

LCH: A local clustering H-index centrality measure for identifying and ranking influential nodes in complex networks

Gui-Qiong Xu(徐桂琼), Lei Meng(孟蕾), Deng-Qin Tu(涂登琴), and Ping-Le Yang(杨平乐)
Citation: Chin. Phys. B, 2021, 30 (**8**): 088901. DOI: 10.1088/1674-1056/abea86
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Detection of influential nodes with multi-scale information

Jing-En Wang(王静恩), San-Yang Liu(刘三阳), Ahmed Aljmiai, and Yi-Guang Bai(白艺光)
Chin. Phys. B, 2021, 30 (**8**): 088902. DOI: 10.1088/1674-1056/abff2d

Exploring individuals' effective preventive measures against epidemics through reinforcement learning

Ya-Peng Cui(崔亚鹏), Shun-Jiang Ni (倪顺江), and Shi-Fei Shen(申世飞)
Chin. Phys. B, 2021, 30 (**4**): 048901. DOI: 10.1088/1674-1056/abcfa5

Finite density scaling laws of condensation phase transition in zero-range processes on scale-free networks

Guifeng Su(苏桂锋), Xiaowen Li(李晓温), Xiaobing Zhang(张小兵), Yi Zhang(张一)
Chin. Phys. B, 2020, 29 (**8**): 088904. DOI: 10.1088/1674-1056/ab8a41

Influential nodes identification in complex networks based on global and local information

Yuan-Zhi Yang(杨远志), Min Hu(胡敏), Tai-Yu Huang(黄泰愚)
Chin. Phys. B, 2020, 29 (**8**): 088903. DOI: 10.1088/1674-1056/ab969f

Identifying influential spreaders in complex networks based on entropy weight method and gravity law

Xiao-Li Yan(闫小丽), Ya-Peng Cui(崔亚鹏), Shun-Jiang Ni(倪顺江)
Chin. Phys. B, 2020, 29 (**4**): 048902. DOI: 10.1088/1674-1056/ab77fe

---

LCH: A local clustering H-index centrality measure for identifying and ranking influential nodes in complex networks*

Gui-Qiong Xu(徐桂琼)[1,†], Lei Meng(孟蕾)[1], Deng-Qin Tu(涂登琴)[1,‡], and Ping-Le Yang(杨平乐)[2,3]

[1] *Department of Information Management, School of Management, Shanghai University, Shanghai* 200444, *China*
[2] *Business School, University of Shanghai for Science and Technology, Shanghai* 200093, *China*
[3] *School of Electrical and Information Engineering, Jiangsu University of Science and Technology Zhangjiagang* 215600, *China*

(Received 26 November 2020; revised manuscript received 2 February 2021; accepted manuscript online 1 March 2021)

Identifying influential nodes in complex networks is one of the most significant and challenging issues, which may contribute to optimizing the network structure, controlling the process of epidemic spreading and accelerating information diffusion. The node importance ranking measures based on global information are not suitable for large-scale networks due to their high computational complexity. Moreover, they do not take into account the impact of network topology evolution over time, resulting in limitations in some applications. Based on local information of networks, a local clustering H-index (LCH) centrality measure is proposed, which considers neighborhood topology, the quantity and quality of neighbor nodes simultaneously. The proposed measure only needs the information of first-order and second-order neighbor nodes of networks, thus it has nearly linear time complexity and can be applicable to large-scale networks. In order to test the proposed measure, we adopt the susceptible-infected-recovered (SIR) and susceptible-infected (SI) models to simulate the spreading process. A series of experimental results on eight real-world networks illustrate that the proposed LCH can identify and rank influential nodes more accurately than several classical and state-of-the-art measures.

Keywords: complex networks, influential nodes, local structure, susceptible infected recovered model, susceptible infected model

PACS: 89.75.Fb **DOI:** 10.1088/1674-1056/abea86

1. Introduction

Identifying and ranking influential nodes are one kind of the most fundamental problems in modern network science, which may help us to understand the structures, functions, and characteristics of networks.[1,2] For instance, removing critical nodes may cause relatively large damage to complex communication networks or transportation networks,[3,4] thus identification of vital nodes may contribute to research of network robustness and vulnerability. Locating influential nodes helps us to understand and control the processes such as epidemic disease outbreak,[5] virus propagation[6] and rumor spreading,[7] and it also plays an important role in viral marketing or accelerating information or knowledge diffusion.[8–11]

So far, many achievements have been made in searching vital nodes and a variety of effective methods have been proposed. Among them, degree centrality (DC), betweenness centrality (BC) and closeness centrality (CC) are three classical measures.[12] Like BC and CC, Katz centrality is defined by considering all the paths in the network. The random-walk-based PageRank[13] and its enhancement LeaderRank[14] can perform well in directed networks. Other representative centrality indices include K-shell[15] and its variants or extensions.[16–18] In addition, multi-attribute decision making (MADM) method is proven to be an alternative tool in evaluating nodes' importance.[19–23] Local centrality measures are widely used in various applications due to their low computational demands, such as the semi-local centrality (SLC) developed by Chen and his coworkers.[24]

The topological connections among neighbors have great impact in spreading abilities of nodes. Based on this, the ClusterRank[25] and the local structural centrality[26] take into account the influence of neighbors and node itself as well as the role of local clustering. Li *et al.*[27] developed a new centrality named as clustered local degree (CLD) by combining degree with topological connections of nodes. Berahmand *et al.*[28] presented a new local centrality measure by considering the negative role of the first-order neighbors and positive role of second-order neighbors. In this direction, researchers proposed a great number of local centrality measures from different perspectives.[29–32]

H-index is widely used to measure scientific achievements of scholars.[33] By analyzing the dynamic process of network structures, Lü *et al.*[34] pointed out that the H-index can be regarded as a good balance between coreness and degree centrality in identifying influential nodes. H-index focuses on the high-influence neighbors of the node, while some of the low-influence neighbors are ignored. To overcome

*Project supported by the National Natural Foundation of China (Grant No. 11871328) and the Shanghai Science and Technology Development Funds Soft Science Research Project (Grant No. 21692109800).

†Corresponding author. E-mail: xugq@staff.shu.edu.cn
‡Corresponding author. E-mail: shumse724@shu.edu.cn

© 2021 Chinese Physical Society and IOP Publishing Ltd

the limitation, some enhancements and extensions of H-index have been developed, including the local H-index centrality (LH),[35] the extended H-index centrality (EHC) measure.[36] However, there is still a lot of room to improve the local centrality measures.

Inspired by the above work, we propose a local clustering H-index (LCH) centrality measure for ranking and identifying influential nodes in complex networks, which aims at finding a compromise between accuracy, extensive applicability and low computational cost. The proposed measure is based on three aspects. Firstly, LCH only needs the information of neighboring nodes in two hops and thus it can be applicable to large-scale networks due to its near linear time complexity. Secondly, we consider that the clustering coefficient of a node has negative effect on information spreading. In other words, the higher the clustering coefficient of a node is, the more limited the node spreading is. Lastly, the H-index puts emphasis on neighbor nodes with high-influence and neglects other neighbor nodes with low influence, so we consider both the quantity and quality of neighbor nodes in the new measure.

The following sections are organized as follows. In Section 2, we first introduce several centrality measures for comparison analysis. In Section 3, the detail of the proposed LCH is illustrated by an example network. Section 4 is devoted to the data sets and evaluation methods. The experimental results are presented and analyzed in Section 5. Finally, in Section 6, we discuss and summarize our results.

2. Centrality measure

Before proceeding with the discussion of the proposed method, we devote this section to the concepts of several centrality measures, including degree, clustered local-degree centrality, K-shell, H-index as well as local h-index centrality for undirected networks. For a given network $G = \langle V, E \rangle$ with n nodes, where V and E being the node set and edge set, the connections between nodes are expressed by adjacency matrix $A = \{a_{ij}\}_{n\times n}$. If there exists a link between node i and node j, $a_{ij} = 1$. On the contrary, $a_{ij} = 0$.

2.1. Degree centrality

The degree centrality (DC) of node i is defined as

$$\mathrm{DC}_i = \sum_{j=1}^{n} a_{ij}, \tag{1}$$

which is the number of the direct neighbors of node i.[12] DC is a simple but effective measure for evaluating nodes' importance in networks.

2.2. Clustered local-degree centrality

The clustered local-degree centrality (CLD)[27] measures the importance of a node according to the degree of neighbors and topological connections among neighbors. The CLD of node i is expressed as

$$\mathrm{CLD}_i = (1 + C_i) \sum_{j\in\Gamma_i} \mathrm{DC}_j, \tag{2}$$

where C_i denotes the clustering coefficient of node i, Γ_i is the neighbor set of node i, and DC_j is the degree of node j.

2.3. K-shell decomposition

As a global measure, the K-shell (KS) decomposition[15] takes full advantage of position information of nodes. The key idea of this method is that the nodes at the core are regarded as the most influential nodes. The K-shell decomposition can be proceeded as follows: Firstly, remove these nodes whose degrees are equal to 1. When some new nodes with $\mathrm{DC} = 1$ appear, continue to remove them from the network until there are only remained nodes whose degree is larger than 1. At this point, the K-shell values of all removed nodes are set to 1. Next, all nodes with two direct neighbors are removed in the same manner. Finally, repeat procedures until all nodes are removed from the network. After decomposition, each node is assigned with a K-shell value, which denotes the location attribute of the node in the network.

2.4. H-index

H-index (H) was initially used to evaluate the influence of scholars based on the citation of their publications. Lü *et al.*[34] extended it to detect vital nodes in complex networks. Suppose that node i has k_i neighbors, its H is given by

$$H_i = \mathscr{H}(k_{j1}, k_{j2}, \ldots, k_{js}, \ldots, k_{jk_i}), \tag{3}$$

where k_{js} denotes the degree of the s-th neighbor of node i. In Eq. (3), the operator $\mathscr{H}$ returns the maximum integer v such that there are at least v neighbors whose degree is not lower than v.

2.5. Local h-index centrality

The local h-index (LH) centrality[35] quantifies the influence of a node by accumulating the H of the node itself and its direct neighbors, which is defined as

$$\mathrm{LH}_i = H_i + \sum_{j\in\Gamma_i} H_j \tag{4}$$

with H_i and Γ_i being the H-index and neighbor set of node i, respectively.

3. The proposed method

Generally speaking, global centrality measures are always computationally demanding and not suitable for large-scale networks. Christakis and Fowler[37] argued that, for most social networks, the behavior of a node is highly related to its first-order neighbors, second-order neighbors and up to the

third-order neighbors. As a good balance between degree centrality and coreness, the H-index[34] can rank nodes more accurately because it takes the influence of the nearest neighbors into account. As an extension of H-index, the LH[35] is defined through accumulating the H-index values of all direct neighbor nodes and the node itself. Even though the LH performs well in terms of discrimination, there is still room for improvement and can be further extended.

On the one hand, the topological connections among neighbors always have great impact on information spreading. Usually, if neighbors of a node are closely connected, its influence may be limited to a small scope. In this sense, the clustering coefficient has a negative effect on spreading in a local area.[25,38–40] For the example network shown in Fig. 1, nodes 2 and 6 have the same degree centrality and H, but the spreading capabilities of nodes 2 and 6 are 4.06 and 4.36 by averaging 100 independent experiments of the SIR model. That is, node 6 with lower clustering coefficient is more influential than node 2.

On the other hand, LH only considers the high-influence neighbors and neglects other neighbor nodes with low influence. Take node 7 as an example in Fig. 1, node 7 has four neighbors. Except node 6, the H-index values of the other three neighbor nodes are 1. Nevertheless, as observed, these nodes are affected by more than one node. That is to say, the influence of node 7 will be underestimated.

Based on the above two points, we put forward a novel measure called local clustered H-index centrality (LCH), which considers the number and influence of neighbor nodes, and the topological connections among neighbors. The LCH of node i is defined as

$$\mathrm{LCH}_i = \frac{1}{\langle H\rangle}\times\frac{H_i}{1+C_i} + \sum_{j\in\Gamma_i}\left(\frac{1}{\langle H\rangle}\times\frac{H_j}{1+C_j}+\frac{1}{\langle k\rangle}\times \mathrm{DC}_j\right), \quad (5)$$

where H_i and C_i are the H-index and the clustering coefficient of node i. Γ_i denotes the neighbor set of node i, and DC_j is the degree value of node j; $\langle H\rangle$ and $\langle k\rangle$ represent the average H and average degree of the given network, respectively.

Without loss of generality, one may multiply $\langle H\rangle$ for each term on the right side of Eq. (5). Thus, the proposed LCH can be simplified to

$$\mathrm{LCH}_i = \frac{H_i}{1+C_i} + \sum_{j\in\Gamma_i}\left(\frac{H_j}{1+C_j}+\mu\times \mathrm{DC}_j\right), \quad (6)$$

where $\mu = \langle H\rangle/\langle k\rangle$. The LCH contains two parts, one is the node own influence, and the other part is the influence of the nearest neighbors on it. Since the values of DC and H-index are not comparable in order of magnitude, the parameter μ can adjust the trade-off between the two centrality measures.[18,41]

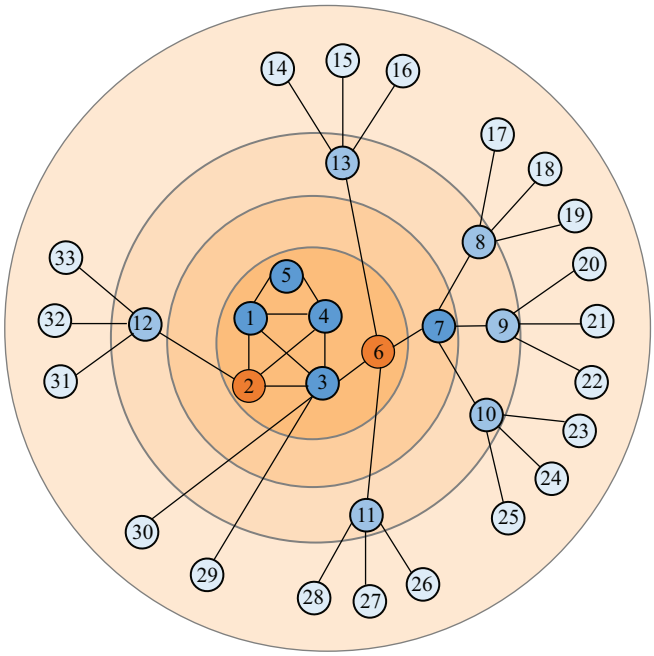


Fig. 1. An example network with 33 nodes and 36 edges. Note that the spreading ability is obtained by averaging 100 independent experiments of the SIR model, the infection probability β is set as the epidemic threshold β_{th}.

In order to further illustrate the proposed measure, we take nodes 2 and 6 as an example. These two nodes have the same H-index value ($H_2 = H_6 = 4$), and the sum of degree of node-2's neighbors is equal to that of node-6's ($\sum_{i\in\Gamma_2}\mathrm{DC}_i = \sum_{j\in\Gamma_6}\mathrm{DC}_j = 18$). At the same time, the sum of H-index of node-2's neighbors is slightly larger than that of node-6's neighbors ($\sum_{i\in\Gamma_2}H_i = 11 > \sum_{j\in\Gamma_6}H_j = 10$). After some calculations, it is easily obtained that $\mathrm{LCH}_2 = 23.1$, $\mathrm{LCH}_6 = 25.83$. Node 6 is expected to have more influence than node 2. By averaging 100 independent experiments of the SIR model, the spreading capabilities of nodes 2 and 6 are 4.06 and 4.36, which is consistent with the rank obtained by LCH. In fact, node 6 has larger spreading ability mainly due to two reasons. Firstly, node 6 with lower clustering coefficient can spread the information out quickly, while the influence of node 2 is confined in a small scope. Secondly, if the hub node 6 is removed, the example network will be divided into two large components. However, it will have a relatively smaller influence on the network structure if node 2 is removed.

The LCH is an easily computable index, its computation mainly depends on the H, clustering coefficient as well as degree of nodes. Suppose that n and $\langle k\rangle$ are the node number and average degree in a given network. The time costs for H and clustering coefficient of a node are both $O(\langle k\rangle^2)$, and the time cost of degree centrality is $O(\langle k\rangle)$. Then the total time cost is $O(2n\langle k\rangle^3 + 3n\langle k\rangle^2)$. Therefore, the asymptotic time complexity of the proposed measure is $O(n\langle k\rangle^3)$. Generally speaking, the average degree $\langle k\rangle$ is far less than the total number n of nodes, so the near linear computational complexity makes the proposed measure LCH scalable for large-scale networks.

4. Experimental setup

4.1. Data sets

In this work, eight real-world networks are chosen to test the performance of the proposed measure: (i) Lesmis: the co-appearance network of characters in the novel Les Miserables. (ii) Netscience: the co-author network in network science consists of 1589 nodes. We choose the largest component of this network with 379 nodes. (iii) Roget: the co-appearance network containing the cross-references in Roget's Thesaurus. (iv) Email: the email network of the University Rovirai Virgili. (v) CSphd: the genealogy network of Ph.D. students and their advisors in theoretical computer science. (vi) Yeast: the protein-protein interaction network. (vii) Power: the network of the power grid. (viii) Ca-hepth: the collaboration network of Arixv in the area of high energy physics theory. The datasets can be downloaded from http://www-personal.umich.edu/~mejn/netdata/, http://vlado.fmf.uni-lj.si/pub/networks/data. Table 1 gives the statistical features of the eight networks with different sizes.

Table 1. Some statistical properties of the eight real-world networks. The topological characteristics include: node number ($|V|$), edge number ($|E|$), average degree ($\langle k\rangle$), maximum degree ($k_{\max}$), clustering coefficient ($\langle C\rangle$) and epidemic threshold $\beta_{\rm th} = \langle k\rangle/\langle k^2\rangle$.[42]

Network	$\lvert V\rvert$	$\lvert E\rvert$	$\langle k\rangle$	$k_{\max}$	$\langle C\rangle$	$\beta_{\rm th}$
Lesmis	77	254	6.5974	36	0.5731	0.0830
Netscience	379	914	4.8232	34	0.7412	0.1247
Roget	1022	5075	7.1389	28	0.1498	0.0950
Email	1133	5451	9.6222	71	0.2202	0.0535
CSphd	1882	1740	1.8491	46	0.0051	0.2080
Yeast	2361	7182	5.6289	64	0.1301	0.0603
Power	4941	6594	2.6691	19	0.0801	0.2583
Ca-hepth	9877	25998	5.2593	65	0.4714	0.0798

4.2. The spreading model

To compare different measures for identifying crucial nodes, the susceptible-infected-recovered (SIR)[43,44] and susceptible-infected (SI) models[45] are adopted to simulate the spread and evolution in real networks. The SIR model contains three types of nodes, susceptible (S), infected (I) and recovered (R). The final state of SIR dynamics is (i) all nodes are either in S or R state or (ii) all nodes had been infected and got recovered. The SIR model contains two parameters, one is the recovery probability α, the other is infection probability β. For simplicity, we take $\alpha = 1$. When $\alpha = 0$, the SIR model is converted to the SI model. Different from the SIR model, the SI model only contains two types of nodes, susceptible (S) and infected (I). Once the node is infected, it will never be recovered.

In this paper, $F(t)$ denotes the sum of the number of the infected nodes and the recovered nodes at time t. With the increase of time t the spreading process will reach stable gradually. $F(t_c)$ stands for the spreading capability of the source node at the time t_c in a steady state. Since there is no node recovery process in the SI model, the number of recovery nodes is set to 0. Here, $F(t)$ $(t < t_c)$ represents the spreading capability of the source node in the SI model.

4.3. Evaluation method

Kendall's tau correlation coefficient[46] is chosen to determine the consistency between the ranking list obtained by a specific method and the $F(t_c)$ ranking list obtained by the SIR model. Given two ranking lists X and Y, (x_i, y_i) and (x_j, y_j) are two node pairs in these lists. The Kendall tau correlation coefficient is defined as

$$\tau(X,Y) = \frac{n_1 - n_2}{\frac{1}{2}N(N-1)}, \tag{7}$$

where n_1 and n_2 denote the number of concordant pairs and discordant pairs, and N is the length of the list X. Obviously, τ lies down between -1 and 1. Here, $\tau = -1$ indicates that the two lists are exactly in the opposite order. By contrast, the order of the two lists is identical when $\tau = 1$.

In some applications, one may focus on the top-rank nodes instead of all nodes. Imprecision function can measure how close the top-rank nodes by a specific method is to the spreading ability list by the SI model, which is given by

$$\varepsilon(p) = 1 - \frac{M(p)}{M_{\rm eff}(p)}, \tag{8}$$

where p is the proportion of the inspected nodes to the network scale N. In addition, $M(p)$ and $M_{\rm eff}(p)$ represent the average spreading abilities of pN nodes obtained by a specific method and the SI model. Note that a lower imprecision value indicates that the considered method is more accurate.

Monotonicity is employed to evaluate the discrimination capacity of different measures, which is defined by

$$M(R) = \left(1 - \frac{\sum_{r\in R} n_r(n_r - 1)}{n(n-1)}\right)^2, \tag{9}$$

where n is the length of the ranking list R, and n_r denotes the number of nodes with the same ranking value r. The M value ranges from 0 to 1. The best value for M is 1, which means that each node in a network has a unique ranking value. In contrast, the worst value for M is 0, which indicates that all nodes in a network have the same ranking.

5. Experimental results and analysis

In the experiments, we compare the proposed LCH with five centrality measures including DC,[12] CLD,[27] KS,[15] H[34] and LH.[35] The EHC,[36] another extension of the H, also takes into account the degrees of the neighbors of nodes and its accuracy is improved. The LCH will not be compared with EHC because the latter one needs to specify two parameters.

5.1. Kendall's tau correlation coefficient

To test the accuracy of the proposed measure, we will compare it with other five different centrality measures, with the SIR model. In principle, the closer the value of τ is to 1, the ranking method has better performance. Note that the infection probability β is a key parameter in the SIR model. If β is too small, the spreading of source node may be restricted in a relatively limited scope. By contrast, if β is too large, the spreading may cover the entire network. Thus, the β value changes around the epidemic threshold β_{th}. The β_{th} value is closely related with the topological characteristic of a given network, so the abscissa is not the same for different networks in Fig. 2.

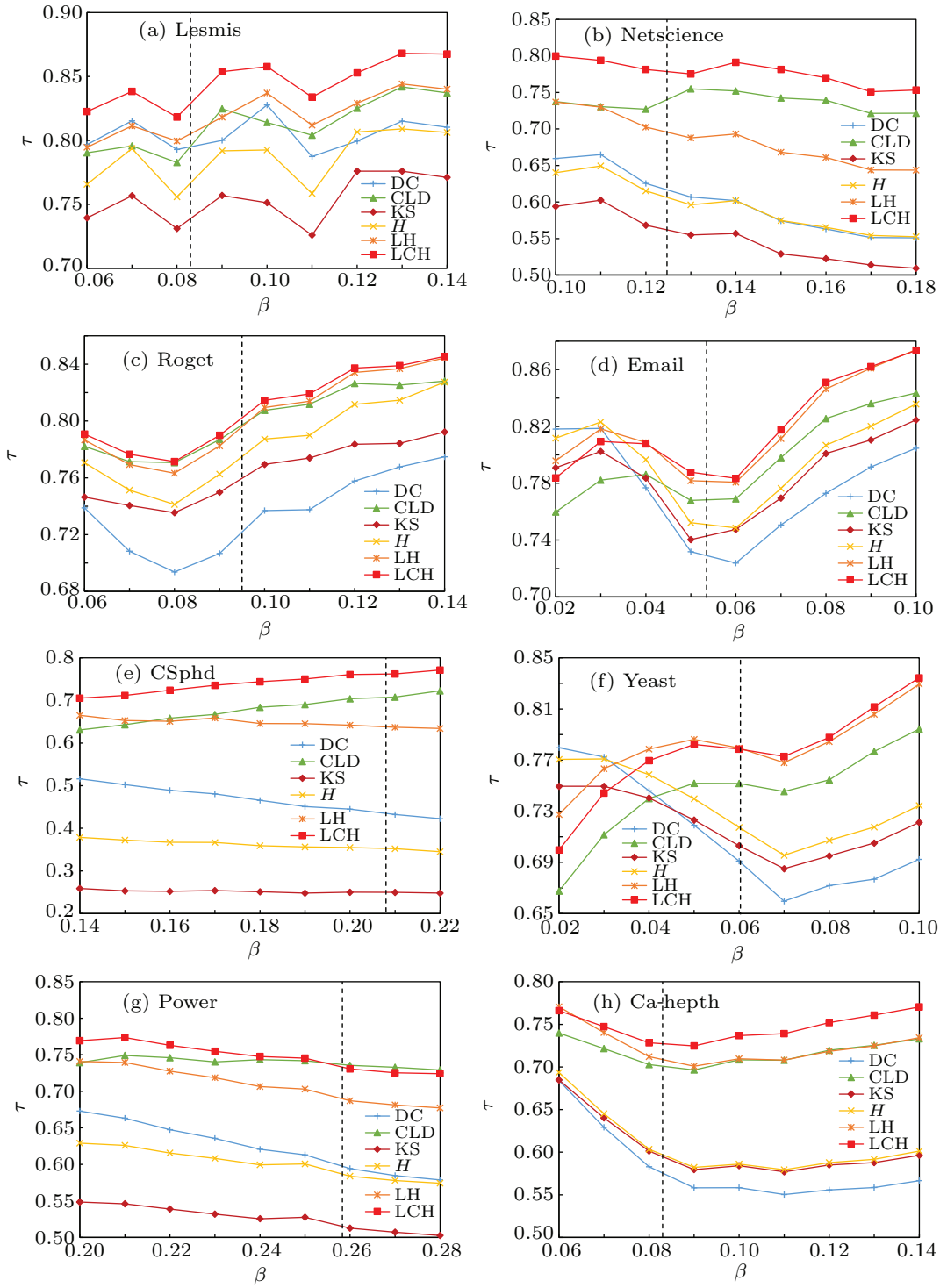


Fig. 2. Kendall's tau correlation coefficient (τ) between the ranking list obtained by six different measures and the $F(t_c)$ ranking list by the SIR model with different infection probability β, and β is near β_{th}. The results are obtained by averaging 100 independent experiments of the SIR model, and the vertical dotted line represents the epidemic threshold β_{th}. (a) Lesmis, (b) Netscience, (c) Roget, (d) Email, (e) CSphd, (f) Yeast, (g) Power, (h) Ca-hepth.

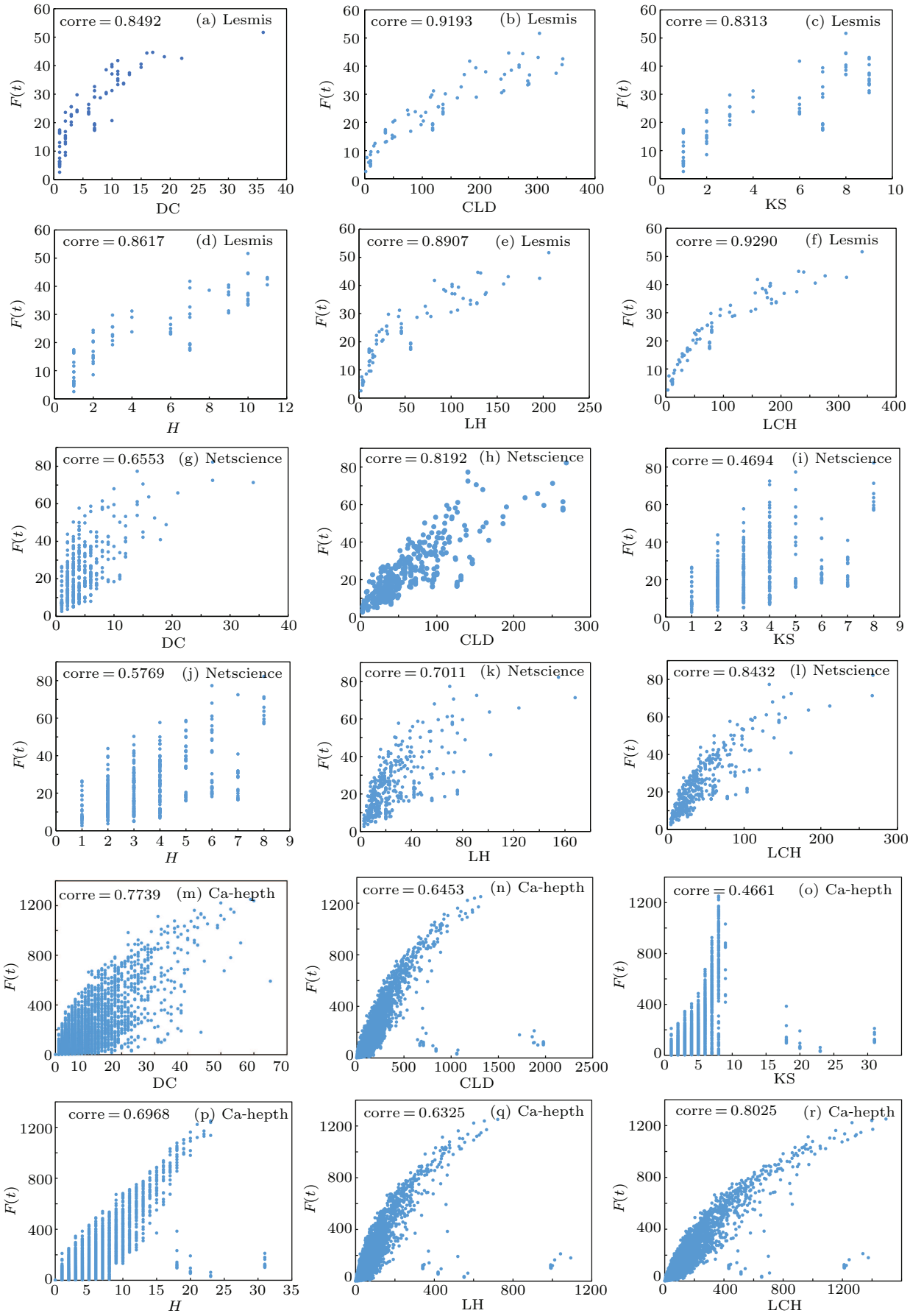


Fig. 3. The correlation between nodes' centrality values and their spreading abilities on three networks. The horizontal axis corresponds to nodes' centrality values, and the vertical axis denotes the stable spreading abilities $F(t)$ $(t = 10)$. The results are obtained by averaging 100 independent experiments of the SI model for the particular case $\beta = \beta_{\text{th}}$.

Figure 2 depicts the Kendall tau correlation coefficient of the six centrality measures with the change of infection probability. As shown in Figs. 2(a), 2(b), 2(e), and 2(h) the value of τ for LCH is higher than other five measures within the entire infection probability range. Thus, the LCH gives the best performance in the Lesmis, Netscience, CSphd and Ca-hepth networks. It can be seen in Fig. 2(c) that the values of τ for CLD, LH and LCH are very close, but LCH is always superior to the other two measures. In the Email and Yeast networks as shown in Figs. 2(d) and 2(f), DC, H and LH outperform other measures when β is not larger than 0.03. This is because the influence of nodes is limited to a smaller scope if β is taken as a relatively small value. The performance of DC and H decrease gradually with increasing values of β, but the performance of the LCH and LH is improved. Moreover, the LCH is slightly better than the LH when β exceeds the epidemic threshold. In the Power network as shown in Fig. 2(g), LCH and CLD are superior to the other four centrality measures. However, CLD performs slightly better than our propose method when β is larger than 0.26. On the whole, the LCH can identify influential nodes more accurately in most cases compared with the other five ranking methods.

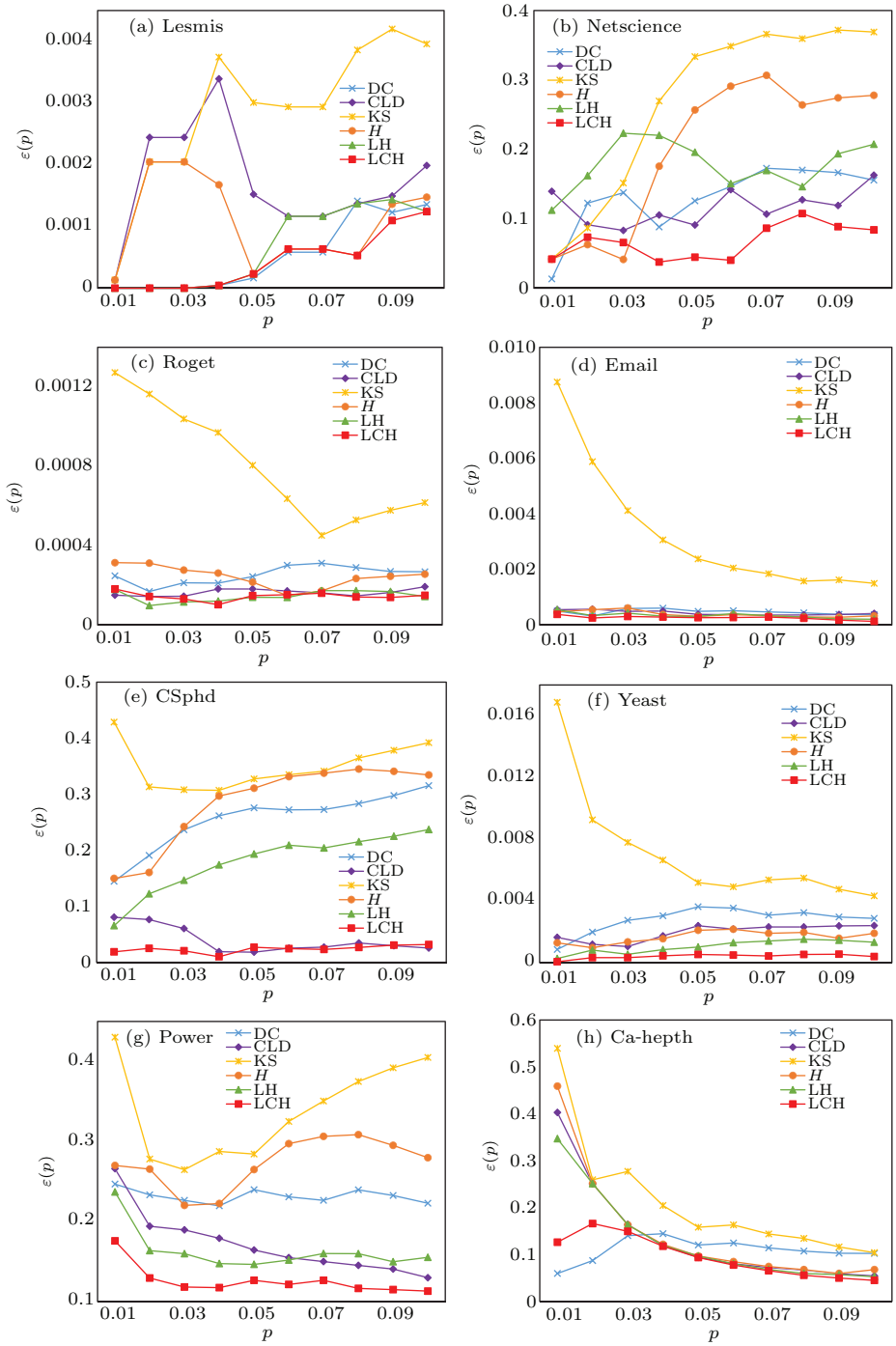


Fig. 4. The imprecision value of six centrality measures in the eight real-world networks. The results are obtained by averaging 100 experiments using the SI model with $\beta = \beta_{\text{th}}$. (a) Lesmis, (b) Netscience, (c) Roget, (d) Email, (e) CSphd, (f) Yeast, (g) Power, (h) Ca-hepth.

5.2. Correlation analysis

To further analyze the correlation between nodes' centrality values and the spreading ability of the SI model, we choose the Lesmis, Netscience and Ca-hepth networks as examples. The experimental results are shown in Fig. 3, where each point stands for a node in the network. In these three networks, there are weak correlations between the centrality values and $F(t)$ in the cases of DC, H and KS. Because many nodes with different spreading abilities are assigned with the same centrality values, which means these three measures cannot well evaluate the influence of nodes. Comparatively speaking, CLD, LH and LCH have strong positive correlations with $F(t)$. However, the points in CLD and LH graphs are relatively discrete compared with LCH, which means that the proposed measure can better detect influential nodes in networks. To summarize, the LCH outperforms other five ranking methods from this perspective.

5.3. Imprecision function

The imprecision function is used to compare the spreading ability of top-rank nodes obtained by different measures, with the SI model. As shown in Fig. 4, one can observe that KS has significantly higher imprecision than the other measures in all studied networks, especially when the p is small. Therefore, the influential nodes identified by KS have poor spreading ability in the actual spreading process. Furthermore, in the Roget, Email, Yeast and Ca-hepth networks as shown in Figs. 4(c), 4(d), 4(f), and 4(h), the imprecisions of DC, CLD, H, LH and LCH are relatively low at most situations. Note that the imprecision values by these four measures are very close, so their performance is relatively close. However, it can be seen that the imprecision of LCH is the lowest. Furthermore, in the Netscience network as shown in Fig. 4(b), the imprecision of LCH is significantly lower than that of the other measures when p is larger than 0.03. At the same time, in the CSphd network as shown in Fig. 4(e), the imprecision of LCH is close to CLD when p is larger than 0.04. On the whole, the imprecision of LCH is still lower. It is concluded that the proposed LCH has the lowest imprecision.

5.4. Top-10 nodes

In some scenarios, researchers are more interested in identifying the top-ranked nodes in a given network. We further analyze the spreading ability of the top-10 ranked nodes, with the SIR model. For the sake of simplicity, this group of experiments is also conducted in three networks, including Lesmis, Netscience and Ca-hepth. Figures 5(a)–5(c) give the comparison between the CLD and the LCH. In the Lesmis and Ca-hepth networks, the $F(t)$ obtained by LCH is always larger than that obtained by CLD throughout the spreading process, which shows the superiority of LCH. In the Netscience network, the two curves overlap completely, indicating that CLD and LCH have similar performance in identifying the top-10 nodes.

The comparison between LH and LCH are shown in Figs. 5(d)–5(f). In the Lesmis network, the initial spreading process for these two measures is almost synchronous. However, LCH speeds up and exceeds LH in the later period, which means that LCH performs better than LH. In the Netscience and Ca-hepth networks, LCH is clearly superior to LH for the spreading speed and the $F(t)$. To summarize, compared with CLD and LH, the measure proposed here can identify the top-rank nodes more effectively in the studied networks.

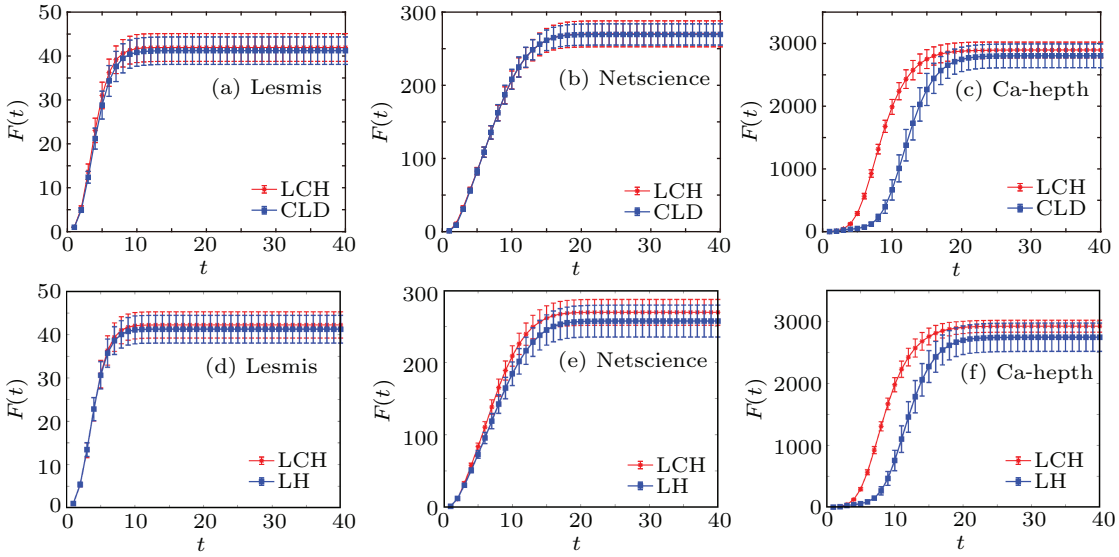


Fig. 5. The average number of $F(t)$ (at time t) of the top-10 nodes in three networks. The first row corresponds to the comparison between CLD and LCH, and the second row is the comparison between LH and LCH. Note that the average number of infected nodes are obtained through 100 independent runs of the SIR model.

5.5. Monotonicity

A good centrality measure should have the ability to discriminate nodes with different spreading abilities. Table 2 shows the monotonicity value of the six centrality measures in eight different networks. According to Eq. (9), the closer the M is to 1, the better discrimination capacity the measure has. Among the six centrality measures, the monotonicity value of LCH is always close to 1 in all studied networks. To sum up, LCH can better discriminate nodes with different spreading abilities.

Table 2. The monotonicity values (M) for the ranking list obtained by six centrality measures on eight real-world networks.

Network	DC	CLD	KS	H	LH	LCH
Lesmis	0.8147	0.9534	0.7441	0.7625	0.9447	**0.9581**
Netscience	0.7642	0.9795	0.6421	0.6825	0.9513	**0.9949**
Roget	0.8672	0.9961	0.5588	0.8055	0.9827	**0.9990**
Email	0.8874	0.9975	0.8088	0.8583	0.9900	**0.9997**
CSphd	0.2432	0.8723	0.0128	0.0757	0.5855	**0.9362**
Yeast	0.7341	0.9871	0.6385	0.6977	0.9693	**0.9960**
Power	0.5927	0.9006	0.2460	0.3930	0.8262	**0.9795**
Ca-hepth	0.7627	0.9830	0.6741	0.7071	0.9552	**0.9922**

6. Conclusions

Identifying and ranking influential nodes are one kind of the most significant and challenging research issues in network science. Global centrality measures are always computationally demanding and not suitable for large-scale networks. With the increase of network scale and increasingly complicated network structure, some node ranking methods based on local information of networks are widely used in various applications due to their low computational demands.

In this work, a local clustering H-index (LCH) centrality measure is presented for identifying and ranking influential nodes in complex networks. The proposed LCH only needs the information of first-order and second-order neighbor nodes of networks, thus it has near linear time complexity and can be applicable to large-scale networks. On the basis of H and its variants, the proposed LCH considers the topological connections among neighbors, the quantity and quality of neighbor nodes simultaneously. In order to test the performance of the LCH, the SIR and SI models are used to simulate the spreading process. According to Kendall's tau correlation coefficient, the imprecision function and monotonicity, we compare the LCH with other five ranking methods, including DC, CLD, KS, H and LH. The experimental results illustrate that the LCH can distinguish nodes more effectively and rank them more accurately. Moreover, another advantage of the proposed measure is its simplicity and low time cost and thus it can be applied to large-scale networks in potential applications. Further studies on identifying influential nodes in weighted networks, multi-level networks and temporal networks are worthy in future works.

References

[1] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[2] Newman M E J 2003 *SIAM Rev.* **45** 167
[3] Albert R, Jeong H and Barabási A L 2000 *Nature* **406** 378
[4] Reis S D, Hu Y Q, Babino A, Andrade J S, Canals S, Sigman M and Makse H A 2014 *Nat. Phys.* **10** 762
[5] Schadt E E 2009 *Nature* **461** 218
[6] Barabási A L, Gulbahce N and Loscalzo J 2010 *Nat. Rev. Genet.* **12** 56
[7] Zhang J P, Guo H M, Jing W J and Jin Z 2019 *Acta. Phys. Sin.* **68** 150501 (in Chinese)
[8] Wu L R, Li J J and Qi J Y 2019 *Acta. Phys. Sin.* **68** 078901 (in Chinese)
[9] Liu F, Wang Z and Deng Y 2020 *Knowl-Based Syst.* **193** 105464
[10] Sun Y, Yao P Y, Wan L J, Shen J and Zhong Y 2017 *Chin. Phys. B* **26** 020201
[11] Yang P L, Xu G Q, Yu Q and Guo J W 2020 *Chaos* **30** 093106
[12] Freeman L C 1978 *Soc. Netw.* **1** 215
[13] Brin S and Page L 1998 *Comput. Netw. ISDN Sys.* **30** 107
[14] Lü LY, Zhang Y C, Yeung C H and Zhou T 2011 *PLoS ONE* **6** e21202
[15] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E and Makse H A 2010 *Nat. Phys.* **6** 888
[16] Bae J and Kim S 2014 *Physica A* **395** 549
[17] Wang Z X, Du C J, Fan J P and Xing Y 2017 *Neurocomputing* **260** 466
[18] Maji G 2020 *J. Comput. Sci.* **39** 101055
[19] Hu J T, Du Y X, Mo H M, Wei D J and Deng Y 2016 *Physica A* **444** 73
[20] Yang Y Z, Yu L, Wang X, Zhou Z L, Chen Y and Kou T 2019 *Physica A* **526** 121118
[21] Yang P L, Liu X and Xu G Q 2018 *Mod. Phys. Lett. B* **32** 1850216
[22] Yan X L, Cui Y P and Ni S J 2020 *Chin. Phys. B* **29** 048902
[23] Yang Y Z, Hu M and Huang T Y 2020 *Chin. Phys. B* **29** 088903
[24] Chen D B, Lü L Y, Shang M S, Zhang Y C and Zhou T 2012 *Physica A* **391** 1777
[25] Chen D B, Gao H, Lü L Y and Zhou T 2013 *PLoS ONE* **8** e77455
[26] Gao S, Ma J, Chen Z M, Wang G H and Xing C M 2014 *Physica A* **403** 130
[27] Li M T, Zhang R S, Hu R J, Yang F, Yao Y B and Yuan Y N 2018 *Int. J Mod. Phys. B* **32** 1850118
[28] Berahmand K, Bouyer A and Samadi N 2018 *Chaos Soliton Fract.* **110** 41
[29] Wang Z X, Sun C C, Yuan G, Rui X B and Yang X D 2020 *J. Comput Sci.* **43** 101129
[30] Tang J X, Zhang R S, Yao Y B, Yang F, Zhao Z L, Hu R J and Yuan Y N 2019 *Physica A* **513** 477
[31] Wen T and Deng Y 2020 *Inform. Sci.* **512** 549
[32] Ruan Y R, Lao S Y, Xiao Y D, Wang J D and Bai L 2016 *Chin. Phys. Lett.* **33** 28901
[33] Hirsch J E 2005 *Proc. Natl. Acad. Sci.* **102** 16569
[34] Lü L Y, Zhou T, Zhang Q M and Stanley H E 2016 *Nat. Commun.* **7** 10168
[35] Liu Q, Zhu Y X, Jia Y, Deng L, Zhou B, Zhu J X and Zou P 2018 *Physica A* **512** 379
[36] Zareie A and Sheikhahmadi A 2019 *Physica A* **514** 141
[37] Christakis N A and Fowler J H 2012 *Stat. Med.* **32** 556
[38] Eguiluz V M and Klemm K 2002 *Phys. Rev. Lett.* **89** 108701
[39] Petermann T and Rios P 2004 *Phys. Rev. E* **69** 066116
[40] Zhou T, Yan G and Wang B H 2005 *Phys. Rev. E* **71** 046141
[41] Wang J Y, Hou X N, Li K Z and Ding Y 2017 *Physica A* **475** 88
[42] Castellano C and Pastor-Satorras R 2010 *Phys. Rev. Lett.* **105** 218701
[43] May R M and Anderson R M 1979 *Nature* **280** 455
[44] Kermack W O and McKendrick A G 1991 *Bull. Math. Biol.* **53** 57
[45] Dorogovtsev S N, Goltsev A V and Mendes J F 2008 *Rev. Mod. Phys.* **80** 1275
[46] Kendall M G 1938 *Biometrika* **30** 81