symmetry*, vol. 11, no. 6, p. 763, 2019.

[32] S. Wang, J. Liu, and Y. Jin, "Finding influential nodes in multiplex networks using a memetic algorithm," *IEEE Transactions on Cybernetics*, vol. 99, pp. 1–13, 2019.

[33] G. Fu, C. Hou, and X. Yao, "Learning topological representation for networks via hierarchical sampling," in *Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Budapest, Hungary, 2019.

[34] S. S. Singh, K. Singh, A. Kumar, and B. Biswas, "MIM2: multiple influence maximization across multiple social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 526, Article ID 120902, 2019.

[35] X. Huang, D. Chen, D. Wang, and T. Ren, "MINE: identifying top-k vital nodes in complex networks via maximum influential neighbors expansion," *Mathematics*, vol. 8, no. 9, p. 1449, 2020.

[36] F. Emmert-Streib and M. Dehmer, "Identifying critical financial networks of the DJIA: toward a network-based index," *Complexity*, vol. 16, no. 1, pp. 24–33, 2010.

[37] F. Morone, G. Del Ferraro, and H. A. Makse, "The k-core as a predictor of structural collapse in mutualistic ecosystems," *Nature Physics*, vol. 15, no. 1, pp. 95–102, 2019.

[38] F. Morone and H. A. Makse, "Correction: corrigendum: Influence maximization in complex networks through optimal percolation," *Nature*, vol. 527, no. 544, pp. 65–68, 2015.

[39] N. Musmeci, V. Nicosia, T. Aste, T. Di Matteo, and V. Latora, "The multiplex dependency structure of financial markets," *Complexity*, vol. 2017, Article ID 9586064, 13 pages, 2017.

[40] M. Zhou, Y. Long, W. Zhang et al., "Adaptive genetic algorithm-aided neural network with channel state information tensor decomposition for indoor localization," *IEEE Transactions on Evolutionary Computation*, p. 1, 2021.

[41] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 78, no. 4, Article ID 046110, 2008.

[42] Y. Du, C. Gao, Y. Hu, S. Mahadevan, and Y. Deng, "A new method of identifying influential nodes in complex networks based on TOPSIS," *Physica A: Statistical Mechanics and Its Applications*, vol. 399, pp. 57–69, 2014.

[43] T. Nie, Z. Guo, K. Zhao, and Z.-M. Lu, "Using mapping entropy to identify node centrality in complex networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 453, pp. 290–297, 2016.

[44] C. Castellano and R. Pastor-Satorras, "Thresholds for epidemic spreading in networks," *Physical Review Letters*, vol. 105, no. 21, Article ID 218701, 2010.

[45] K. Järvelin and J. Kekäläinen, "IR evaluation methods for retrieving highly relevant documents," *ACM SIGIR Forum*, vol. 51, no. 2, pp. 243–250, 2017.