Hindawi
Complexity
Volume 2020, Article ID 5903798, 15 pages
https://doi.org/10.1155/2020/5903798

WILEY | Hindawi

Research Article

Ranking Influential Nodes in Complex Networks with Information Entropy Method

Nan Zhao,[1] **Jingjing Bao**,[1,2] **and Nan Chen**[1]

[1]*State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China*
[2]*College of Computer Science and Technology, Inner Mongolia Normal University, Hohhot 010022, China*

Correspondence should be addressed to Nan Zhao; zhaonan@xidian.edu.cn

Received 23 March 2020; Accepted 22 April 2020; Published 8 June 2020

Guest Editor: Hongshu Chen

The ranking of influential nodes in networks is of great significance. Influential nodes play an enormous role during the evolution process of information dissemination, viral marketing, and public opinion control. The sorting method of multiple attributes is an effective way to identify the influential nodes. However, these methods offer a limited improvement in algorithm performance because diversity between different attributes is not properly considered. On the basis of the k-shell method, we propose an improved multiattribute k-shell method by using the iterative information in the decomposition process. Our work combines sigmod function and iteration information to obtain the position index. The position attribute is obtained by combining the shell value and the location index. The local information of the node is adopted to obtain the neighbor property. Finally, the position attribute and neighbor attribute are weighted by the method of information entropy weighting. The experimental simulations in six real networks combined with the SIR model and other evaluation measure fully verify the correctness and effectiveness of the proposed method.

1. Introduction

Multidimensional information flows rapidly on the network, while different nodes have different effects on information transmission [1], viral marketing [2], public opinion guidance [3], and social recommendation [4, 5] due to their different influences. From the perspective of information transmission, different social networks have different modes of information transmission because of the diversity of functional focuses and user structures. From the perspective of marketing, by providing influential user rankings for different hobbies and groups, it can help new users quickly and effectively obtain relevant information sources of interest so as to achieve a smooth cold start. From the perspective of public opinion guidance and control, the event evolution process of hot public opinions often includes the forwarding and comments of users with different influences on different platforms. These simple operations often lead to an enormous development of public opinions in different directions.

The influence of nodes is evaluated from global structure information, such as betweenness centrality [6], closeness centrality [7], and Katz centrality [8]. These methods show good performance in node sorting. However, because of $O(n^2)$ or even higher computational complexity, these methods are not suitable for large-scale networks. The influence of nodes is quantified by local information, such as degree centrality [9], semilocal centrality [10], hybrid degree centrality [11], average shortest path centrality [12], and h index [13]. Local measures are less efficient because they only consider local neighborhood information. There are many heuristic algorithms [14, 15] combined with local neighborhood information. Research based on random walk evaluates the influence of nodes through multiple iterative operations with high-computational complexity such as feature vector centrality [16], PageRank [17], LeaderRank [18], and Hits [19].

Kitsak et al. [20] argues that the most influential nodes are those located at the core of the network. Each node is assigned a fixed shell value after k-shell decomposition.

However, k-shell decomposition tends to assign the same shell value to many nodes so that the influence of these nodes with same shell value cannot be further distinguished. On this basis, plenty of methods have been proposed to further improve the performance of k-shell method. Zeng and Zhang [21] propose a mixed degree decomposition method, which combines the residual degree and depletion degree to update the nodes. In each step of decomposition, the nodes are removed and decomposed based on the mixed degree. However, the λ parameter is difficult to optimize. Liu et al. [22] proposes an improved ranking method to generate a more differentiated ranking list. This method is realized by calculating the shortest distance between the target node and the core node of the network. The core nodes of the network are in a node set with the highest shell value in k-shell decomposition. The computational cost of this method is relatively expensive by calculating the shortest distances to the core nodes. Bae and Kim [23] propose a new measurement of neighborhood coreness centrality, which calculates the diffusion influence of nodes in the network by summing all neighborhood shell values. The influence of nodes with the same *ks* value can be further distinguished by using the iterative information of removal nodes to identify the position difference of nodes in the network. The degree decomposition method based on iteration factor [24] is to improve the performance of the traditional method by using the iteration information and node degree in the decomposition. In addition, some other node-sorting algorithms were introduced to improve the sorting performance.

On the basis of the k-shell method, our work makes full use of the iterative information in the decomposition process and proposes an improved multiattribute k-shell method. First, the iteration information is processed by sigmod function to obtain the position index. Then, the position attribute is captured by combining the shell value and the position index. The local information of the node is adapted to obtain the neighbor property. Finally, the position attribute and neighbor attribute are weighted by the method of information entropy weighting. In the experiment, the SIR model, Kendall coefficient, and imprecision function are used to, respectively, evaluate the propagation capability of different probabilities of propagation, the imprecision of ranking and the correlation coefficient of different probability of propagation. Furthermore, we evaluate ranking results of the proposed method by selecting seeds in influence maximization problem and measuring the ranking uniqueness and distribution. The experimental results prove that the proposed method can effectively distinguish the differences of different attributes and significantly promote the performance of identifying the influence of nodes.

The following parts are organized as follows. We briefly review the definition of related algorithms used for comparison in Section 2. In Section 3, our improved multiattribute k-shell method is proposed and a meaningful example is illustrated to show how the proposed measure works. In Section 4, we present the details of the data, the spreading model, and the evaluation measure that are used to evaluate the performance of our measure. The experimental results are presented in Section 5. Finally, we expose the conclusion of the work in Section 6.

2. Related Work

Kitsak et al. proposed the k-shell method to determine the influence of nodes in the network. This method considers that the closer the node is to the center of the network, the higher the influence of the node will be. This method uses node degree to rank the importance of nodes. The following details show the decomposition principle of the k-shell method.

First, we need to remove the nodes and edges with a medium degree as 1 in the network. At this time, nodes with a degree as 1 still exist in the remaining network. We should continue to remove them until no nodes with a degree as 1 exist in the network. At this point, the removed nodes form a layer and its *ks* value is assigned to 1. According to the abovementioned method, continue to remove nodes with a degree value as 2 in the network and repeat this operation until there are no nodes in the network. As can be seen in Figure 1, the *ks* value allocated for nodes 8, 9, 10, 11, 12, 13, 14, 15, and 16 is 1. The value allocated for nodes 5, 6, and 7 is 2. The *ks* value allocated for nodes 1, 2, 3, and 4 is 3.

K-shell decomposition method is suitable for large networks because of low-computational complexity. Of course, the disadvantages of this method are obvious. First, most nodes are assigned the same *ks* value so that the importance of these nodes cannot be further distinguished. For example, from the perspective of degree, the degree of node 8 is 3 and the degree of node 11 is 1. The influence of node 8 is obviously greater than 11, but they have the same *ks* value. Second, in the process of removing nodes, the edges that have been removed are not considered and the influence of residue is only concerned about. In this way, it is considered that nodes with the same *ks* have the same number of edges in the outer layer, which is obviously not consistent with common sense. For example, node 1 and node 2 have abundant first-order and second-order neighbors in the outer layer, while node 1 has no neighbors in the outer layer. The same *ks* value assigned to them does not reflect the difference between them. Third, in the regular network, the *ks* value of most nodes is 1, which is obviously not suitable for this background.

The traditional K-shell decomposition method only updates the nodes according to the residual degree of the remaining nodes and completely ignores the depletion degree of the removed nodes. In the mixed degree decomposition method, the decomposition process is based on the residual degree of the remaining nodes and the depletion degree of the removed nodes. For node i, its residual degree and depletion degree are represented by k_i^r and k_i^e, respectively. In each step of the mixing degree decomposition method, the node removal is determined by the mixing degree k_i^m:

$$k_i^m = k_i^r + \lambda k_i^e, \tag{1}$$

where λ is the adjustable parameter between 0 and 1. When λ is 0, the mixing degree decomposition method is consistent

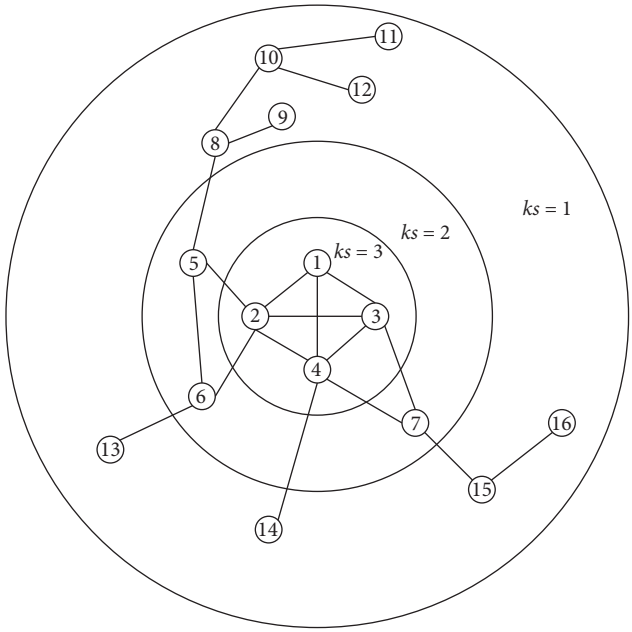

Figure 1: The schematic diagram of basic principles of the k-shell method.

as the K-shell decomposition method. When λ is 1, the mixed degree decomposition method is equivalent to the degree centrality method. Different from the traditional K-shell decomposition method, in the mixed degree decomposition method, the value of ks of all nodes can be decimal. The parameter λ is usually set to 0.7.

The traditional K-shell decomposition method is improved by using the shortest distance from the source node to the core node of the network. The propagation capability of nodes with the same ks value can be further distinguished by the following methods:

$$\theta(i) = \left(ks_{\max} - ks_i + 1\right) \sum_{j \in S_c} d_{ij}, \tag{2}$$

where $ks_{\max}$ is the maximum ks value in K-shell decomposition, ks_i is the ks value of node i, and S_c is the set of nodes whose ks value is maximum. Although this nonparametric method can identify nodes with the same ks value, the computational complexity is high because of calculation of the shortest path distance to the core. More seriously, if the network is not fully connected, the shortest path distance between partial node pairs cannot be obtained.

If the iterative information is used to identify the position difference of nodes in the network, the propagation ability of nodes with the same ks value can be further distinguished. Degree decomposition method based on the iteration factor is proposed to improve the performance of the traditional method by using iteration information and node degree in decomposition. It is worth noting that the degree is a local variable and the iteration factor is a global variable. This method fully combines local and global factors to identify influential nodes more effectively. The iterative factor δ_i of node i is

$$\delta_i = ks_i * \left(1 + \frac{\text{iter}(i)}{m}\right), \tag{3}$$

where m is the total number of iterations in K-shell decomposition and $\text{iter}(i)$ is the number of times that node i has been removed from the decomposition. The influence of node i in the degree decomposition method based on iterative factor is

$$\text{IC}_i = \delta_i * d_i + \sum_{j \in \Gamma(i)} \delta_j * d_j. \tag{4}$$

The influence of node will be great if a node has many neighbors at the core of the network. Based on this assumption, the neighbor core of the node is

$$C_{\text{nc}}(i) = \sum_{j \in \Gamma(i)} ks(j), \tag{5}$$

where $\Gamma(i)$ is the neighbor of node i. Recursively, the extended neighbor kernel is defined as

$$C_{\text{nc+}}(i) = \sum_{j \in \Gamma(i)} C_{\text{nc}}(j). \tag{6}$$

3. Materials and Methods

Practice has proved that the combination of multiple attributes can further improve the sorting effect. In recent years, researchers have combined different attributes and strategies to mine the influence of nodes. The performance of these methods proves that considering multiple attributes is an effective strategy to evaluate the impact capability of nodes. At present, there are many attribute weighting methods, such as least squares weighting method and principal component analysis method. Among the many attributes of nodes, location attributes play a significant role in the sorting process of nodes. At the same time, the influencing ability of nodes depends largely on the neighbor attributes. Combining these two attributes, it is an effective strategy to use the attribute weighting method to further identify the influence of nodes.

In the K-shell decomposition process, the number of iterations reveals very important location information, and it can further distinguish the location differences of the removed nodes. A node with a higher number of iterations is closer to the core of the network, or it is closer to the edge of the network. The number of iterations here refers to the number of global iterations decomposed from K-shell to the end. In this paper, sigmod function is used to further process the number of iterations when the node is deleted, so as to define the node position index $p(i)$:

$$p(i) = \frac{3}{4} \frac{1}{1 + e^{-\sqrt{\text{iter}(i)}}}. \tag{7}$$

The relationship between the position index and the number of iterations is shown in Figure 2. As the number of iterations increases, the position index increases in a downward slope with a critical value of 0.75.

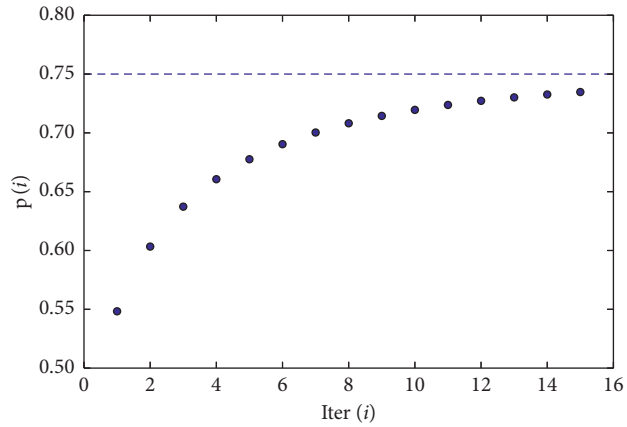

Figure 2: The diagram of the relationship between iteration number and position index.

The position attribute of the node is represented by $\mathrm{PN}_p(i)$, which is composed of the ks value of the node and the sum of the location index of the neighbor of the node:

$$\mathrm{PN}_p(i) = ks_i + \sum_{j\in\Gamma(i)} p(j), \tag{8}$$

where ks_i is the ks value of the node in the K-shell decomposition.

The position attribute of a node cannot effectively distinguish the influence of a node in the same position. For example, all nodes on the edge of a network have the same position information, so their influence should be the same. In fact, due to the difference of local structure, the influence of edge nodes in the same position will vary greatly. The local attributes of nodes should be further used to distinguish the influence differences of nodes with the same location attributes. If a node has more neighbors, it can have a greater impact on the network. Furthermore, the influence of a node's neighbor is also impacted by its neighbors. Consider the second-order neighbor can improve the capability of measuring the influence of a node.

The neighbor attribute of a node is represented by $\mathrm{PN}_n(i)$, and it is represented by the second-order neighbor number of the node:

$$\mathrm{PN}_n(i) = \sum_{j\in\Gamma(i)} \sum_{l\in\Gamma(j)} k_l, \tag{9}$$

where k_l represents the degree of node l.

Both position attribute and local attribute play a significant role in identifying the influence of nodes. Combining these two key attributes, the influence of nodes can be accurately calculated and the performance of node influence ranking can be further improved. Many traditional multiattribute sorting methods treat all attribute weights to be consistent. At the same time, there are many weighting methods, such as analytic hierarchy process, multiobjective programming, principal component analysis, and weighted least square method. Information entropy weighting method is an excellent weighting method, which has been verified by many examples. Our method adapts the method of information entropy weighting to avoid the defects of the traditional weighting method:

$$\mathrm{PN}(i) = w_1 * \mathrm{PN}_p(i) + w_2 * \mathrm{PN}_n(i), \tag{10}$$

where w_1 represents the weight of the location attribute and w_2 represents the weight of the neighbor attribute.

The calculation process of information entropy weighting method is as follows. First, the entropy value of each attribute is calculated:

$$H_i = -\frac{1}{\ln n} \sum_{j=1}^{n} r_{ij} \ln r_{ij}, \quad i = 1, 2, \tag{11}$$

where H_i represents the entropy of the ith attribute, and r_{ij} represents the normalized value of the ith attribute of the jth node. Because this method has only location and neighbor attributes, i is set as 1 and 2:

$$r_{1j} = \frac{\mathrm{PN}_p(j)}{\sum_{j=1}^{n} \mathrm{PN}_p(j)}, r_{2j} = \frac{\mathrm{PN}_n(j)}{\sum_{j=1}^{n} \mathrm{PN}_n(j)}. \tag{12}$$

Then, the information entropy is combined to calculate the weight of the two attributes:

$$w_i = \frac{1 - H_i}{2 - \sum_i H_i}, \quad i = 1, 2. \tag{13}$$

Whether the multiattribute-improved K-shell algorithm proposed in this section is feasible, the feasibility can be explained with the help of diagrams. The graph is an undirected graph with 16 nodes and 20 edges. The PN value of each node is calculated according to the algorithm. The values in the calculation process are shown in Table 1. As we can see from the table, the importance of all nodes can be sorted in the descending order according to the PN value. The PN values of node 11 and node 12 are the same and their importance cannot be distinguished. The PN values of nodes 1, 2, 3, and 4 are in the first gradient, and the PN value of node 2 is the largest. Its importance can be seen in Figure 1 in the network. PN values of nodes 5, 6, 7, and 8 are located in the second gradient. These nodes are not outer edge nodes.

The PN value of node 14 is larger than that of edge nodes 9, 11, 12, 13, and 16. It can be seen from the figure that since it is directly connected to core node 4, its influence has been enhanced. From the calculation results shown in Table 1, it can be preliminarily concluded that the improved K-shell method based on multiple attributes is feasible to some extent.

4. Experimental Setup

4.1. Data Description. We conduct several experiments on six different real networks to evaluate the performance of our proposed centrality measure. The real networks are drawn from disparate fields. CA-HepTh [25] is a collaboration network of Arxiv High Energy Physics Theory. Netscience [25] is the network of co-authorship of scientists in network theory and experiments. Cond-Mat [25] is from the e-print arXiv and covers scientific collaborations

Table 1: Example verification results of K-shell-improved algorithm based on multiple or inner core nodes.

Node	ks	Iter (i)	PNp (i)	PNn (i)	PN (i)
1	3	5	5.032744485944649	50	36.79944673550679
2	3	5	6.353940102911473	66	48.490348105799676
3	3	5	5.693342294428061	59	43.35133818263464
4	3	5	6.241636228400565	64	47.04449874200889
5	2	4	3.9754352186156736	34	25.186011640937103
6	2	4	3.8864732377707982	32	23.747014505664833
7	2	4	3.958485252509983	36	26.593917345765504
8	1	3	2.812214004336133	19	14.247925877904956
9	1	1	1.6372559148173789	7	5.425716932665104
10	1	2	2.7338437827623867	13	9.986275003487945
11	1	1	1.6033222618802176	5	4.002873871402858
12	1	1	1.6033222618802176	5	4.002873871402858
13	1	1	1.6605978084834119	9	6.8454506851464405
14	1	1	1.6775814953148829	16	11.795521737181323
15	1	2	2.208891742455916	13	9.832170482539274
16	1	1	1.6033222618802176	4	3.2964331094214456

between authors papers submitted to Condense Matter category. DNC Email [26] is the network of emails in the 2016 Democratic National Committee email leak. Nodes in the network correspond to persons in the dataset and the edge is the mail exchange between users. Ego-Twitter [27] contains Twitter user-user following information. A node represents a user and an edge indicates that the user follows the other user. Route Views [28] is the network of autonomous systems of the Internet connected with each other. Nodes are autonomous systems (AS), and edges denote communication. A brief overview of the networks is shown in Table 2.

Table 2: The basic topological features of four real network datasets.

Network	N	M	K	$K_{\max}$	C	r	β_{th}	β
CA-HepTh	9877	51971	5.26	65	0.47	0.268	0.087	0.12
Netscience	379	914	4.82	43	0.74	−0.082	0.142	0.25
Cond-Mat	16264	47595	5.85	107	0.62	0.185	0.084	0.14
DNC Email	2029	5598	4.72	404	8.9	−0.307	0.014	0.08
Ego-Twitter	23370	33101	2.83	239	2.15	−0.478	0.027	0.08
Route Views	6474	13895	4.3	1459	0.96	−0.182	0.007	0.06

N and M are the numbers of nodes and edges, respectively. K and $K_{\max}$ denote the average degree and the maximum degree. C and r are the clustering coefficient [29] and assortative coefficient [30]. β_{th} and β are the epidemic threshold of network and the infection probability used in our experiment.

4.2. Spreading Model. To evaluate the lists ranked by all the centrality measures, we need to know the list ranked by the real spreading process of the nodes. In the spreading process, the probability of accepting a message from another user depends on the user's influence [31]. So, the spreading efficiency of nodes is used to measure the ranking result of influential nodes. There are many information diffusion models, such as SIR model, Independent Cascade model, and Linear Threshold model, and some information diffusion models independent of network topology [32, 33]. In this paper, we employ the standard SIR model [34] to simulate the spreading process on networks and record the spreading efficiency for every node. In the SIR model, every node belongs to one of the susceptible states, the infected state or the recovered state. In detail, we set one node as an infected node and the other nodes are susceptible nodes. At each step, for every infected node, it can infect its susceptible neighbors with infection probability β and then can be removed with probability λ. Generally, we set $\lambda = 1.0$. The appropriate propagation probabilities are needed to be chosen, in case too small or too large propagation probability makes the propagation effect not ideal and leads to the failure to distinguish the influence of nodes. According to the heterogenous mean-field method, the epidemic threshold of network is $\beta_{th} = <k> / (<k^2> - <k>)$. k and k^2 are degree and second-order degree of node. The propagation probabilities are set just larger than the epidemic threshold. In the experiment, this dynamical process of infection and recovering will repeat until there are no infected nodes. The sum of infected and recovered nodes at time t, denoted by $F(t)$, can be considered as an indicator to evaluate the influence of the initially infected node at time t. Obviously, $F(t)$ increases with the increasing of t and will reach stable state at time t_c, denoted by $F(t_c)$, where t_c represents the final time and $F(t_c)$ represents the eventual influence of the initially infected node. To guarantee the reliability of the results, all of them are averaged over a large number of realizations.

4.3. Evaluation. In order to evaluate the performance of the centrality measures, we use Kendall's coefficient τ [35] to measure the correlation between one topology-based ranking list and the one generated by the SIR model, which is approached by a large number of simulations. Let (x_i, y_i) and (x_j, y_j) be a randomly selected pair of joint from two ranking list, X and Y, respectively. If both $(x_i > x_j)$ and

$(y_i > y_j)$ or if both $(x_i < x_j)$ and $(y_i < y_j)$, they are said to be concordant. If $(x_i > x_j)$ and $(y_i < y_j)$ or $(x_i < x_j)$ and $(y_i > y_j)$, they are said to be discordant. If $(x_i = x_j)$ or $(y_i = y_j)$, the pair is neither concordant nor discordant. Kendall's coefficient τ is defined as

$$\tau = \frac{n_c - n_d}{0.5n(n-1)}, \tag{14}$$

where n_c and n_d denote the number of concordant and discordant pairs, respectively. The value τ lies between +1 and −1. The higher the τ value indicates, the more accurate ranked list a centrality measure could generate. The most ideal case is $\tau = 1$, where the ranked list generated by the centrality measure is exactly the same as the ranked list generated by the real spreading process.

To measure the imprecision [17] of methods in ranking influential nodes, we compare propagation capability of influential nodes obtained by ranking result with nodes which have largest propagation capability and the propagation capability of nodes generated from the SIR model. Kendall's correlation coefficient considers the correlation between the ranking order of all nodes in the network and the order of propagation capability of all nodes. However, the imprecision function is used to evaluate the cumulative propagation capability of top-ranked nodes in different proportions. The imprecision function $\varepsilon(p)$ is defined as

$$\varepsilon(p) = 1 - \frac{\sum_{i \in \phi_m(p)} F_i(t_c)}{\sum_{j \in \phi_s(p)} F_j(t_c)}, \tag{15}$$

where p is the proportion of top-ranked nodes and $\phi_m(p)$ and $\phi_s(p)$ denote the set of nodes at the top when proportion is p. $F_i(t_c)$ is propagation capability of node i. The value of ε lies between 0 and 1. The lower ε value indicates, the more precise the centrality measure is in ranking propagation capability of node.

5. Results and Discussion

In this chapter, SIR model, Kendall coefficient, and imprecision function are used to verify the performance of the proposed method. Degree, Ks (K-shell), MDD (Mixed Degree Decomposition), ksIF (K-shell iteration factor method), and Cnc+ (Extended Neighborhood coreness centrality measure) were compared with the proposed multiattribute PN method.

5.1. Evaluate the Spreading Capability of Nodes. This section verifies that different propagation probabilities are selected under the fixed proportion of transmission sources to calculate the propagation capability of the influence node set. By comparing the propagation capability of different algorithms under the same propagation probability, the performance of different methods can be compared. The simulation set the propagation source ratio as 0.1. The max time step of infection process is set as 100, but the infection will stop when there is no user in the infection state. The results were based on an average of 500 independent experiments. The simulation results are shown in Figure 3. The abscissa is the propagation probability and the ordinate is the propagation capability, which is expressed as a percentage.

Firstly, with the increasement of the propagation probability, the node propagation ability of the six methods is improved. The performance of the six methods is relatively close when the propagation probability is small and obviously different when the propagation probability is large. Specifically, in the CA-HepTh, DNC Email, and Cond-Mat, the performance of PN maintained the best under various propagation probabilities. In the Ego-Twitter dataset, the performance of PN and Cnc+ is significantly higher than that of the other four methods. Moreover, when the transmission probability is 0.03, the node transmission ability of Cnc+ is higher, and in other cases, the node transmission ability of PN method is higher. In the Netscience and Route Views dataset, the performance of PN is relatively better than the other method, but Cnc + outperforms the PN method when the transmission probability is less than 0.1 in the Netscience. PN is not the best in several points in Route Views, but it had highest propagation ability in most cases, as shown in Figure 3(f). In general, the proposed PN in this chapter has better performance. The difference in performance is related to the specific network structure.

5.2. Evaluate the Imprecision of Ranking. This section verifies that different proportions of top-ranked nodes are selected under the fixed propagation probability to calculate the imprecision of ranking the influence of nodes. By comparing the imprecision of different algorithms under the same proportion of top-ranked, the performance of different methods in ranking the influence of nodes can be distinguished. The simulation set the propagation probability of node, as shown in Table 2, and the proportion top-ranked nodes varies from 0.01 to 0.2. The max time step of infection process set as 100 and the results were based on an average of 500 independent experiments. Simulation results in six networks are shown in Figure 4. The abscissa is the proportion of top-ranked nodes that are considered and the ordinate is the imprecision of ranking.

In Figure 4(a), the method PN, KsIF, and Cnc + have low imprecision when p is higher than 0.02 and they have poor performance in identifying the 2% top-ranked nodes. However, the imprecision of PN is much lower than the other two in CA-HepTh dataset. The imprecision of Degree to rank top 1% nodes is low about 0.07 than the method PN. However, Degree has highest ε when p is greater than 0.04. In Netscience and Cond-Mat, Degree, K-shell, and MDD have higher value, and PN is lowest in most p except that p is in the range of 0.02 to 0.04, KsIF outperforms PN in Netscience. When p is larger than 0.04 and smaller than 0.06, Cnc + outperforms PN in Cond-Mat. In the dataset, DNC Email, K-shell, and PN are better compared with other method, and K-shell has a slight edge when p between 0.04 and 0.1 or 0.13 and 0.15. However, K-shell performs poor in 3% top-ranked nodes. In the dataset, Ego-Twitter and Route Views, it is obvious that the method we proposed performs

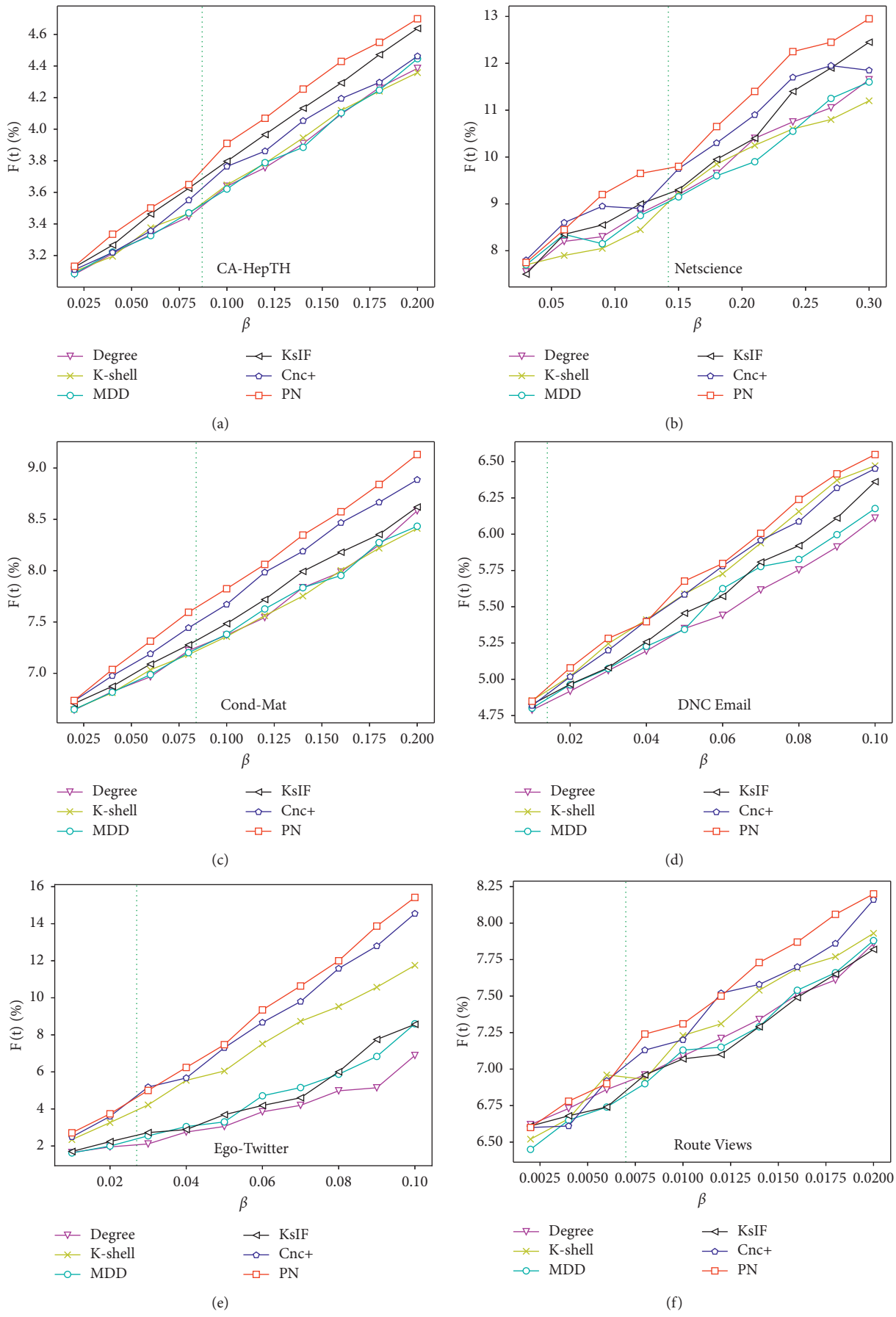

Figure 3: The propagation capability graph of the five methods under different propagation probabilities. The experiments are simulated on six different datasets: CA-HepTh (a), Netscience (b), Cond-Mat (c), DNC Email (d), Ego-Twitter (e), and Route Views (f). The green vertical line is the epidemic threshold of the corresponding network.

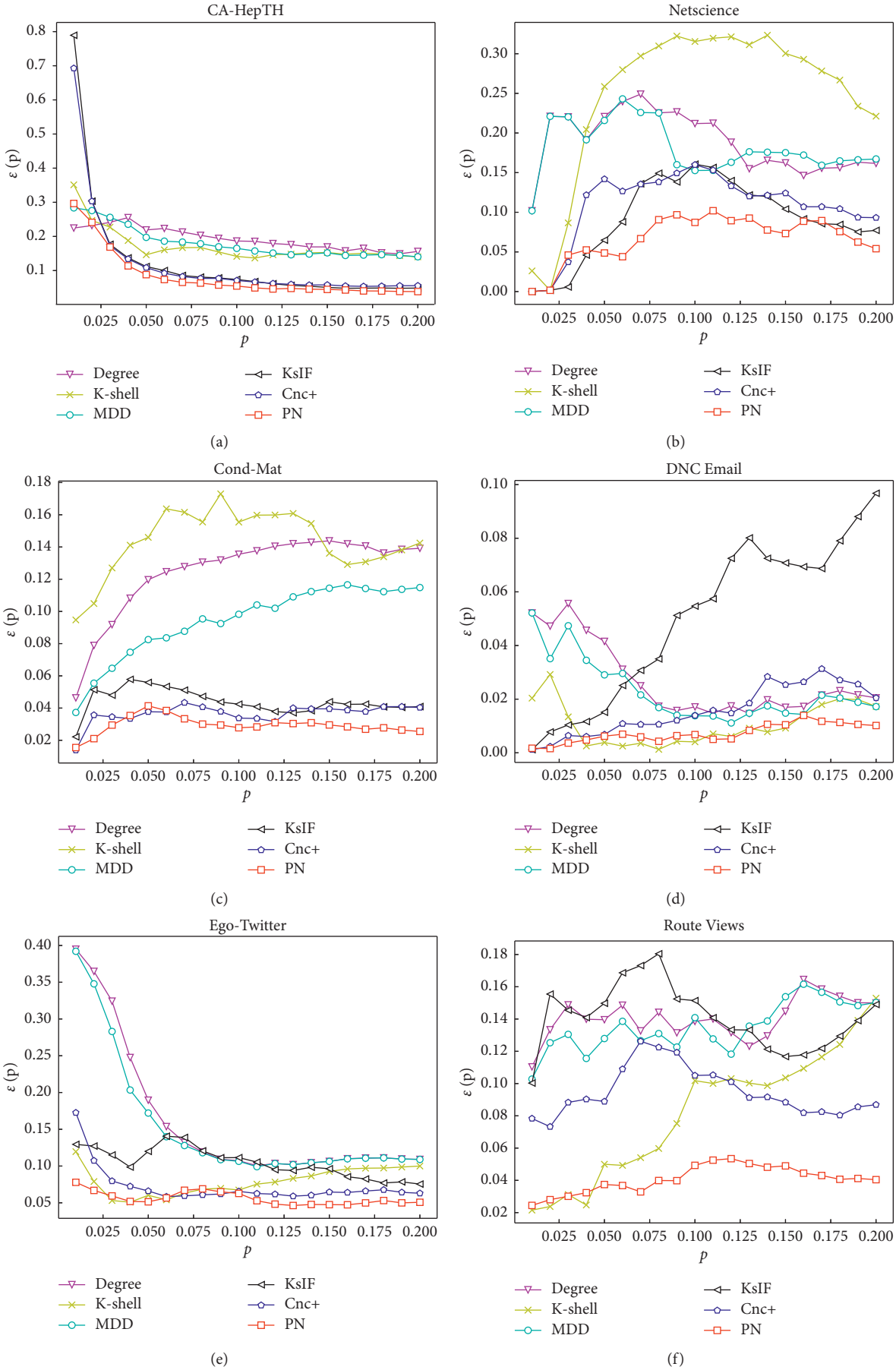

Figure 4: The imprecision of six ranking methods with different proportion of top-ranked nodes. The experiments are simulated on six different datasets: CA-HepTh (a), Netscience (b), Cond-Mat (c), DNC Email (d), Ego-Twitter (e), and Route Views (f).

well and is more stable in all cases. From the simulation results of the abovementioned six datasets, it can be seen that the method PN can not only precisely identify most of the top 1% to 4% nodes but also rank the following nodes by their influence stably.

5.3. Evaluate the Correlation Coefficients of Method. This section verifies the correlation between the influence value and propagation ability of nodes under different propagation probabilities by Kendall's coefficient τ. The value range of propagation probability in this simulation is 0.02 to 0.2 in the datasets CA-HepTh, Cond-Mat, DNC Email, Ego-Twitter, and Route Views, and extended to 0.3 in the dataset Netscience because of higher epidemic threshold of it. In the DNC Email and Route Views, we also put the simulation result when the value of propagation probability is equal to their epidemic threshold. The max time step of infection process is 100. The results were based on an average of 500 independent experiments. The simulation results are shown in Figure 5. The abscissa is the propagation probability and the ordinate is the correlation coefficient. The epidemic threshold of each network is also drawn into the figure as a vertical line.

The correlation of ranking result of different methods and real propagation abilities obtained by the SIR model has different manifestations in different datasets, but the general trend is the same. When the propagation probability is small, the ranking correlation of Degree, K-shell, and MDD with real propagation ability is higher than the other three methods and the thresholds are roughly the same as the epidemic threshold β_{th} of network. It can be seen in Figure 5 that the Degree has high Kendall's coefficient τ when β is less than 0.06 in the CA-HepTh and Cond-Mat and less than 0.07 in the Netscience, and it is not obvious in other three networks because of the small epidemic threshold. However, in the cases where β is higher than β_{th}, these methods that have considered both degree and other aspects such as the influence of neighbors to the spreading ability of nodes perform better than the method that mainly take degree into account. This is because the very small propagation probability will make the infection behaviour in a small range near the initial infected node, unable to spread over a large area, so the degree is the decisive factor. As the probability of propagation increases, the act of infection becomes easier, so the extent of infection depends not only on the number of neighbors of a node but also on the ability of its neighbors to propagate, or even the neighbors' neighbors. On the contrary, when the propagation probability increases to a point much larger than β_{th} of network, infection becomes too easy so that the nodes in the core of the network or with high degree had largest scope of infection. It is clearly shown in the networks with small epidemic threshold such as DNC Email and Route Views in the figure. When $\beta > \beta_{th}$, the method PN has the highest Kendall's coefficient t in CA-HepTh, Netscience, and Cond-Mat. Our proposed method also performs best when β is between 0.06 to 0.14 in DNC Email and between 0.06 to 0.18 in Route Views. In short, compared with the other five methods, the PN method proposed in this chapter has good performance under the appropriate propagation probability value range.

5.4. Evaluate the Performance of Selecting Seeds in Influence Maximization Problem. This section verifies the reliability of PN when it is applied in the influence maximization problem. Maximization of influence is widely used in real life. For example, viral marketing is a typical application that can promote new products or ideas for merchants or publicity departments. It aims to select a group of nodes in the network called seed node set as initial propagation nodes and spread in the network as widely as possible according to a certain diffusion mode. There are many related studies. We use the selection of top from the ranking to specify seed nodes to measure the propagation range of seed nodes selected by several ranking methods. We examine the spreading efficiency of different seed node set with six real networks: CA-HepTh, Netscience, Cond-Mat, DNC Email, Ego-Twitter, and Route Views.

In the experiment, we set P as the proportion of seed nodes in the whole network, ranging from 0.01 to 0.05. We also used the SIR Model to simulate the propagation process and calculated the influence range by using the number of users who were finally infected. The propagation probability in the simulation is determined according to the epidemic threshold and is shown in Table 2. The max time step of infection process is set as 100 and the results were based on an average of 500 independent experiments. The simulation results are shown in Figure 6. The abscissa is the ratio of the seed set in all users, and the ordinate is the propagation scope of initial seed nodes and expressed as a percentage.

The results manifest that the PN method performs best in the datasets Cond-Mat, DNC Email, Ego-Twitter, and Route Views. In the CA-HepTh dataset, the seed set selected by the method PN when P is greater than 0.02 can infect wilder than others. When P is 0.02 and below, degree is most effective and the PN is better than KsIF an Cnc+. In the Netscience network, KsIF is best when P is 0.03 and 0.04, and the infected rate of seeds selected by PN is almost same as KsIF in other cases.

5.5. Measure the Ranking Uniqueness and Distribution. This section verifies the monotonicity of the new method on the sorting by using Bae and Kim's ranking monotonicity method [23]. Since the K-shell method will calculate the K-shell value of many nodes into the same value, it is difficult to distinguish their differences in influence. In this respect, the method we proposed can do better. According to the definition of Bae and Kim, the monotonicity of the ranking result is expressed as follows:

$$M(R) = \left[1 - \frac{\sum_{r \in R} n_r (n_r - 1)}{n(n-1)}\right]^2. \tag{16}$$

In (16), R represents the ranking list and n is the size of it, every element of the ranking list is a set of nodes that have the same ranking value, and n_r is the number of nodes in R that have the same ranking position r. The

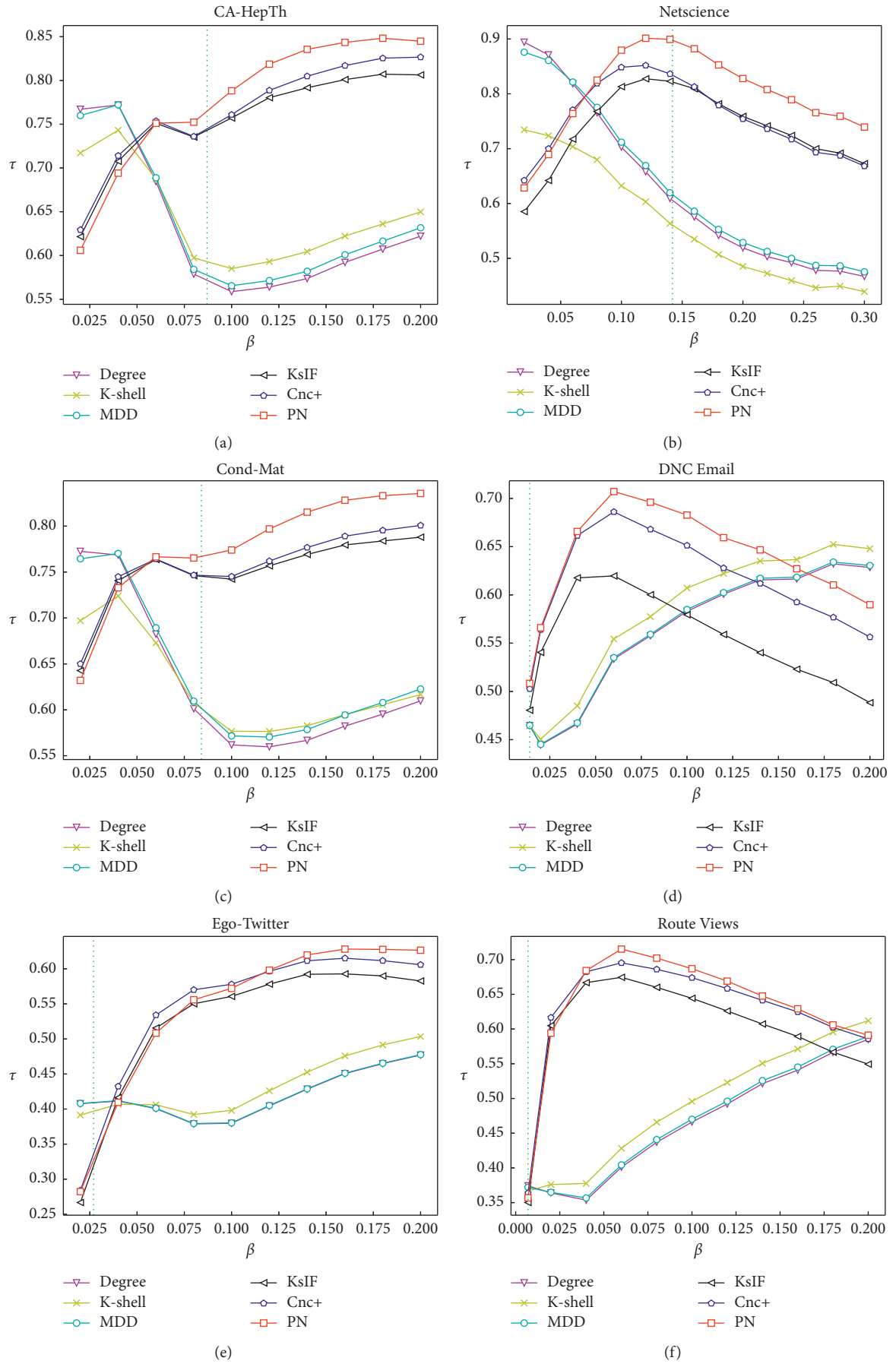

Figure 5: The correlation coefficient variation diagram of the six methods under different propagation probabilities. The experiments are simulated on six different datasets: CA-HepTh (a), Netscience (b), Cond-Mat (c), DNC Email (d), Ego-Twitter (e), and Route Views (f). The green vertical line is the epidemic threshold of the corresponding network.

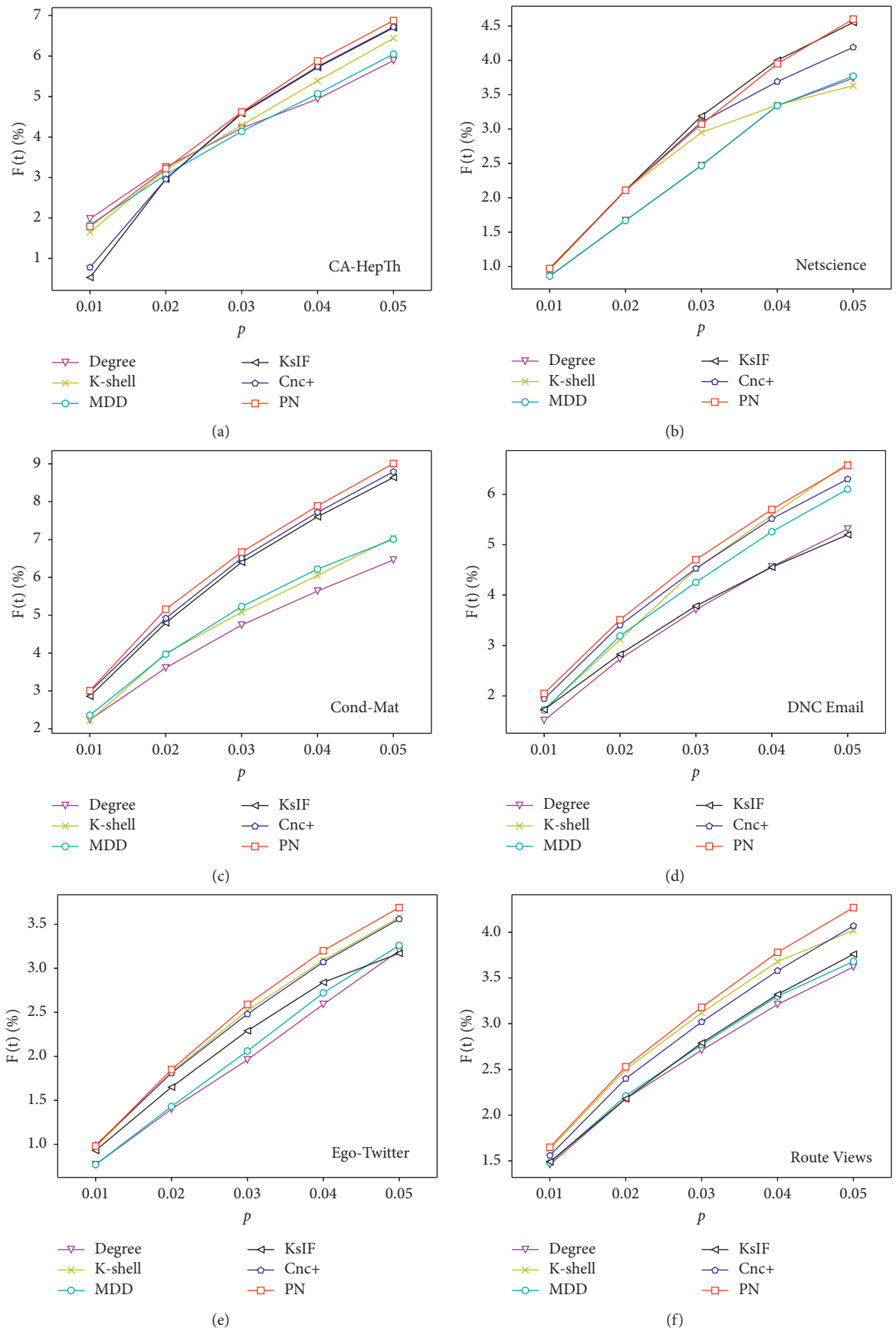

Figure 6: The performance of selecting seeds of the six methods in influence maximization problem. The experiments are simulated on six different datasets: CA-HepTh (a), Netscience (b), Cond-Mat (c), DNC Email (d), Ego-Twitter (e), and Route Views (f).

Table 3: The monotonicity M (R) of six real networks, R is representing the ranking vector of different methods.

Network	M (Degree)	M (Ks)	M (M DD)	M (KsIF)	M (Cnc+)	M (PN)
CA-HepTh	0.762683	0.674237	0.81057	0.992416	0.988876	0.993566
Netscience	0.764206	0.642083	0.821527	0.994702	0.989307	0.995036
Cond-Mat	0.809054	0.740879	0.86048	0.994138	0.991407	0.994788
DNC Email	0.339762	0.32344	0.344565	0.936082	0.936028	0.936465
Ego-Twitter	0.131463	0.122307	0.131656	0.995574	0.991322	0.996321
Route Views	0.587174	0.549622	0.608671	0.993285	0.992548	0.993893

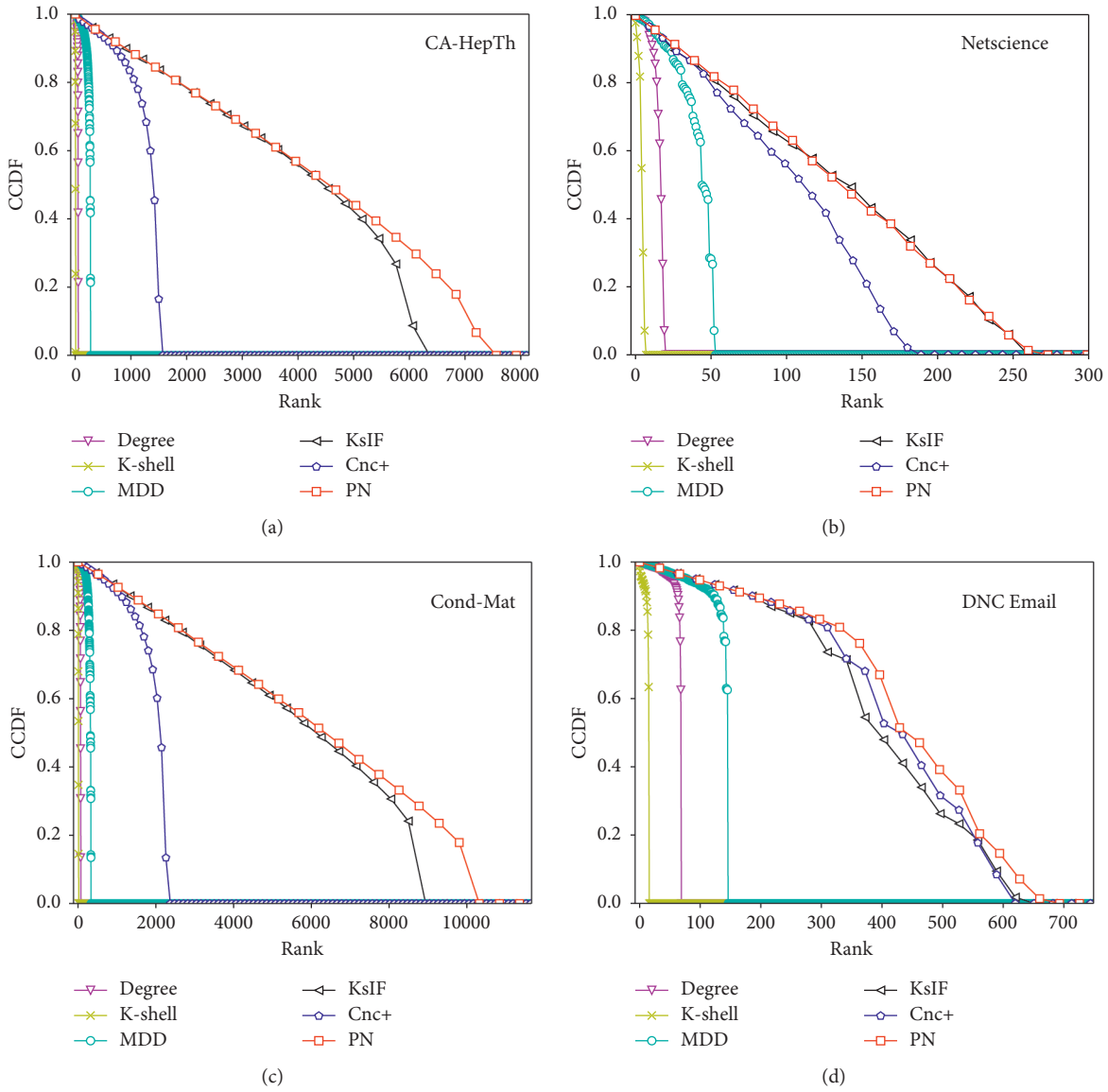

Figure 7: Continued.

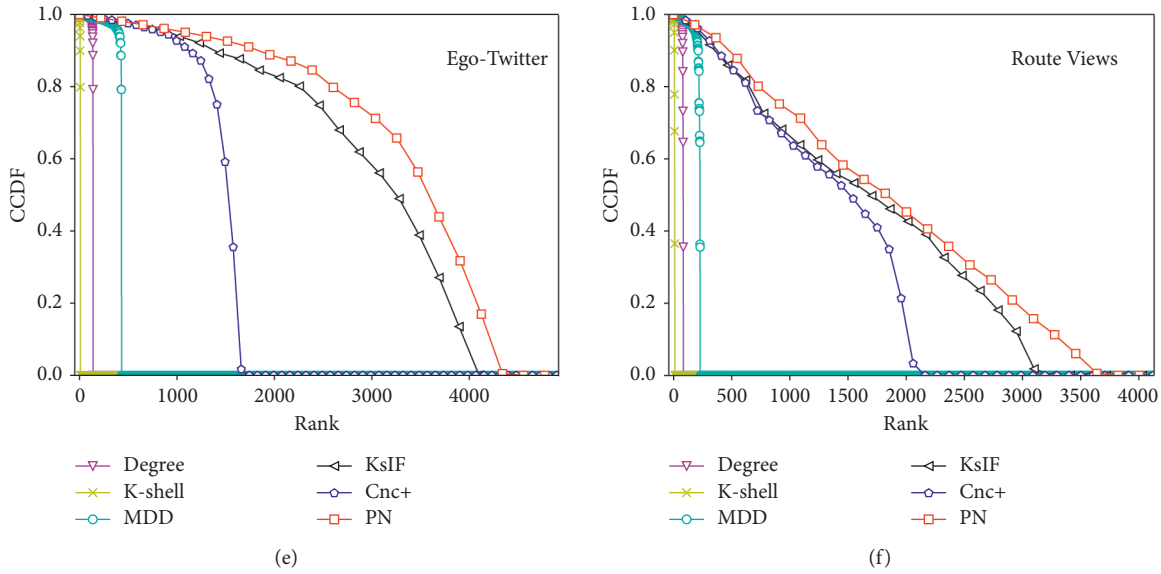

Figure 7: The complementary cumulative distribution (CCDF) of the six methods in six different datasets: CA-HepTh (a), Netscience (b), Cond-Mat (c), DNC email (d), Ego-Twitter (e), and Route Views (f).

value of $M(R)$ fluctuates between $[0, 1]$ and the higher the value, the stronger the uniqueness. In extreme cases, 1 means that each node is assigned a different sort value, whereas 0 is the opposite and all nodes are in the same rank.

We examine the monotonicity of different methods with the same six datasets as above. The calculation results are shown in Table 3. It can be seen from the table that the monotonicity result of the PN is apparently higher than the degree, *ks* and MDD, and approximate to KsIF and Cnc+.

In order to clarify the ranking distribution of the different measures more clearly, a complementary cumulative distribution function (CCDF) is plotted. According to the CCDF principle, if many nodes are in the same rank, the CCDF plot will decrease rapidly; otherwise, the CCDF plot will slow down. Figure 7 shows the ranking distribution in six networks.

The line representing Degree, *ks*, or MDD drop sharply, as can be seen on the left side of each graph. This is especially true when the number of nodes in the dataset is large. For the method PN, the ranking distribution is slightly improved compared with Cnc+ and KsIF in datasets DNC Email and Route Views. The curves of KsIF and PN in the dataset Netscience are basically overlapping and drop off more slowly than that of Cnc+. In the dataset CA-HepTh and Cond-Mat, the method Cnc+ also does not perform well compared to the method KsIF and PN. The KsIF and PN are equally good at identifying the influential nodes, while in the latter part of the ranking, the downward trend of KsIF curve is more obvious than that of PN curve. It is indicated that the ability of method PN to distinguish the nodes' spreading capability is better than Cnc+. It can be seen in the Ego-Twitter that the performance of PN is better than the method Cnc+. So, we can say that PN performs well in most networks.

6. Conclusions

On the basis of the K-shell method, we proposed a new multiattribute ranking method based on node position and neighborhood. We made full use of the iterative information in the decomposition process. First, the iteration information is processed by sigmod function to obtain the position index. The position attribute is obtained by combining the shell value and the position index. Then, the local information of the node is adopted to obtain the neighbor property. Furthermore, the position attribute and neighbor attribute are weighted by the method of information entropy weighting. Finally, we evaluated the propagation capability of different propagation probabilities, the imprecision of different proportions, the correlation coefficient of different propagation probabilities, and the propagation capability of selected seed nodes in influence maximization problem. At the same time, we also verified the good performance of our method in distinguishing influence of nodes. Compared with other K-shell decomposition and its improved algorithms, the method proposed in this paper had better performance. Through simulation experiments, it is found that the PN method can make full use of the iterative information in the decomposition process and the influence of neighbors to further distinguish the difference of nodes with the same *ks* value. Experiments with SIR model, Kendall's coefficient, and imprecision function fully verified the correctness and effectiveness of the proposed method. In a word, the effectiveness of the proposed method in the identification of influence nodes was verified by various forms of experiments.

Data Availability

Previously reported network datasets were used to support this study and are available at http://networkrepository.com

and http://konect.uni-koblenz.de. These datasets are cited at relevant places within the text as references [22–25].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Key Industry Projects in Shaanxi Province (Grant no. 2019ZDLGY09-03), Natural Science Foundation of Shaanxi Province (Grant nos. 2018JM6053 and 2018JZ6006), National Natural Science Foundation of China (Grant nos. 61372076, 61301171, and 61771296), and 111 Project (Grant no. B08038).

References

[1] L. Gao, W. Wang, P. Shu, H. Gao, and L. A. Braunstein, "Promoting information spreading by using contact memory," *EPL (Europhysics Letters)*, vol. 118, no. 1, p. 18001, 2017.

[2] V. P. T. Menta and P. K. Singh, "Efficient selection of influential nodes for viral marketing in social networks," in *Proceedings of the IEEE International Conference on International Conference on Current Trends in Advanced Computing*, Bangalore, India, March 2017.

[3] Y. Cho, J. Hwang, and D. Lee, "Identification of effective opinion leaders in the diffusion of technological innovation: a social network approach," *Technological Forecasting and Social Change*, vol. 79, no. 1, pp. 97–106, 2012.

[4] Z. Li, F. Xiong, X. Wang, H. Chen, and X. Xiong, "Topological influence-aware recommendation on social networks," *Complexity*, vol. 2019, Article ID 6325654, 12 pages, 2019.

[5] F. Xiong, X. Wang, S. Pan, H. Yang, H. Wang, and C. Zhang, "Social recommendation with evolutionary opinion dynamics," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–13, 2018.

[6] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[7] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.

[8] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.

[9] L. C. Freeman, "Centrality in social networks' conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1979.

[10] B. Tian, J. Hu, and Y. Deng, "Identifying influential nodes in complex networks based on AHP," *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 4, pp. 1777–1787, 2017.

[11] Q. Ma and J. Ma, "Identifying and ranking influential spreaders in complex networks with consideration of spreading probability," *Physica A: Statistical Mechanics and Its Applications*, vol. 465, pp. 312–330, 2017.

[12] Z. Lv, N. Zhao, F. Xiong, and N. Chen, "A novel measure of identifying influential nodes in complex networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 523, pp. 488–497, 2019.

[13] L. Lü, T. Zhou, Q. Zhang et al., "The H-index of a network node and its relation to degree and coreness," *Nature Communications*, vol. 7, no. 1, p. 10168, 2016.

[14] A. Sheikhahmadi and M. A. Nematbakhsh, "Identification of multi-spreader users in social networks for viral marketing," *Journal of Information Science*, vol. 43, no. 3, pp. 412–423, 2017.

[15] X. Wang, X. Zhang, C. Zhao et al., "Maximizing the spread of influence via generalized degree discount," *Plos One*, vol. 11, no. 10, Article ID e0164393, 2016.

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

[17] 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.

[18] L. Lu, Y. C. Zhang, C. H. Yeung et al., "Leaders in social networks, the delicious case," *PLoS One*, vol. 6, no. 6, Article ID e21202, 2011.

[19] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[20] 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.

[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] 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.

[23] 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, no. 4, pp. 549–559, 2014.

[24] Z. Wang, Y. Zhao, J. Xi, and C. Du, "Fast ranking influential nodes in complex networks using a k-shell iteration factor," *Physica A: Statistical Mechanics and Its Applications*, vol. 461, pp. 171–181, 2016.

[25] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, Austin, TX, USA, January 2015.

[26] *DNC Emails Network Dataset*, KONECT, Landau Germany, 2017.

[27] J. Leskovec and J. McAuley, "Learning to discover social circles in ego networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, pp. 548–556, Lake Tahoe, NE, USA, December 2012.

[28] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph Evolution: densification and shrinking diameters," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, 2007.

[29] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

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

[31] F. Xiong, W. Shen, H. Chen, S. Pan, X. Wang, and Z. Yan, "Exploiting implicit influence from information propagation for social recommendation," *IEEE Transactions on Cybernetics*, pp. 1–14, 2019.

[32] S. Fang, N. Zhao, N. Chen, F. Xiong, and Y. Yi, "Analyzing and predicting network public opinion evolution based on group persuasion force of populism," *Physica A: Statistical Mechanics and Its Applications*, vol. 525, pp. 809–824, 2019.

[33] N. Zhao, S. Fang, N. Chen, and C. Pei, "Modeling and analyzing the influence of multi-information coexistence on attention," *IEEE Access*, vol. 7, pp. 117152–117164, 2019.

[34] L. Guan, D. Li, K. Wang, and K. Zhao, "On a class of nonlocal SIR models," *Journal of Mathematical Biology*, vol. 78, no. 6, pp. 1581–1604, 2019.
[35] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1-2, pp. 81–93, 1938.