HHa Centrality Algorithm: A Complex Network Node Ranking Algorithm Based on h-index and Ha-index Postprint

Authors: Liu Jiacheng, Ma Tingcan, Yue Mingliang

Date: 2023-04-01T16:02:57+00:00

Abstract

[Purpose/Significance] This study designs a node centrality algorithm for identifying important nodes in complex networks, with applications in epidemic prevention and control, public opinion monitoring, product marketing, talent discovery, and related domains. [Method/Process] By simultaneously considering both the quantity of a node's high-influence neighbors and their aggregate impact, we propose the HHa node centrality algorithm. Utilizing the SIR epidemic model to simulate information propagation processes on both real-world and synthetic networks, we employ monotonic function M and Kendall's tau coefficient as evaluation metrics to validate the effectiveness, accuracy, and stability of the HHa centrality algorithm. [Results/Conclusion] Experimental results indicate that, compared with seven classical centrality algorithms, the HHa centrality algorithm achieves an M value of 0.999 for its ranking results, ranking second; its Kendall's tau coefficient reaches 0.845, approximately 0.15 higher than other algorithms, ranking first with stable performance. The adoption of the HHa centrality algorithm for identifying important nodes in networks is thus feasible.

Full Text

HHa Centrality Algorithm: A Node Centrality Algorithm Based on the h-index and Ha-index for Complex Networks

Liu Jiacheng[1,2], **Ma Tingcan**[1,2,3], **Yue Mingliang**[1,2,3]

[1]Department of Library, Information and Archives Management, University of Chinese Academy of Sciences, Beijing 100190
[2]Wuhan Library of Chinese Academy of Sciences, Wuhan 430071
[3]Hubei Key Laboratory of Big Data in Science and Technology, Wuhan 430071

Abstract: [Purpose/Significance] This paper designs a node centrality algorithm for identifying important nodes in complex networks, which can play a role in infectious disease prevention and control, public opinion monitoring, product marketing, talent discovery, and other applications. [Method/Process] Considering both the number of a node's high-influence neighbors and their overall influence, we propose the HHa node centrality algorithm. Using the SIR epidemic model to simulate information propagation processes on both real and artificial networks, we employ the monotonic function M and Kendall correlation coefficient as evaluation metrics to verify the effectiveness, accuracy, and stability of the HHa centrality algorithm. [Results/Conclusion] Experiments show that compared with seven classical centrality algorithms, the HHa centrality algorithm achieves an M value of 0.999 for its ranking results (ranking 2nd) and a Kendall coefficient of 0.845 (approximately 0.15 higher than other algorithms, ranking 1st) with stable performance. Using the HHa centrality algorithm to identify important nodes in networks is feasible.

Keywords: complex networks; node centrality; node influence; h-index; HHa centrality algorithm
Classification Codes: TP393.0, TP301.6
DOI: 10.13266/j.issn.0252-3116.2021.20.010

Complex networks are ubiquitous in both the natural world and human society, where individuals and relationships in various systems can be abstracted into networks comprising nodes and edges [1-3], such as social networks and scientific collaboration networks in which humans participate, power grids and transportation networks in the macroscopic world, and protein interaction networks, gene networks, and virus transmission networks in the microscopic world. The development of network science provides a new perspective for understanding the objective world, helping people better comprehend relationship changes [3], information dissemination [4-6], contagion diffusion [7-8], and disease treatment [9].

Regardless of network structure or function, key nodes play crucial roles in information propagation. Therefore, identifying influential nodes in complex networks has attracted increasing attention and finds widespread application across numerous domains. For instance, when an infectious disease is about to outbreak, precise identification of risk nodes directly affects the formulation of immunization policies and the difficulty of subsequent prevention and control efforts [10]; when promoting new products, correct selection of spokespersons and precise marketing strategies can quickly create commercial value [11]; in rumor propagation and public opinion monitoring, opinion leaders' statements can rapidly curb rumors and guide positive public discourse [12]; in scientific collaboration networks, timely identification and support of important talents can promote knowledge flow and increase academic exchange [13].

To this end, this paper applies h-type indices from scientometrics to complex network analysis with appropriate modifications, proposing a new centrality algorithm that considers both the number and overall quality of a node's high-influence neighbors. We aim to achieve a balance between computational complexity and ranking accuracy to more efficiently discover high-influence important nodes in complex networks. We then conduct experiments on real and artificial networks with different structures and functions to verify the effectiveness, accuracy, and stability of this centrality algorithm, with the expectation that it will play a role in infectious disease prevention and control, public opinion monitoring, product marketing, and talent discovery.

2 Related Research

Centrality algorithms for identifying influential nodes in networks can be roughly divided into four categories. The first category is the simplest and most direct, based on neighbor information, such as degree centrality [14]. The degree of node i is defined as the number of nodes directly connected to node i, i.e., the number of its direct neighbors. This centrality is simple, intuitive, and easy to compute, assuming that nodes with more neighbors are more important. However, it ignores environmental information such as the node's position in the network and the quality of its neighbors. k-shell centrality [15] determines node importance based on its position in the network, positing that nodes located at the network core are typically highly important even if they have few neighbors. k-shell centrality has low computational complexity and is suitable for large-scale networks. However, k-shell centrality produces ranking results with low discrimination, as many nodes share the same core number, and it does not perform optimally in certain networks (such as star graphs or BA artificial networks). Moreover, k-shell does not comprehensively consider neighbor information during layer-by-layer decomposition, only considering the remaining neighbor count.

The second category focuses on nodes' ability to control information flow in networks, i.e., the paths they occupy. In transportation or power networks, important nodes often play the role of "bridges." Betweenness centrality [16] starts from the shortest paths between any node pair in the network—the more shortest paths passing through a node, the greater its influence. Closeness centrality [17] considers nodes with smaller average distances to other nodes in the network as more important, meaning that the closer a node is to others, the greater its impact on information flow. Closeness centrality can be understood as using information propagation speed in networks to define node importance, but it can only be used in connected networks. Notably, both betweenness and closeness centrality are global information-based algorithms that must traverse the entire network to obtain all nodes' influence, resulting in high time complexity.

The third category comprises eigenvector-based centrality algorithms, with eigenvector centrality [18] as the typical representative. This method considers not only the number of neighbors but also their importance when judging node importance. However, when the network contains high-degree nodes, eigenvector centrality tends to rank them as important while showing low discrimination for other low-degree nodes.

The fourth category applies relevant metrics from scientometrics to complex networks to measure node influence. For example, A. Korn et al. proposed the lobby-index (h-index centrality) [19] by referencing the h-index from scientometrics. Although computationally simple, its ranking results lack discrimination, as many nodes in the network may share the same h-index. L.Y. Lü et al. [20] revealed the intrinsic relationship between degree, h-index, and coreness, which can be connected through a simple operator h, representing the initial, intermediate, and steady states of a series of actions. Q. Liu et al. [21] used the sum of a node's h-index and all its neighbors' h-indices as an indicator to identify important nodes in networks, but this still does not fully consider neighbors' degree information and quality. A. Zareie et al. [22] designed a cumulative function to calculate node neighbor degree information by referencing the h-index definition, with parameters to sum neighbor degree information for ranking node influence. However, this method involves adjustable parameters, is computationally complex, and shows limited accuracy improvement over classical algorithms. S. Zhao et al. [23] and L. Gao et al. [24] applied the h-index to weighted networks, defining h-degree and weighted h-index to measure node importance in weighted networks, but these methods still fail to overcome the original shortcomings of h-index centrality. P.L. Yang et al. [25] combined shortest distance with the h-index, but calculating shortest distances requires computing distances between a node and all other nodes in the network—similar to closeness centrality mentioned above, it must traverse the entire network, has high computational complexity, and can only identify high-influence node groups without further distinguishing influence magnitudes among them. Y.X. Li et al. [26] proposed a centrality based on h-index and n-order neighbors' h-indices to measure group influence in networks (h-index group centrality), but parameters r, n, and threshold conditions need adjustment for different network structures, resulting in low applicability and high computational load. Lu Pengli et al. [27] proposed a node centrality measure combining degree and h-index, but this method requires parameter debugging and does not consider neighbor quality information. A. Abbasi et al. proposed the a-index in 2013, defined as the average degree of neighbor nodes contributing to the node's h-index [28]. Using this centrality alone yields low accuracy in identifying important nodes and has not been combined with h-index for node centrality ranking.

Through the above analysis, existing node centrality algorithms each have advantages but remain improvable: (1) Current centrality algorithms suffer from being "simple but not accurate enough, or accurate but too complex." Balancing simplicity and accuracy should be the key consideration for new node centrality algorithms, ensuring that as much node information as possible is considered while maintaining low computational complexity. (2) Network structures and functions are complex and diverse; the effectiveness and accuracy of node centrality algorithms across multiple networks should be verified through extensive experiments on real and artificial networks to ensure universality.

This paper's proposed centrality algorithm belongs to the fourth category. Therefore, we next focus on discussing important node mining methods based on the h-index. The HHa centrality algorithm addresses the above issues by migrating h-type indices from scientometrics to complex network analysis for node centrality ranking, considering both the number of high-influence neighbors and their overall influence.

3 HHa Centrality and Its Calculation Method

As shown in Figure 1 [Figure 1: see original paper], let network G(V, E) represent a network with V vertices and E edges. Node v's degree is defined as the number of neighbors directly connected to node v, denoted by d(v). N(v) represents v's neighbors, and I(v) represents nodes in v's neighbors whose degree is greater than or equal to v's h-index. The h-index definition in networks: node v has an h-index of h if it has h neighbors with degree no less than h, and all other neighbors have degree no greater than h, denoted by h(v).

This paper defines node v's Ha-index: among node v's neighbors, sum the degrees of nodes whose degree is greater than or equal to h(v). The resulting number is node v's Ha-index, denoted by Ha(v). The calculation method is shown in formula (1):

$$Ha(v) = \sum_{w \in I(v)} d(w) \quad (1)$$

HHa centrality does not need to traverse the entire network to obtain relevant node information, has low algorithmic time complexity, and is simple to compute. Nodes with higher h-index and Ha-index have greater influence. The algorithm for calculating node HHa centrality uses the h-index as the primary key to sort all nodes in the network in descending order, then uses each node's Ha-index as the secondary key for descending order.

For example, in Figure 1, let HHa(v) represent node v's HHa-index. The HHa-indices of nodes v_1, v_2, v_3 are shown in formulas (2)-(4):

$$HHa(v_1) = (h(v_1), Ha(v_1)) = (3, 3+4+3+3) = (3, 13) \quad (2)$$

$$HHa(v_2) = (h(v_2), Ha(v_2)) = (3, 4+3+4) = (3, 11) \quad (3)$$

$$HHa(v_3) = (h(v_3), Ha(v_3)) = (2, 6+4) = (2, 10) \quad (4)$$

Therefore, $HHa(v_1) > HHa(v_2) > HHa(v_3)$, meaning the influence and importance of nodes v_1, v_2, v_3 decrease in order.

4 Effectiveness and Accuracy Verification of the HHa Centrality Algorithm

4.1 Datasets

To test the performance of the HHa centrality algorithm, this paper uses eight real networks of different scales: Karate [29], Dolphins [30], Jazz [31], USair97 [32], Email [33], Netscience [34], Yeast [35], and Power [36]. Among them, Karate, Dolphins, Yeast, and Power are undirected and unweighted networks; USair97, Jazz, and Netscience are undirected weighted networks; Email is a directed unweighted network. Table 1 shows the number of nodes, edges, average degree, and propagation threshold for each network.

Table 1 Basic information of experimental datasets

Network	Nodes	Edges	Average Degree	Propagation Threshold
Karate				
Dolphins				
USair97				
Email				
Netscience				
Yeast				
Power				

4.2 Evaluation Methods

This paper uses the monotonic function M [37] to quantify the monotonicity of ranking results from different centrality algorithms—that is, whether the algorithm can differentially identify the influence capacity of various nodes. The monotonic function M is defined as shown in formula (5):

$$M(R) = 1 - \frac{\sum_r n_r(n_r - 1)}{n(n-1)} \quad (5)$$

where n is the number of elements in ranking table R, and n_r is the number of elements with the same rank in R. The monotonic function M quantifies the proportion of elements with identical rankings. If ranking table R is completely monotonic, then M(R) = 1; if all nodes in the ranking table share the same rank, then M(R) = 0.

Currently, the Susceptible-Infected-Recovered (SIR) propagation model [38] is widely used to explain and simulate epidemic or information spreading processes [15, 21-22, 24-27, 37, 41]. Therefore, to evaluate the performance of ranking methods, this paper adopts the SIR propagation model to examine the real influence of ranked nodes. In the SIR model, to obtain node v's real influence, node v is set to the infected state in the initial stage while all other nodes are set to susceptible. In each subsequent step, every infected node attempts to infect its susceptible neighbors with infection probability β, and each infected node transitions to the recovered state with probability γ at each step. This process repeats until no infected nodes remain in the network. The number of nodes in the recovered state at the end of the propagation process serves as an indicator to estimate the influence of the initially infected node v. Node v's real influence is defined as the average number of recovered nodes after a sufficiently large number of simulation runs. In this paper, the number of simulation runs is set to 100, meaning the average number of recovered nodes after 100 SIR propagations serves as the node's real influence.

To quantify the accuracy of various centrality algorithms, this paper uses the Kendall correlation coefficient [39] as a ranking correlation measure. Assuming two ranking results X = (x_1, x_2, ..., x) and Y = (y_1, y_2, ..., y), the element pair set formed by corresponding elements is U = {(x_1, y_1), (x_2, y_2), ..., (x , y)}. For any two element pairs (x , y) and (x , y) in this set, if x > x and y > y , or x < x and y < y , then these two elements are concordant; if x > x and y < y , or x < x and y > y , then they are discordant; if x = x or y = y , they are neither concordant nor discordant. The Kendall correlation coefficient is calculated as: T(X, Y) = (n_c - n_d) / (0.5n(n-1)), where n_c and n_d are the numbers of concordant and discordant element pairs respectively, and n is the number of ranked elements in table X or Y.

In this paper, the baseline ranking result is the real influence ranking of nodes obtained through SIR propagation simulation, while the other ranking result is the node centrality ranking obtained through node centrality algorithms. The Kendall coefficient ranges from [-1, 1]; the closer the coefficient is to 1, the closer the two ranking results are, indicating higher accuracy of the centrality algorithm's ranking compared to the real influence ranking from SIR simulation.

During experiments, DC denotes degree centrality, BC denotes betweenness centrality, CC denotes closeness centrality, KS denotes k-shell centrality, EV denotes eigenvector centrality, H denotes h-index centrality, Ha denotes Ha-index centrality, and HHa denotes HHa centrality.

4.3 Monotonicity Test

For the eight real network datasets, various centrality algorithms are used to obtain corresponding node influence ranking results, and the monotonic function M is used to calculate the monotonicity of each ranking result. Table 2 shows the M values for each centrality algorithm's ranking results.

Table 2 Monotonic function M values of different centrality algorithms on eight real network datasets

Network	M(DC)	M(CC)	M(BC)	M(EV)	M(KS)	M(Ha)	M(HHa)
Karate	0.831	0.962	0.899	0.775	0.974	0.708	

Network	M(DC)	M(CC)	M(BC)	M(EV)	M(KS)	M(Ha)	M(HHa)
Dolphins	0.958	0.998					
USair97	0.966	0.859	0.887	0.738	0.734	0.593	
Email	0.989	0.697	0.940	0.090	0.701	0.832	
Netscience	0.988	0.989	0.999	0.909	0.996	1.000	
Yeast	0.999	0.995	1.000	0.917	0.997	1.000	
Power	0.984	0.881	0.940	0.974	0.999	0.994	

As shown in Table 2, since HHa centrality is based on nodes' h-index and Ha-index, its monotonic M value shows significant improvement over h-index centrality. Compared with the other seven centrality algorithms, HHa centrality's monotonic M value ranks at the upper level, second only to eigenvector centrality or betweenness centrality, placing it 2nd or 3rd. This indicates that the proposed HHa centrality can effectively identify and distinguish nodes with different influence levels in networks, assigning different index values to each node, giving our algorithm a competitive advantage.

4.4 Accuracy Test

Next, we compare the ranking results from different centrality algorithms with the node influence ranking obtained from SIR simulation, calculating the Kendall correlation coefficient between them to evaluate the accuracy of different centrality algorithms. A larger Kendall coefficient indicates higher accuracy of the corresponding centrality algorithm's ranking results.

In SIR simulation, the infection probability β is first set near the network's propagation threshold [40] β_{th} = k / k^2 , where k is the network's average degree. If β is much smaller than the propagation threshold, the propagation process cannot proceed normally; if β is much larger than the threshold, the propagation process spreads rapidly and intensely throughout the network, eventually infecting all nodes. Therefore, setting β near the propagation threshold is reasonable and feasible. Second, the recovery probability from infected to recovered state is set as $\gamma = 1/$ k [41]. Compared with the traditional SIR model, this personalized setting of infection and recovery rates better reflects actual network conditions and can more effectively evaluate nodes' real influence.

Table 3 shows the Kendall correlation coefficients between ranking results from various centrality algorithms and the ranking results from the SIR propagation model.

Table 3 Kendall correlation coefficients between different node centrality algorithms and real node influence on eight real network datasets

Network	T(DC)	T(BC)	T(CC)	T(EV)	T(KS)	T(Ha)	T(HHa)
Karate	0.611	0.571	0.507	0.727	0.706	0.863	0.754
Dolphins	0.392	0.647	0.575	0.543	0.801	0.702	0.849
USair97	0.712	0.595	0.569	0.538	0.542	0.773	0.796
Email	0.614	0.744	0.315	0.579	0.664	0.586	0.776
Netscience	0.819	0.624	0.750	0.345	0.572	0.703	0.576
Yeast	0.802	0.790	0.598	0.697	0.711	0.587	0.834
Power	0.829	0.713	0.845	0.542	0.569	0.523	0.354

As shown in Table 3, across the eight real networks, HHa centrality's Kendall correlation coefficients are all greater than those of h-index centrality. In Karate, Dolphins, USair97, Email, and Yeast networks, HHa centrality's Kendall correlation coefficients rank first among the eight centrality algorithms, demonstrating the highest accuracy. In the Jazz network, it ranks second only to eigenvector centrality; in the Power network, it ranks third after eigenvector centrality and Ha-index centrality; in the Netscience network, it ranks fourth.

To test the impact of infection probability β, we set β values from 0.01 to 0.20 in the SIR propagation model. As shown in Figure 2 [Figure 2: see original paper], in Dolphins, USair97, and Yeast networks, as β increases, HHa's Kendall correlation coefficient T values remain significantly higher than the other seven centrality algorithms. In the Email network, HHa centrality's Kendall T values are superior when β is small, though the advantage becomes less obvious as β increases, it remains among the top performers. We can infer that network structural characteristics affect information propagation across the entire network. Therefore, we next use artificial networks to study the impact of network structure.

This paper adopts the LFR artificial network [42] as the experimental network model. The LFR network has numerous parameters characterizing network structure features. By setting different parameter values, we can construct networks with different structures to examine whether these changes affect HHa's applicability. The initial default parameters are set as: number of nodes = 1000, average degree = 5, maximum degree = 50, node degree follows a power-law distribution with exponent $\gamma = 2$, and community mixing parameter = 0.2. We evaluate the impact of changes in node count, average degree, degree distribution exponent, and community mixing parameter.

First, we set network node counts to N = 500, 1000, 1500. As shown in Figure 3 [Figure 3: see original paper], when infection rate is small, closeness centrality, eigenvector centrality, and Ha-index centrality are effective methods for measuring node importance. However, as infection rate increases, the performance of closeness centrality, eigenvector centrality, and Ha-index centrality declines rapidly, while HHa centrality shows significant advantages over other centrality algorithms.

Second, we set network average degrees to k = 5, 10, 15. As shown in Figure 4 [Figure 4: see original paper], notably, as network average degree increases, HHa centrality gradually approaches degree centrality. When average degree k = 15, the performance of all eight centrality algorithms tends to overlap, and their accuracy all declines significantly as infection rate increases. This may be because when network average degree and infection probability are large, selecting any node as the initial infected state can cause information to spread rapidly and extensively, weakening node importance under such conditions. However, HHa centrality can still achieve optimal performance when infection rate is small.

Third, we set node degree distribution exponents to t = 2, 2.5, 3. As shown in Figure 5 [Figure 5: see original paper], regardless of degree distribution exponent values, HHa centrality algorithm performs relatively stably. Although it may only rank at upper-middle level when infection rate is small, as infection rate increases, HHa centrality algorithm's performance always surpasses other centrality algorithms, ranking first.

Finally, we set community mixing parameters to = 0.2, 0.5, 0.8, as shown in Figure 6 [Figure 6: see original paper]. Community structure is a crucial network characteristic. When the community mixing parameter is low, the LFR benchmark creates random networks with clear community structures. Experimental results show that regardless of community structure clarity, HHa centrality algorithm can always significantly identify influential nodes in networks, with performance far exceeding the other seven centrality algorithms.

Overall, on LFR networks with varying network scales or parameters, although network topological properties change, HHa centrality algorithm can always more accurately identify important nodes in networks. When infection rate is small, propagation is difficult to diffuse and remains limited to local neighbor nodes. In this case, nodes with more neighbors can typically influence more nodes, so degree centrality performs well when infection rate is low. As infection rate increases, the probability of infecting nodes beyond neighbors rises, and the limitations of degree centrality and other algorithms gradually become apparent, making HHa centrality algorithm's advantages more pronounced. Compared with other centrality algorithms, HHa centrality algorithm demonstrates higher stability, i.e., lower sensitivity to parameter changes. On LFR networks with different structural features, we verify the effectiveness and accuracy of HHa centrality algorithm.

5 Conclusion and Discussion

This paper proposes an effective ranking method—the HHa centrality algorithm—which uses the h-index and Ha-index to measure node influence in complex networks. Existing algorithms suffer from the drawback of being "simple but not accurate, or accurate but too complex." The HHa centrality algorithm migrates metrics from scientometrics to complex network analysis, simultaneously considering both the number of a node's high-influence neighbors and their overall influence when ranking node centrality, achieving a good balance between algorithmic simplicity and accuracy. Experimental results demonstrate that HHa centrality is a simple yet powerful method for evaluating network information propagation capability and identifying important nodes.

First, we use the monotonic function M to measure whether centrality algorithms can distinguish node importance. Experiments on eight real networks show that HHa centrality produces more monotonic rankings than h-index centrality and Ha-index centrality, ranking at the upper-middle level compared with the other seven metrics. Additionally, to measure the accuracy of node influence rankings obtained by HHa centrality, we use SIR epidemic model propagation results as nodes' real influence and employ Kendall correlation coefficient to measure consistency between rankings from centrality algorithms and SIR simulation. Compared with other centrality algorithms, HHa centrality demonstrates the highest accuracy. Based on the LFR benchmark, experiments on artificial networks with different topological properties show that HHa centrality's accuracy and stability are significantly superior to other centrality methods.

Our experimental network set includes directed/undirected and weighted/unweighted networks, on which HHa centrality achieves excellent results, indicating its broad applicability. However, since HHa centrality is based on h-index centrality and Ha-index centrality—neither of which considers edge direction or weight—HHa centrality also does not incorporate edge direction or weight. Future work could improve HHa centrality based on these two aspects to consider more network information, enhancing its application specificity and selectivity while ensuring universality.

Notably, node importance manifests through network structure and function, so evaluating centrality algorithms' effectiveness and accuracy must consider specific network structures and functions. Although this paper conducts relatively complete tests of HHa centrality algorithm's applicability and generality on LFR networks, the selected real network datasets remain insufficient and do not cover all functional types. Additionally, the SIR model is only a highly simulated and abstracted representation of the real world. While it has certain guiding significance, results may differ from real situations due to individual heterogeneity and high uncertainty in real networks.

Finally, after conducting small-scale real experiments, the proposed HHa centrality algorithm can be applied more broadly in practical scenarios to assist in efficiently discovering important nodes, such as identifying distinguished scholars and leading talents, retrieving important domain papers, finding high-influence Weibo users, identifying key disease-causing genes, extracting article keywords, controlling infectious disease spread, and searching for criminals and terrorists.

References

[1] Barabási A L. Network Science[M]. Cambridge: Cambridge University

Press, 2016.
[2] Strogatz S H. Exploring complex networks[J]. Nature, 2001, 410(6825): 268-276.
[3] Newman M. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167-256.
[4] Borge-Holthoefer J, Moreno Y. Absence of influential spreaders in rumor dynamics[J]. Physical Review E, 2012, 85(2): 026116.
[5] Hosni A, Li K, Ahmad S. Minimizing rumor influence in multi-sign online social networks based on human individual and social behaviors[J]. Information Sciences, 2020, 512: 1458-1480.
[6] Ahmed W, Vidal-Alaball J, Downing J, et al. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data[J]. Journal of Medical Internet Research, 2020, 22(5): e19458.
[7] Colizza V, Barrat A, Barthélemy M, et al. The modeling of global epidemics: stochastic dynamics and predictability[J]. Bulletin of Mathematical Biology, 2006, 68(8): 1893-1921.
[8] Wang W, Tang M, Eugene S H, et al. Unification of theoretical approaches for epidemic spreading on complex networks[J]. Reports on Progress in Physics, 2017, 80(3): 036603.
[9] Liu X R, Hong Z Y, Liu J, et al. Computational methods for identifying critical nodes in biological networks[J]. Briefings in Bioinformatics, 2020, 21(2): 486-497.
[10] Liu Z J, Zhao S, Chen J, et al. A new immunization strategy based on node importance[J]. Journal of Nanjing University (Natural Sciences), 2017, 53(2): 350-356.
[11] Lü L Y, Zhang Y C, Yeung C H, et al. Leaders in social networks, the delicious case[J]. PLoS ONE, 2011, 6(6): e21202.
[12] Chen F, Fu X, He Y, et al. Research on microblog opinion leader discovery integrating social network analysis and influence diffusion model[J]. Data Analysis and Knowledge Discovery, 2018, 2(12): 60-67.
[13] You X M, Ma Y H, Liu Z Y, et al. Representation method of cooperative social network features based on node2vec model[J]. Computer Communications, 2021, 173: 21-26.
[14] Freeman L C. Centrality in social networks: conceptual clarification[J]. Social Networks, 1979, 1(3): 215-239.
[15] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6(11): 888-893.
[16] Freeman L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40(1): 35-41.
[17] Sabidussi G. The centrality index of a graph[J]. Psychometrika, 1966, 31(4): 581-603.
[18] Bonacich P. Factoring and weighting approaches to status scores and clique identification[J]. Journal of Mathematical Sociology, 1972, 2(1): 113-120.
[19] Korn A, Schubert A, Telcs A. Lobby index in networks[J]. Physica A: Statistical Mechanics and its Applications, 2009, 388(11): 2221-2226.
[20] Lü L Y, Zhou T, Zhang Q M, et al. The h-index of a network node and its

relation to degree and coreness[J]. Nature Communications, 2016, 7: 10168.
[21] Liu Q, Zhu Y X, Jia Y, et al. Leveraging local h-index to identify and rank influential spreaders in networks[J]. Physica A: Statistical Mechanics and its Applications, 2018, 512: 379-391.
[22] Zareie A, Sheikhahmadi A. EHC: Extended h-index centrality measure for identifying influential spreaders in complex networks[J]. Physica A: Statistical Mechanics and its Applications, 2019, 514: 141-155.
[23] Zhao S, Rousseau R, Ye F Y. H-degree as a basic measure in weighted networks[J]. Journal of Informetrics, 2011, 5(4): 668-677.
[24] Gao L, Yu S, Li M, et al. Weighted h-index for identifying influential spreaders[J]. Symmetry, 2019, 11(10): 1263.
[25] Yang P L, Liu X, Xu G Q. An extended clustering method using h-index and minimum distance for searching multiple key spreaders[J]. International Journal of Modern Physics C, 2019, 30(7): 1940008.
[26] Li Y X, Sheng Y Q, Ye X Z. Group centrality algorithms based on the h-index for identifying influential nodes in large-scale networks[J]. International Journal of Innovative Computing, Information and Control, 2020, 16(4): 1183-1201.
[27] Lu P L, Yu Z. Node ranking method in complex networks based on degree and H-index extension[J]. Journal of Lanzhou University of Technology, 2020, 46(5): 100-106.
[28] Abbasi A. H-type hybrid centrality measures for weighted networks[J]. Scientometrics, 2013, 96(2): 633-640.
[29] Zachary W W. An information flow model for conflict and fission in small groups[J]. Journal of Anthropological Research, 1977, 33(4): 452-473.
[30] Lusseau D, Schneider K, Boisseau O J, et al. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations[J]. Behavioral Ecology & Sociobiology, 2003, 54(4): 396-405.
[31] Gleiser P M, Danon L. Community structure in jazz[J]. Advances in Complex Systems, 2003, 6(4): 565-573.
[32] Batagelj V, Mrvar A. Pajek datasets[DB/OL]. [2021-04-07]. http://vlado.fmf.uni-lj.si/pub/networks/data.
[33] Guimerà R, Danon L, Díaz-Guilera A, et al. Self-similar community structure in a network of human interactions[J]. Physical Review E, 2003, 68(6): 065103.
[34] Newman M. Finding community structure in networks using the eigenvectors of matrices[J]. Physical Review E, 2006, 74(3): 036104.
[35] Bu D, Zhao Y, Cai L, et al. Topological structure analysis of the protein-protein interaction network in budding yeast[J]. Nucleic Acids Research, 2003(9): 2443-2450.
[36] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. Nature, 1998, 393(6684): 440-442.
[37] Bae J, Kim S, et al. Identifying and ranking influential spreaders in complex networks by neighborhood coreness[J]. Physica A: Statistical Mechanics and its Applications, 2014, 395(4): 549-559.
[38] Kermack W O, McKendrick A G. A contribution to the mathematical

theory of epidemics[J]. Proceedings of the Royal Society A, 1927, 115(772): 700-721.
[39] Kendall M. A new measure of rank correlation[J]. Biometrika, 1938, 30(1/2): 81-93.
[40] Castellano C, Pastor-Satorras R. Thresholds for epidemic spreading in networks[J]. Physical Review Letters, 2010, 105(21): 218701.
[41] Zou Q, Zhang Y Y, Chen Y F, et al. A fast and accurate node influence ranking algorithm in social networks[J]. Computer Engineering & Science, 2014, 36(12): 2346-2354.
[42] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms[J]. Physical Review E, 2008, 78(4): 046110.

Author Contributions:
Liu Jiacheng: Experimental design, data processing, and paper drafting;
Ma Tingcan: Experimental design and paper revision;
Yue Mingliang: Paper revision.

Note: Figure translations are in progress. See original paper for figures.