Identifying Influential Spreaders in Complex Networks by Considering the Impact of the Number of Shortest Paths*

LUAN Yangyang · BAO Zhongkui · ZHANG Haifeng

DOI: 10.1007/s11424-021-0111-7
Received: 20 May 2020 / Revised: 19 July 2020

Abstract The study on how to identify influential spreaders in complex networks is becoming increasingly significant. Previous studies demonstrate that considering the shortest path length can improve the accuracy of identification, but which ignore the influence of the number of shortest paths. In many cases, even though the shortest path length of two nodes is rather larger, their interaction influence is also significant if the number of shortest paths between them is considerable. Inspired by this fact, the authors propose an improved centrality index (ICC) based on well-known closeness centrality and a semi-local iterative algorithm (semi-IA) to study the impact of the number of shortest paths on the identification of the influential spreaders. By comparing with several traditional centrality indices, such as degree centrality, k-shell decomposition, betweenness centrality and eigenvector centrality, the experimental results on real networks indicate that the ICC index and semi-IA have a better performance, regardless of the identification capability or the resolution.

Keywords Complex networks, influential spreader, number of shortest paths.

1 Introduction

Many real systems such as social, biological, technological systems can be described in terms of complex networks, therefore, the research on complex networks has gained attention in many fields. In particular, how to effectively identify influential spreaders is a central topic of interest in complex networks[1–9], because it is of great significance for developing efficient strategies to control epidemic spreading, accelerate information diffusion, promote new products, and so on. Thus, a lot of centrality indices have been proposed to identify influential spreaders in networks, including degree centrality (DC)[10], k-shell decomposition (KS)[11], betweenness

LUAN Yangyang · BAO Zhongkui (Corresponding author) · ZHANG Haifeng
School of Mathematical Science, Anhui University, Hefei 230601*, China.*
Email: yyluan@stu.ahu.edu.cn; zkbao@ahu.edu.cn; haifengzhang1978@gmail.com.
*This research was supported by the National Natural Science Foundation of China under Grant Nos. 61973001, 61806001, the Natural Science Foundation of Anhui Province under Grant No. 1808085MF201, the State Key Laboratory for Ocean Big Data Mining and Application of Zhejiang Province under Grant No. OBDMA201502, and Anhui University Foundation under Grant No. 01005102.
◇ *This paper was recommended for publication by Editor LIU Guoping.*

centrality (BC)[12], eigenvector centrality (EC)[10, 13], etc. Degree centrality is a simple and intuitive index only by considering the number of nearest neighbors which ignores the position and higher-order neighbors' information. K-shell method shows that the nodes with the same k-core values have the equal influence. However, this method will be invalid in tree-like networks, regular networks, as well as BA networks, and has a low resolution because only the impact of the residual degree is considered[14]. Betweenness centrality is a global measure and is not suitable for large-scale networks owing to its higher computation complexity. Eigenvector centrality is an appropriate measure where the importance of a node is assumed to depend on the number and importance of neighbor nodes.

The closeness centrality (CC) assumes that a node is more influential if the average shortest path length from it to other nodes is smaller[15], namely, the impact of shortest path length is considered in CC index. Later, some methods were proposed to identify the influential spreaders in complex networks by considering the impact of the shortest path length from different perspectives[16, 17]. For instance, Ma, et al. designed a gravity centrality index by taking into account the k-shell value and the shortest path length[16]. Liu, et al. proposed a generalized closeness centrality to identify influential spreaders[17]. Bao, et al. also proposed an SP (spreading probability) index by considering the effects of the shortest path length as well as spreading capability[18]. The results in these works indicate that considering the impact of shortest path length can improve the identification capability of influential spreaders, however, most of them ignore the impact of the number of shortest paths (NSPs). Supposing that node j and node k are the first-order and the second-order neighbors of node i, if the NSPs between node i and k is rather large, the impact of node k on node i is considerable and may be larger than that of node j. Inspired by this fact, we think that the impact of the NSPs on the identification of influential spreaders cannot be ignored. In doing so, we proposed two indices to address the impact of the NSPs, the first one is the improved closeness centrality (ICC) index, in which the CC index is improved by incorporating the effect of the NSPs. The other one is the semi-local iterative algorithm (semi-IA), where the importance of a node is affected by the importance of other nodes weighted by the shortest path length between them and the NSPs. In semi-IA, we just consider the shortest path length within 3-hops when the scale of the network is less than 1000 and within 5-hops when the scale of the network is larger than 1000, so it is called a semi-local algorithm. We employ the SIR (susceptible-infectious-recovered) epidemic model to estimate the effectiveness of proposed methods, the experimental results indicate that our methods can better identify the influential nodes than that of CC, DC, KS, BC, EC, and so on.

The layout of the paper is organized as follows. In Section 2, we first briefly review several typical centrality indices which are used to compare in this work, and the descriptions of our methods are presented. Then the experimental results are presented in Section 3. Finally, conclusions are summarized in Section 4.

2 Method

An undirected network is represented by $G = (V, E)$ with V nodes and E edges, and its structure can be described by an adjacency matrix $A = (a_{ij})_{N\times N}$, where N is the number of nodes, $a_{ij} = 1$ if node i is connected to node j, and $a_{ij} = 0$ otherwise.

Here we briefly review the definitions of several centrality indices that will be discussed in this work.

2.1 Existing Methods

The degree centrality (DC) of node i is defined as the number of nearest neighbors (i.e., the first-order neighbors), namely

$$\mathrm{DC}(i) = \sum_{j=1}^{N} a_{ij}. \tag{1}$$

The betweenness centrality (BC) of node i is defined as the fraction of all shortest paths travelling through the node, which is denoted as

$$\mathrm{BC}(i) = \sum_{s\neq i\neq l} n_{sl}^{i}/n_{sl} \tag{2}$$

with n_{sl} and n_{sl}^{i} be the NSPs between nodes s and l, and NSPs between s and l passing through node i, respectively.

The closeness centrality (CC) of node i is defined as the reciprocal of the sum of the shortest path length to all other nodes:

$$\mathrm{CC}(i) = \frac{N-1}{\sum_{j\neq i} d_{ij}}, \tag{3}$$

where d_{ij} represents the distance from node i to node j. Closeness can be considered as a measure of how long it will spread information from a given node to other reachable nodes in the network.

The eigenvector centrality (EC) holds that the importance of a node not only depends on the number of its neighbors but also the importance of them. The eigenvector centrality of node i is denoted as

$$\mathrm{EC}(i) = x_i = c\sum_{j=1}^{N} a_{ij}x_j, \tag{4}$$

where x_i represents the importance value of node i, c is a proportional constant.

The HITs algorithm gives each node two metrics: Authorities and hubs[19]. We define a_i^t and h_i^t as the authorities and hubs of node i at time t, therefore, in each iteration of time step:

$$a_i^t = \sum_{j=1}^{N} a_{ji}h_j'^{t-1}, \quad h_i^t = \sum_{j=1}^{N} a_{ij}a_j'^{t}, \tag{5}$$

and normalization is needed after each time step:

$$a_i'^t = \frac{a_i^t}{||a^t||}, \quad h_i'^t = \frac{h_i^t}{||h^t||}, \tag{6}$$

when the iteration reaches convergence, we use the current a_i and h_i as the authorities and hubs of node i. In this paper, we use the authorities as the importance of nodes.

The k-shell (KS) decomposition method is implemented by the following steps: Firstly, all the nodes with degree $k = 1$ are removed until all nodes' degrees are larger than one. All of these removed nodes are 1-shell. Then the nodes with $k = 2$ is recursively removed, and the existing nodes are deleted unceasingly until all nodes' degrees are larger than two, and these removed nodes are 2-shell. This procedure continues until all nodes have been assigned to a k-shell[11, 20].

2.2 Proposed Methods

In fact, nodes with the same shortest path length from other nodes in the network may have different importance. The node importance is not only related to the shortest path length between nodes, but also to the NSPs, and the influence increases with the NSPs. In order to better identify the influential nodes, we define an improved centrality index based on the CC, which further considers the contribution of NSPs to the centrality. Our improved closeness centrality (labeled as ICC) is defined as

$$\mathrm{ICC}(i) = \frac{N-1}{\sum_{j\neq i} d_{ij}\left(\frac{1}{n_{ij}}\right)^{\alpha}}, \tag{7}$$

where n_{ij} is the NSPs between nodes i and j, $0 \leq \alpha \leq 1$ is a free parameter and small value of α means less influence of the NSPs is considered. The ICC index degenerates to CC index when $\alpha = 0$.

Enlightened by the EC and HITs algorithm, we also consider an iterative algorithm (labeled as semi-IA) by considering both the shortest path length and the NSPs to measure the different impacts of nodes on the given nodes, which is defined as

$$X_i\,(t+1) = \sum_{j=1, j\neq i, j\in\varphi_i}^{N} \frac{n_{ij}^{\gamma}}{d_{ij}} X_i'\,(t)\,, \quad X_i\,(0) = 1, \quad i = 1, 2, \cdots, N, \tag{8}$$

where φ_i is the neighborhood set whose distance to node i is less than or equal to a given value r, $X_i\,(t)$ is the influence of node i at time t, $0 \leq \gamma \leq 1$ is the weighting factor of NSPs, in addition, normalization is needed after each time step:

$$X_i'^t = \frac{X_i\,(t)}{||X\,(t)\,||}, \quad i = 1, 2, \cdots, N, \tag{9}$$

when the iteration process of Equation (8) and Equation (9) reaches steady state, we use these scores as the influence of nodes in network. To reduce the calculation and time complexity, in this paper, we set $r = 3$ when $N < 1000$, otherwise, $r = 5$.

3 Experimental Results

3.1 Data Description

In this section, we compare the effectiveness of several typical indices with our ICC index and semi-IA from different aspects on different real networks. In this paper, the 10 real networks are as follows: Word (adjacency relation in English text)[21], Celegans (network of the nematode worm C.elegans)[22], USAir (the network of the US air transportation system)[23], Facebook (Slavo Zitnik's friendship network in Facebook)[24], Metabolic (metabolic network of C.elegans)[25], Email (e-mail network of University at Rovira i Virgili, URV)[23], TAP (yeast protein-protein binding network generated by tandem affinity purification experiments)[14], Yeast (network or protein interaction)[26], Router (the router-level topology of the Internet)[27], HEP (collaboration network of high-energy physicists)[28]. For simplicity, these networks are treated as undirected and unweighed networks in this work. Structural information about these 10 real networks are summarized in Table 1.

Table 1 Basic structural properties of network

Network	N	M	β_c	H	$\widetilde{r}$	C	L	D
Word	112	425	0.078	1.815	−0.129	0.173	2.536	5
Celegans	297	2148	0.040	1.801	−0.163	0.292	2.455	5
USAir	332	2126	0.023	3.464	−0.208	0.749	2.738	6
Facebook	324	2218	0.049	1.567	0.247	0.465	3.054	7
Metabolic	453	2025	0.026	4.485	−0.226	0.647	2.664	7
Email	1133	5451	0.057	1.942	0.078	0.220	3.606	8
TAP	1373	6833	0.065	1.644	0.579	0.529	5.224	12
Yeast	2375	11693	0.030	3.476	0.454	0.306	5.096	15
Router	5022	6258	0.079	5.503	−0.138	0.012	6.449	15
HEP	5835	13815	0.123	1.926	0.185	0.506	7.026	19

In Table 1, N and M are the number of nodes and edges, respectively. β_c is the epidemic threshold. H is the degree heterogeneity, given by $\langle k^2\rangle/\langle k\rangle^2$. $\widetilde{r}$ is assortativity coefficient. C is the clustering coefficient. L is the average shortest path length. D is the diameter of network.

3.2 Evaluation Metrics

As in many literatures, we employ the SIR epidemic model[29] to measure the real spreading capability of the nodes (labeled by σ). To check the spreading influence of one given node, we firstly set this node as the infected node and the other nodes are susceptible nodes. At each time step, each infected node can infect its susceptible neighbors with infection probability β, and then it recovered from the diseases with probability μ (in our work, we set $\mu = 1.0$). This process repeats until there are no any infected nodes. Lastly, the number of recovered nodes is used to reflect the real influence of the nodes. Moreover, to guarantee the reliability of our results, all of these steps are at least averaged over 500 independent realizations.

First, a monotonicity index $M(X)$ for a ranking list X is used to quantify the resolution of different indices [30]:

$$M(X) = \left[1 - \frac{\sum_{c \in V} N_c(N_c - 1)}{N(N-1)}\right]^2, \tag{10}$$

where N_c is the number of nodes with the same index value c. The ranking method is perfectly monotonic and each node is assigned to a different value if $M(X) = 1$; otherwise, all nodes have the same sort value when $M(X) = 0$.

Meanwhile, the Kendall's tau rank correlation coefficient τ is adopted to measure the correlation between one topology-based ranking list and the real spreading capability σ. Let (x_i, y_i) and (x_j, y_j) be a randomly selected pair of joint observations from ranking lists X and Y, respectively. If one has $x_i > x_j$ and $y_i > y_j$ or $x_i < x_j$ and $y_i < y_j$, the observations (x_i, y_i) and (x_j, y_j) 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[31, 32]. τ is defined as

$$\tau = \frac{N_1 - N_2}{0.5N(N-1)}, \tag{11}$$

where N_1 and N_2 are the number of concordant pairs and discordant pairs, respectively.

3.3 Determining Optimal α and γ

In order to obtain the optimal values of α and γ in Equation (7) and Equation (8), simulations on identifying influence are performed in 10 real networks by setting the value of parameter α to increase from 0 to 1 by 0.1, and do the same to parameter γ. The performance of ICC index and semi-IA with the parameter taking different values are shown in Figures 1 and 2, respectively.

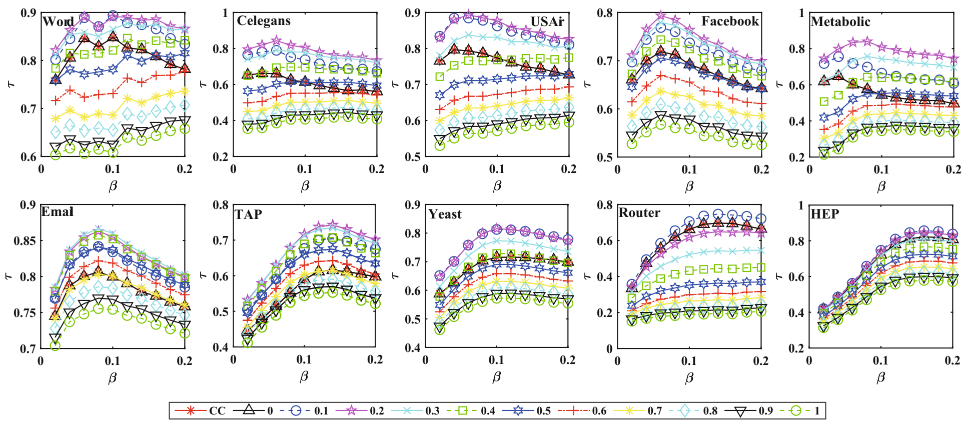

Figure 1 Performance of original CC index and ICC index on different α based on Word, Celegans, USAir, Facebook, Metabolic, Email, TAP, Yeast, Router, HEP. The value of legend represents various parameter value α of ICC index

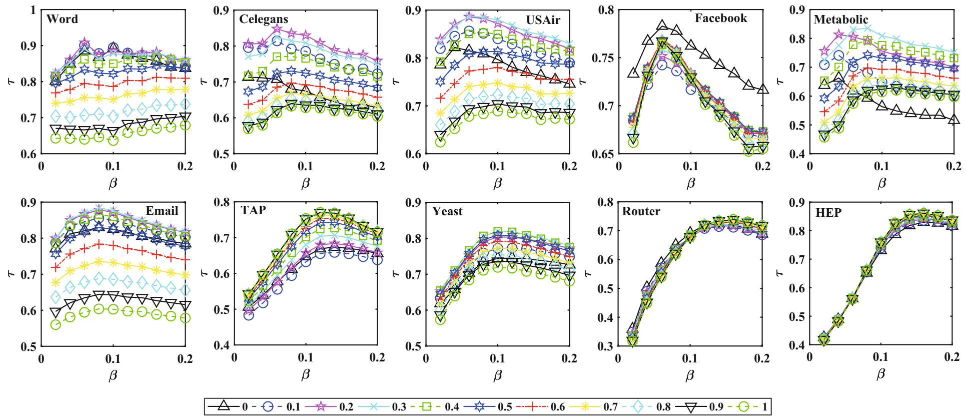

Figure 2 Performance of semi-IA index on different γ based on Word, Celegans, USAir, Facebook, Metabolic, Email, TAP, Yeast, Router, HEP. The value of legend represents various parameter value γ of semi-IA

Generally, it can be observed from Figure 1 and Figure 2 that the ICC index exhibits the better performance when $\alpha = 0.2$ in most situation, although it may perform a bit worse in some specific network when compared to $\alpha = 0.1$ or $\alpha = 0.3$, and semi-IA reveals the better performance when γ is not too large. In order to better explain why we should not choose the larger parameter value α and γ, we first give the definition of the average number of shortest paths length with k as follows:

$$\text{Ave\_num}(k) = \frac{\sum_{i<j} \mathbb{I}(d_{ij} = k) * n_{ij}}{\sum_{i<j} \mathbb{I}(d_{ij} = k)}, \tag{12}$$

where $\mathbb{I}(\cdot)$ is an indicator function. If the predicate is true, $\mathbb{I}(\cdot) = 1$, and $\mathbb{I}(\cdot) = 0$, otherwise.

Taking the schematic network in Figure 3 as an example, in this undirected and unweighted network, we count the NSPs with $d_{ij} = 2$ and $d_{ij} = 3$, and the average number of these shortest paths are computed in Figure 3.

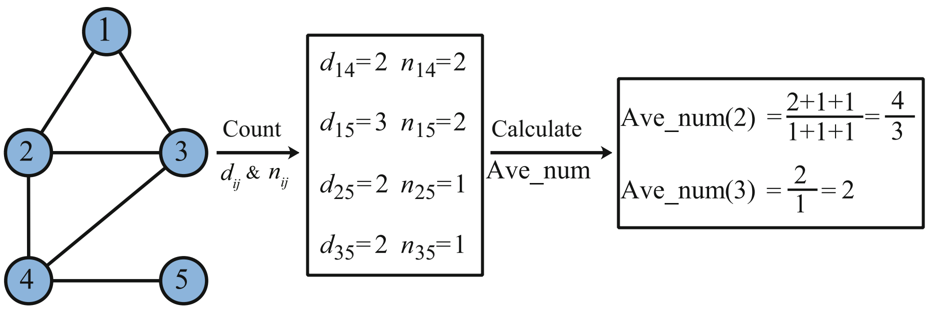

Figure 3 A schematic network is given to illustrate the definition of the average NSPs

According to this calculation method, we count the relationship between the shortest path length and the average NSPs based on above real networks. The results are showed in Figure 4.

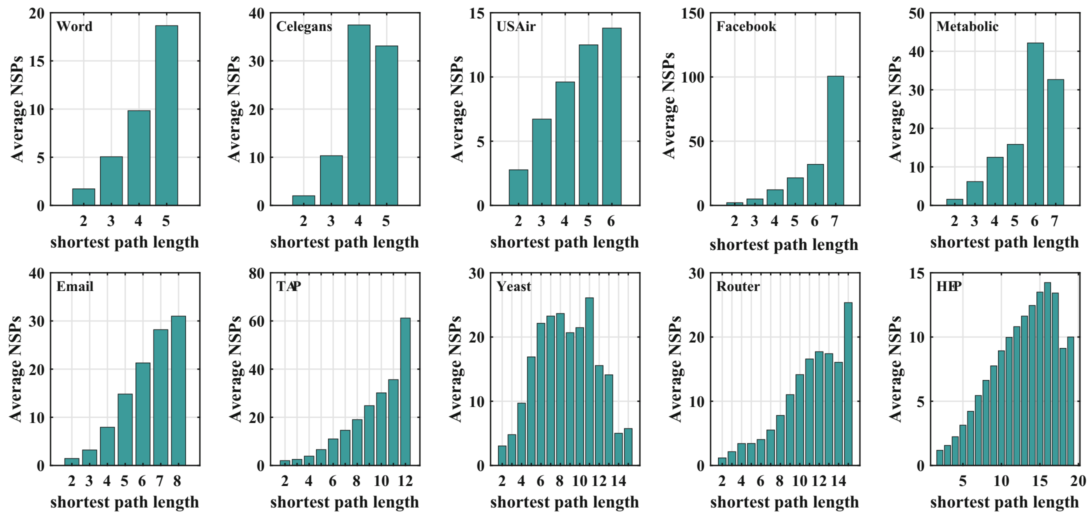

Figure 4 The average NSPs corresponding to each shortest path length in Word, Celegans, USAir, Facebook, Metabolic, Email, TAP, Yeast, Router, HEP

From Figure 4 we can observe that, the average NSPs increases with the shortest path length on the whole. Therefore, we choose $\gamma = 0.2$ as the optimal parameter in a later experiment. If not, a larger parameter means the distant node has a greater impact on this node. As a result, the effectiveness of identifying influential nodes may be worse.

3.4 Convergence Verification of Semi-IA

According to our iterative algorithm, we first give an initial value $X(0)$, then based on the iteration process of Equation (8) and Equation (9) until normalized $X'(t)$ equal to $X'(t-1)$. In this paper, we think the iteration process reaches convergence when the gap between two adjacent iterations is less than 0.00001. We apply the above ten real networks to test whether the semi-IA can reach convergence, and the results in Figure 5 show that the iteration can reach convergence quickly. Therefore, we can use the final iteration values as the influence of the corresponding nodes.

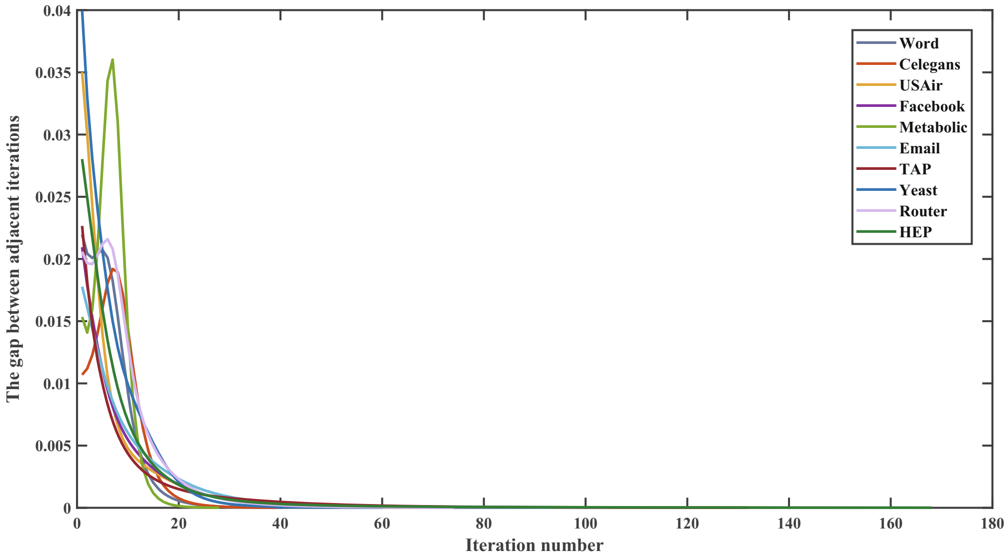

Figure 5 The relationship of the gap between adjacent iterations with the iteration number in Word, Celegans, USAir, Facebook, Metabolic, Email, TAP, Yeast, Router, HEP

3.5 The Performance of Evaluation Metrics

Firstly, the values of $M(\cdot)$ for different indices are compared in Table 2. Results show that

Table 2 $M(\cdot)$ is the monotonicity of the corresponding measures

Network	M(DC)	M(KS)	M(BC)	M(CC)	M(EC)	M(HITs)	M(ICC)	M(semi-IA)
Word	0.8661	0.5990	0.9789	0.9837	**1.0000**	0.9997	0.9997	0.9997
Celegans	0.9217	0.6094	0.9931	0.9893	**0.9999**	0.9977	0.9977	0.9977
USAir	0.8586	0.8114	0.6970	0.9892	**0.9968**	0.9951	0.9952	0.9953
Facebook	0.9315	0.8445	0.9855	0.9953	**1.0000**	0.9999	0.9999	0.9999
Metabolic	0.7922	0.6962	0.8742	0.9900	**0.9993**	0.9989	0.9985	0.9986
Email	0.8874	0.8088	0.9400	0.9988	**0.9999**	**0.9999**	**0.9999**	**0.9999**
TAP	0.8991	0.8380	0.9238	0.9988	**1.0000**	0.9994	0.9994	0.9994
Yeast	0.8314	0.7737	0.8293	0.9988	**0.9993**	0.9992	0.9992	0.9992
Router	0.2886	0.0691	0.3022	0.9961	**0.9977**	0.9965	0.9967	0.9975
HEP	0.7654	0.6303	0.5652	0.9998	**1.0000**	0.9999	0.9999	0.9999

the values of M(ICC), M(semi-IA), M(EC) and M(HITs) are very close to 1 in all 10 networks, and significantly larger than the values of M(DC), M(KS), M(BC) and M(CC). Therefore, ICC index and semi-IA have high resolutions in distinguishing the difference among the nodes in networks. Although the value of $M(\text{EC})$ is the largest, we will find that the performance of EC index in identifying influential nodes is worse than that of ICC index and semi-IA in the following.

Secondly, the correlation value τ as a function of transmission rate β for different indices is compared in Figure 6. We can find that the performances of ICC index and semi-IA are better than other classical indices in a large range of β for most networks. In particular, the advantage of ICC index and semi-IA are more remarkable when β is larger than epidemic threshold β_c (labeled by the dot lines in Figure 6). In Facebook, the performances of the two indices are not as good as EC index, but still better than other indices near β_c. From Figure 2, we can find that $\gamma = 0.2$ is not optimal for Facebook, so the performances of ICC index and semi-IA can be improved through appropriate parameter adjustment. Moreover, the performances of the other indices in ranking the nodes influence, such as BC index, are not significant.

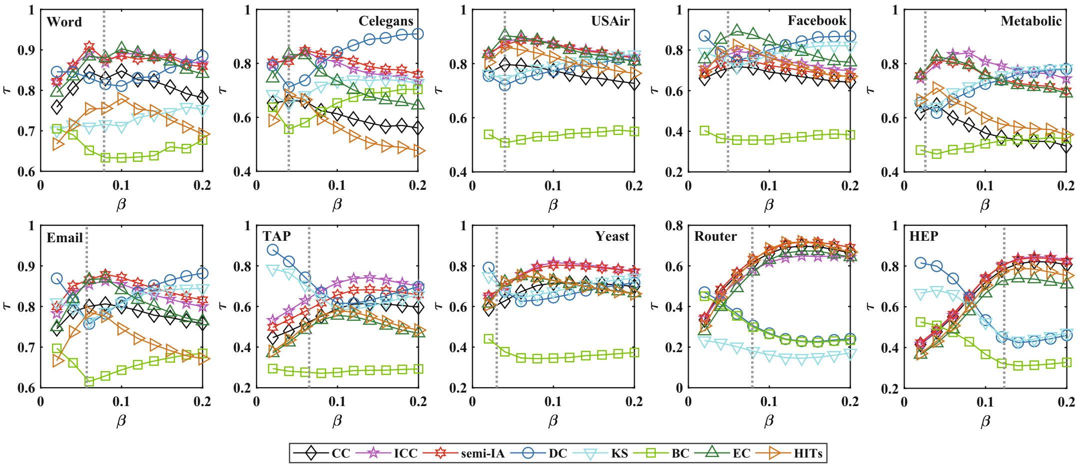

Figure 6 The value of τ obtained by comparing the ranking list generated by the SIR model with the ranking lists generated by the topology-based indices in Word, Celegans, USAir, Facebook, Metabolic, Email, TAP, Yeast, Router and HEP. The dot lines correspond to the epidemic threshold β_c, which are given in Table 1

Thirdly, by setting $\beta = 0.2$ for all networks and all indices, the correlations between the topology-based ranking indices and the real spreading capability (σ) are plotted in Figure 7. Naturally, if one topology-based ranking index can perfectly reflect the real spreading influence of nodes, this index should be significantly correlated with the value of σ. In general, Figure 7 illustrates that DC, KS, BC and CC indices have the lowest correlation with σ (see the left four columns), but EC and HITs indices are a bit better than DC, KS, BC and CC indices (see the fifth and sixth column). Seeing the right two columns in Figure 7, we can find that ICC index and semi-IA have the highest correlation with σ. Therefore, our proposed methods are better than other indices in identifying influential nodes. As mentioned above, ICC index and semi-IA consider the shortest path length and the NSPs simultaneously, as a result, leading to the higher value of correlation, which furtherly explains the rationality and superiority of our methods.

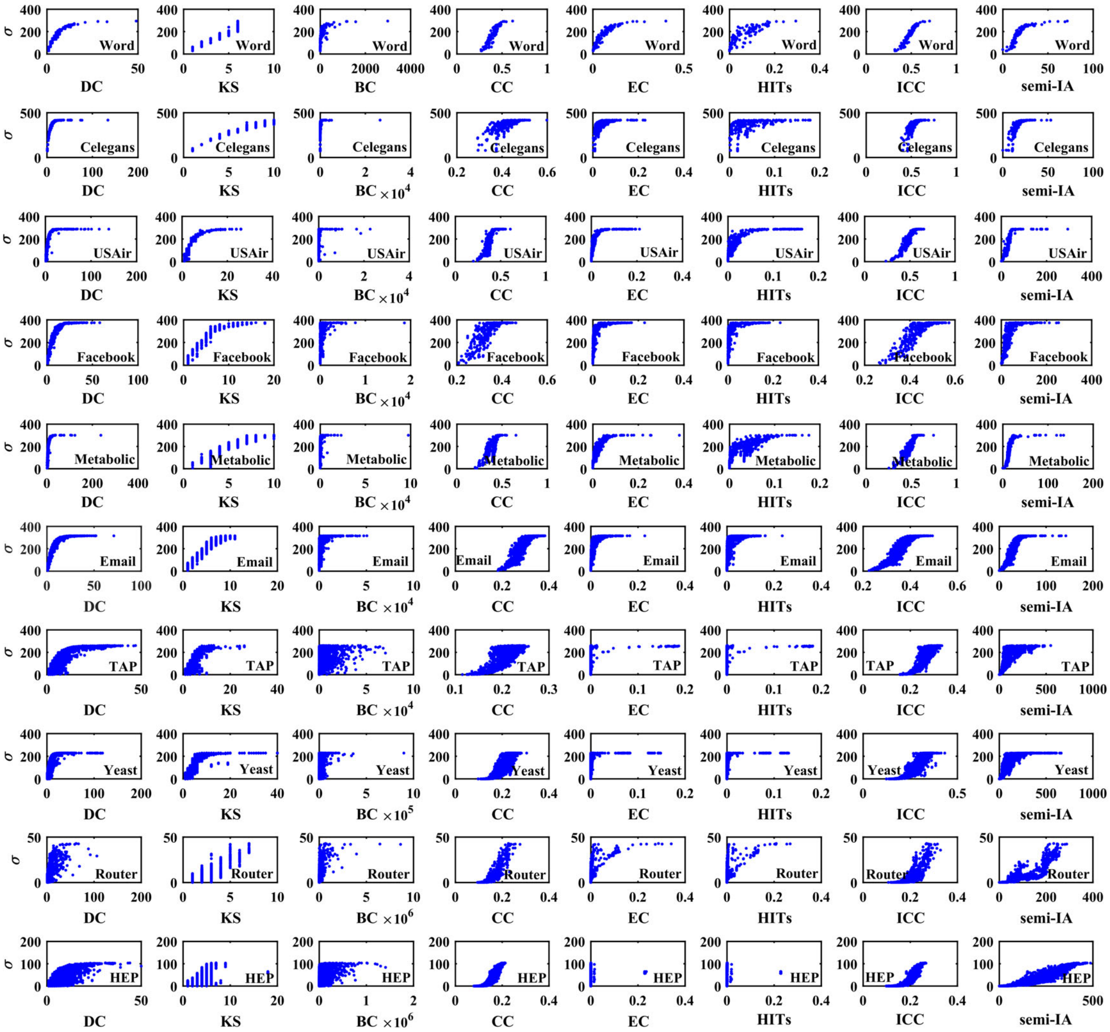

Figure 7 The correlation between the ranking list generated by the SIR model (σ) and the ranking lists generated by topology-based indices in Word, Celegans, USAir, Facebook, Metabolic, Email, TAP, Yeast, Router and HEP. From the left columns to the right columns correspond to the correlations of DC, KS, BC, CC, EC, HITs, ICC and semi-IA with real spreading capability σ, respectively. Here the transmission rate $\beta = 0.2$

Finally, the effect of coverage radius r on the performance of semi-IA is investigated in Figure 8. The results indicate that there exists an optimal value of r (about 2–5) in semi-IA, which breaks the conventional wisdom, namely, the larger value of r leads to the higher value of τ. Through our study, too large value of r may not only decrease the accuracy of identification but also increases the computation complexity. Therefore, our semi-IA by setting $r = 3$ when the scale of the network is less than 1000 and $r = 5$ when the scale of the network is greater than 1000 is sufficient to identify the influential nodes.

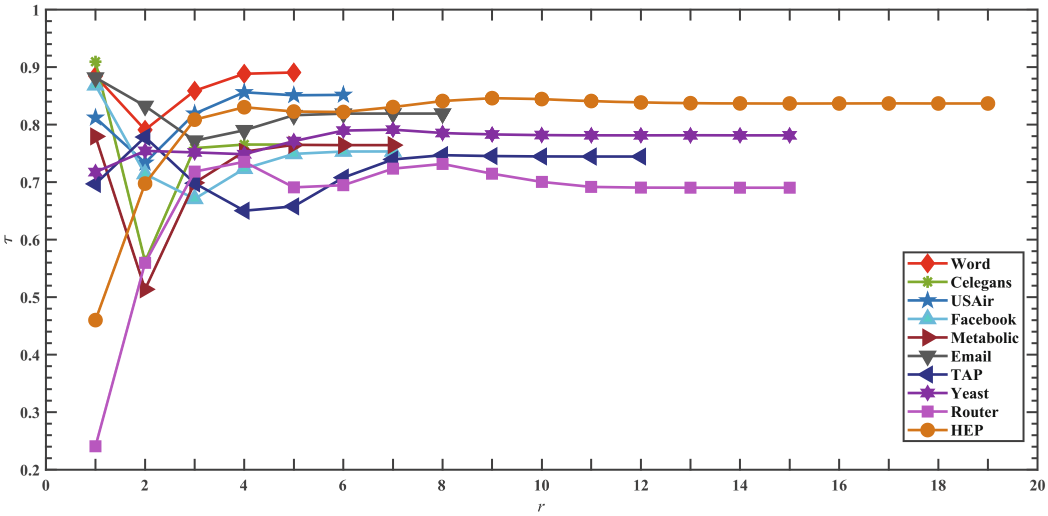

Figure 8 The effect of coverage radius r on the Kendall's tau rank correlation coefficient τ. Here the value of β is 0.2. We should address that the diameter of each network is different, so the changeable range of r for each network is different. The diameter of each network D is given in Table 1

4 Conclusions

In general, based on the closeness centrality, we propose an improved method to identify the influential spreaders by incorporating the shortest path length and the NSPs simultaneously. Enlightened by eigenvector centrality and HITs algorithm, we furtherly propose a semi-local iteration algorithm to measure the influence of nodes. Employing our methods on some real networks by calculating the monotonicity index M, we found that our methods have higher resolutions in distinguishing the different influence of nodes than other indices. Also, by computing Kendall's tau rank correlation coefficient τ and the correlation between the ranking list generated by the SIR model and the ranking lists generated by topology-based indices, we have shown that our methods have a better performance in evaluating and predicting the node's influence in most cases. In addition, our iteration algorithm is a semi-local index, and just considers the node's influence within 3-hops or 5-hops which depend on the size of the networks, because too large value of r may not only decrease the accuracy of identification but also increases the computation complexity. Therefore, our methods provide an effective way to identify the influential spreaders in complex networks.

References

[1] Basaras P, Katsaros D, and Tassiulas L, Detecting influential spreaders in complex, dynamic networks, *Computer*, 2013, **46**(4): 24–29.

[2] Lü L, Zhang Y C, Yeung C H, et al., Leaders in social networks, the delicious case, *PLoS One*, 2011, **6**(6): e21202.

[3] Borge-Holthoefer J and Moreno Y, Absence of influential spreaders in rumor dynamics, *Physical Review E*, 2012, **85**(2): 026116.

[4] Wang P, Tian C, and Lu J, Identifying influential spreaders in artificial complex networks, *Journal of Systems Science and Complexity*, 2014, **27**(4): 650–665.

[5] Chen D B, Xiao R, Zeng A, et al., Path diversity improves the identification of influential spreaders, *EPL* (*Europhysics Letters*), 2013, **104**(6): 68006.

[6] Ren Z M, Zeng A, Chen D B, et al., Iterative resource allocation for ranking spreaders in complex networks, *EPL* (*Europhysics Letters*), 2014, **106**(4): 48805.

[7] Ahmad Z and Amir S, A hierarchical approach for influential node ranking in complex social networks, *Expert Systems with Applications*, 2018, **93**: 200–211.

[8] Sotoodeh H and Falahrad M, Relative degree structural hole centrality, $\mathrm{C}_{\mathrm{RD-SH}}$: A new centrality measure in complex networks, *Journal of Systems Science and Complexity*, 2019, **32**(5): 1306–1323.

[9] Tang J, Zhang R, Wang P, et al., A discrete shuffled frog-leaping algorithm to identify influential nodes for influence maximization in social networks, *Knowledge-Based Systems*, 2020, **187**: 104833.

[10] Bonacich P, Factoring and weighting approaches to status scores and clique identification, *Journal of Mathematical Sociology*, 1972, **2**(1): 113–120.

[11] Kitsak M, Gallos L K, Havlin S, et al.A, Identification of influential spreaders in complex networks, *Nature Physics*, 2010, **6**(11): 888–893.

[12] Freeman L C, A set of measures of centrality based on betweenness, *Sociometry*, 1977, **40**(1): 35–41.

[13] Bonacich P, Power and centrality: A family of measures, *Journal of Mathematical Sociology*, 1987, **92**(1): 1170–1182.

[14] Zeng A and Zhang C J, Ranking spreaders by decomposing complex networks, *Physics Letters A*, 2013, **377**(14): 1031–1035.

[15] Sabidussi G, The centrality index of a graph, *Psychometrika*, 1966, **31**(4): 581–603.

[16] Ma L L, Ma C, Zhang H F, et al., Identifying influential spreaders in complex networks based on gravity formula, *Physica A*, 2016, **451**: 205–212.

[17] Liu H L, Ma C, Xiang B B, et al., Identifying multiple influential spreaders based on generalized closeness centrality, *Physica A*, 2018, **492**(1): 2237–2248.

[18] Bao Z K, Ma C, Xiang B B, et al., Identification of influential nodes in complex networks: Method from spreading probability viewpoint, *Physica A*, 2017, **468**: 391–397.

[19] Benzi M, Estrada E, and Klymko C, Ranking hubs and authorities using matrix functions, *Linear Algebra and Its Applications*, 2013, **438**(5): 2447–2474.

[20] Liu Y, Tang M, Zhou T, et al., Core-like groups result in invalidation of identifying superspreader

by k-shell decomposition, *Scientific Reports*, 2015, **5**: 9602.

[21] Newman M E, Finding community structure in networks using the eigenvectors of matrices, *Physical Review E*, 2006, **74**(3): 036104.

[22] Watts D J and Strogatz S H, Collective dynamics of small-worldnetworks, *Nature*, 1998, **393**(6684): 440–442.

[23] Guimera R, Danon L, Diaz-Guilera A, et al., Self-similar community structure in a network of human interactions, *Physical Review E*, 2003, **68**(6): 065103.

[24] Blagus N, Subelj L, Bajec M, Self-similar scaling of density in complex real-world networks, *Physica A*, 2012, **391**(8): 2794–2802.

[25] Duch J and Arenas A, Community detection in complex networks using extremal optimization, *Physical Review E*, 2005, **72**: 027104.

[26] Mering C V, Krause R, Snel B, et al., Comparative assessment of large-scale data sets of protein-protein interactions, *Nature*, 2002, **417**: 399–403.

[27] Spring N, Mahajan R, Wetherall D, et al., Measuring ISP topologies with rocketfuel, *IEEE/ACM Transactions on Networking*, 2004, **12**(1): 2–16.

[28] Leskovec J, Kleinberg J, and Faloutsos C, Graph evolution: Densification and shrinking diameters, *ACM Transactions on Knowledge Discovery from Data* (*TKDD*), 2007, **1**(1): 2–42.

[29] Moreno Y, Pastor-Satorras R, and Vespignani A, Epidemic outbreaks in complex heterogeneous networks, *The European Physical Journal B*, 2002, **26**(4): 521–529.

[30] Bae J and Kim S, Identifying and ranking influential spreaders in complex networks by neighborhood coreness, *Physica A*, 2014, **395**: 549–559.

[31] Knight W R, A computer method for calculating kendall's tau with ungrouped data, *Journal of the American Statistical Association*, 1966, **61**(314): 436–439.

[32] Zhou T, Lü L, and Zhang Y C, Predicting missing links via local information, *The European Physical Journal B*, 2009, **71**(4): 623–630.