



OPEN

## Identifying influential spreaders by gravity model considering multi-characteristics of nodes

Zhe Li<sup>1</sup>✉ & Xinyu Huang<sup>2</sup>✉

How to identify influential spreaders in complex networks is a topic of general interest in the field of network science. Therefore, it wins an increasing attention and many influential spreaders identification methods have been proposed so far. A significant number of experiments indicate that depending on a single characteristic of nodes to reliably identify influential spreaders is inadequate. As a result, a series of methods integrating multi-characteristics of nodes have been proposed. In this paper, we propose a gravity model that effectively integrates multi-characteristics of nodes. The number of neighbors, the influence of neighbors, the location of nodes, and the path information between nodes are all taken into consideration in our model. Compared with well-known state-of-the-art methods, empirical analyses of the Susceptible-Infected-Recovered (SIR) spreading dynamics on ten real networks suggest that our model generally performs best. Furthermore, the empirical results suggest that even if our model only considers the second-order neighborhood of nodes, it still performs very competitively.

The focus of network science has been shifting from discovering macroscopic statistical regularities to microscopic elements, vital nodes identification has received a huge amount of attention from researchers of network science in recent years. Vital nodes identification can be widely used in disease analysis<sup>1,2</sup>, rumor analysis<sup>3</sup>, information propagation<sup>4</sup>, power grid protection<sup>5</sup>, discovery of candidate drug targets and essential proteins<sup>6</sup>, discovery of important species<sup>7,8</sup>, and so on.

So far, most known methods only use structural information<sup>9</sup>, which can be classified into neighborhood-based centralities and path-based centralities roughly. Typical representatives of neighborhood-based centralities are degree centrality<sup>10</sup> (DC), H-index<sup>11</sup> and *k*-shell decomposition method<sup>12</sup> (KS). For DC, the more neighbors a node has, the greater its influence. For H-index, the more large-degree neighbors a node has, the greater its influence. For KS, the more central a node locates in the network, the greater its influence. Besides, eigenvector centrality<sup>13</sup> (EC) is the representative neighborhood-based iterative method, suggesting that the influence of a node is not only determined by the number of its neighbors, but also determined by the influence of each neighbor. Typical representatives of path-based centralities are betweenness centrality<sup>14</sup> (BC) and closeness centrality<sup>15</sup> (CC). For BC, the more a node is located in shortest paths, the greater its influence. For CC, the closer a node is to other nodes, the greater its influence.

However, a significant number of experiments indicate that depending on a single characteristic of nodes to reliably identify influential spreaders is inadequate<sup>9</sup>. As a result, the methods integrating multi-characteristics of nodes have been proposed. In particular, the methods based on gravity law seem very promising. As several laws behind phenomena in life are similar to the gravity law, the gravity model, which derives from the gravity law, is also favored and exhibited in many real-life scenarios. Representative examples include predicting the population migration between regions in demography<sup>16</sup> and forecasting the trade flows throughout countries in economics<sup>17</sup>. In network science, the gravity model is utilized to evaluate the influence<sup>18–20</sup> of nodes, and so on. Recently, a series of gravity-law-based algorithms<sup>18–30</sup> considering both neighborhood information and path information have been proposed, and their performance is much better than the above well-known state-of-the-art methods. Typical representatives are gravity centrality<sup>18</sup> (GC), improved gravity centrality<sup>19</sup> (IGC) and local gravity model<sup>20</sup> (LGM). For GC, the *k*-shell value of a node is regarded as its mass. For IGC, the focal node uses the *k*-shell value as its mass while its neighbors view the degree value as their masses. For LGM, the degree value of a node is regarded as its mass. However, whether the degree or *k*-shell is regarded as mass, the influence of neighbors is not taken into consideration. In view of this, we propose a gravity model that effectively

<sup>1</sup>Software College, Shenyang University of Technology of China, Shenyang 110870, People's Republic of China. <sup>2</sup>Software College, Northeastern University of China, Shenyang 110819, People's Republic of China. ✉email: gislzneu@163.com; huangxinyu@mail.neu.edu.cn

integrates multi-characteristics of nodes to measure the influence of nodes in spreading dynamics. In our model, the number of neighbors, the influence of neighbors, the location of nodes, and the path information between nodes are all taken into consideration.

## Preliminaries

**Well-known state-of-the-art methods.** Denote  $G = \langle V, E \rangle$  an undirected and unweighted simple network, where  $V$  and  $E$  are the sets of nodes and links. Denote  $|V| = N$  and  $|E| = M$ , then the network has  $N$  nodes and  $M$  links. The adjacent matrix of  $G$  is denoted by  $A = (a_{ij})_{N \times N}$ , if node  $i$  links to node  $j$ ,  $a_{ij} = 1$ , otherwise,  $a_{ij} = 0$ .

The degree centrality<sup>10</sup> (DC) of node  $i$  is measured by

$$DC(i) = k(i), \quad (1)$$

where  $k(i) = \sum_{j=1}^N a_{ij}$ .

The H-index<sup>11</sup> of node  $i$ , denoted by  $H(i)$ , is defined as the maximal integer satisfying that there are at least  $H(i)$  neighbors of node  $i$  whose degrees are all greater than or equal to  $H(i)$ .

The  $k$ -shell decomposition method<sup>12</sup> (KS) works by iterative decomposition of the network into different shells. The first step of KS is to remove the nodes whose degrees are equal to 1 from the network, which will cause a reduction of the degree value to the remaining nodes. Continually remove all the nodes whose residual degrees are less than or equal to 1, until all the remaining nodes' residual degrees are greater than 1. All the removed nodes in the first step form the 1-shell and their  $k$ -shell values are all equal to 1. Repeat this process to obtain 2-shell, 3-shell, . . . , and so on. The decomposition process will continue until there are no more nodes in the network.

The eigenvector centrality<sup>13</sup> (EC) of node  $i$  is measured by

$$x(i) = c \sum_{j=1}^N a_{ij}x(j), \quad (2)$$

where  $c$  is a constant, generally speaking,  $c$  is set to the reciprocal of the largest eigenvalue of  $A$ .

The betweenness centrality<sup>14</sup> (BC) of node  $i$  is measured by

$$BC(i) = \sum_{s \neq i, s \neq t, i \neq t} \frac{g_{st}(i)}{g_{st}}, \quad (3)$$

where  $g_{st}$  is the number of shortest paths between node  $s$  and node  $t$ , and  $g_{st}(i)$  is the number of shortest paths via node  $i$  between node  $s$  and node  $t$ .

The closeness centrality<sup>15</sup> (CC) of node  $i$  is measured by

$$CC(i) = \frac{N-1}{\sum_{j \neq i} d(i, j)}, \quad (4)$$

where  $d(i, j)$  is the shortest distance from node  $i$  to node  $j$ .

The gravity centrality<sup>18</sup> (GC) of node  $i$  is measured by

$$GC(i) = \sum_{j \in \psi_i} \frac{k_s(i)k_s(j)}{d^2(i, j)}, \quad (5)$$

where  $k_s(i)$  is the  $k$ -shell value of node  $i$ , and  $\psi_i$  is the neighborhood set whose distance to node  $i$  is not greater than 3.

An extended version of GC, denoted by  $GC+$ ,  $GC+$  of node  $i$  is measured by

$$GC+(i) = \sum_{j \in \Lambda_i} GC(j), \quad (6)$$

where  $\Lambda_i$  is the neighborhood set whose distance to node  $i$  equals to 1.

The improved gravity centrality<sup>19</sup> (IGC) of node  $i$  is measured by

$$IGC(i) = \sum_{j \in \psi_i} \frac{k_s(i)k(j)}{d^2(i, j)}. \quad (7)$$

An extended version of IGC, denoted by  $IGC+$ ,  $IGC+$  of node  $i$  is measured by

$$IGC+(i) = \sum_{j \in \Lambda_i} IGC(j). \quad (8)$$

The local gravity model<sup>20</sup> (LGM) of node  $i$  is measured by

$$LGM(i) = \sum_{d(i,j) \leq R, j \neq i} \frac{k(i)k(j)}{d^2(i,j)}, \quad (9)$$

where  $R$  is the truncation radius, and the optimal truncation radius  $R^*$  can be estimated by

$$R^* \approx \frac{1}{2} \langle d \rangle, \quad (10)$$

where  $\langle d \rangle$  is the average distance of the network.

**The SIR model.** The SIR model<sup>31</sup> initially considers all the nodes as in the susceptible (S) state except the source node in the infected (I) state. At each time step, each infected node can infect its susceptible neighbors with probability  $\beta$ . Then, each infected node enters the recovered (R) state with probability  $\lambda$ . The propagation process continues until there are no more nodes in the infected state. The influence of node  $i$  can be estimated by

$$F(i) = N_r/N, \quad (11)$$

where  $N_r$  is the number of recovered nodes when dynamic process achieves steady state. For simplicity,  $\lambda$  is set to 1, then the corresponding epidemic threshold<sup>32</sup> can be calculated by

$$\beta_c \approx \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}, \quad (12)$$

where  $\langle k \rangle$  is the average degree, and  $\langle k^2 \rangle$  is the second-order moment of the degree distribution.

**The Kendall's Tau.** The Kendall's Tau<sup>33</sup> is an index describing the strength of correlation between two sequences. Denote  $X = (x_1, x_2, \dots, x_N)$  and  $Y = (y_1, y_2, \dots, y_N)$  are two sequences with  $N$  elements. For any pair of two-tuples  $(x_i, y_i)$  and  $(x_j, y_j)$  ( $i \neq j$ ), if both  $x_i > x_j$  and  $y_i > y_j$  or both  $x_i < x_j$  and  $y_i < y_j$ , the pair is concordant. If both  $x_i > x_j$  and  $y_i < y_j$  or both  $x_i < x_j$  and  $y_i > y_j$ , the pair is discordant. If  $x_i = x_j$  or  $y_i = y_j$ , the pair is neither concordant nor discordant. The Kendall's Tau of  $X$  and  $Y$  can be calculated by

$$\tau = \frac{2(n_+ - n_-)}{N(N-1)}, \quad (13)$$

where  $n_+$  is the number of concordant pairs, and  $n_-$  is the number of discordant pairs.

**The monotonicity.** The monotonicity<sup>34</sup>  $M$  of ranking list  $L$  is used to quantitatively measure the resolution of different indices, and it can be calculated by

$$M(L) = \left[ 1 - \frac{\sum_{r \in L} U_r(U_r - 1)}{U(U-1)} \right]^2, \quad (14)$$

where  $U$  is the size of  $L$ , and  $U_r$  is the number of ties with the same rank  $r$ .

## Results

**Algorithms.** According to previous studies, the degree value of a node indicates the number of its neighbors, the  $k$ -shell value of a node reflects where it locates in the network, the eigenvector centrality value of a node can reflect both the number of its neighbors and the influence of each neighbor, and the distance between two nodes can describe the path information. Individually speaking, nodes with large degree value,  $k$ -shell value and eigenvector centrality value are likely to be more influential. Furthermore, a node is of higher impacts on nearby nodes. According to the above issues and inspired by the gravity law, we regard the sum of degree value,  $k$ -shell value and eigenvector centrality value of a node as its mass, and the shortest distance between two nodes as their distance. Therefore, the influence of node  $i$  can be estimated as

$$MCGM(i) = \sum_{d(i,j) \leq R, j \neq i} \frac{(k(i) + k_s(i) + x(i))(k(j) + k_s(j) + x(j))}{d^2(i,j)}. \quad (15)$$

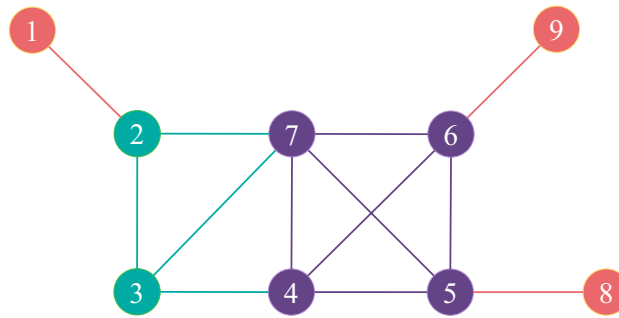
Such method is named as multi-characteristics gravity model (MCGM) as it considers multi-characteristics of nodes and adopts the gravity law.

It is not difficult to find that these three indices (DC, KS, EC) are not in the same order of magnitude, so normalization is required. As a result, Eq. (15) can be rewritten as

$$MCGM(i) = \sum_{d(i,j) \leq R, j \neq i} \frac{\left( \frac{k(i)}{k_{max}} + \frac{k_s(i)}{k_{smax}} + \frac{x(i)}{x_{max}} \right) \left( \frac{k(j)}{k_{max}} + \frac{k_s(j)}{k_{smax}} + \frac{x(j)}{x_{max}} \right)}{d^2(i,j)}, \quad (16)$$

where  $k_{max}$ ,  $k_{smax}$  and  $x_{max}$  denote the maximum of degree value,  $k$ -shell value and eigenvector centrality value, respectively.

However, since the  $k$ -shell index has smaller value space, the normalized  $k$ -shell index is still larger than the other two indices. Therefore, it is necessary to lower the impact of the  $k$ -shell index. Given an index, due to the



**Figure 1.** A toy network. The red nodes are in 1-shell, the green nodes are in 2-shell and the purple nodes are in 3-shell.

scale-free property of networks, the index values of most nodes are relatively small. Therefore, the index with larger value space generally has a smaller ratio between the median and the maximum. In our model, it is obvious that the value space of degree centrality and eigenvector centrality is larger than that of  $k$ -shell index. In view of this, we can lower the impact of  $k$ -shell index by

$$\alpha = \frac{\max\left\{\frac{k_{mid}}{k_{max}}, \frac{x_{mid}}{x_{max}}\right\}}{\frac{k_{smid}}{k_{smax}}}, \tag{17}$$

where  $k_{mid}$ ,  $k_{smid}$  and  $x_{mid}$  denote the median of degree value,  $k$ -shell value and eigenvector centrality value, respectively. The purpose of taking the maximum value of  $\left\{\frac{k_{mid}}{k_{max}}, \frac{x_{mid}}{x_{max}}\right\}$  is to prevent the function of  $k$ -shell index from being excessively weakened.

Finally, Eq. (15) can be rewritten as

$$MCGM(i) = \sum_{d(i,j) \leq R, j \neq i} \frac{\left(\frac{k(i)}{k_{max}} + \frac{\alpha k_s(i)}{k_{smax}} + \frac{x(i)}{x_{max}}\right) \left(\frac{k(j)}{k_{max}} + \frac{\alpha k_s(j)}{k_{smax}} + \frac{x(j)}{x_{max}}\right)}{d^2(i, j)}. \tag{18}$$

The Algorithmic description of MCGM is provided in Algorithm 1. We take a toy network shown in Fig. 1 to illustrate the calculation process of Algorithm 1.

---

**Algorithm 1:** Multi-characteristics gravity model

---

**Input:** graph:  $G < V, E >$ , truncation radius:  $R$ , number of nodes:  $N$   
**Output:**  $Rank[v, MCGM(v)]$   
**for**  $i \leftarrow 1$  **to**  $N$  **do**  
    | Compute  $k(v_i)$  of node  $v_i$ ;  
**end**  
**for**  $i \leftarrow 1$  **to**  $N$  **do**  
    | Compute  $k_s(v_i)$  of node  $v_i$ ;  
**end**  
**for**  $i \leftarrow 1$  **to**  $N$  **do**  
    | Compute  $x(v_i)$  of node  $v_i$ ;  
**end**  
Compute  $k_{max}, k_{smax}, x_{max}, k_{mid}, k_{smid}, x_{mid}$ ;  
Compute  $\alpha$  using Eq. 17;  
**for**  $i \leftarrow 1$  **to**  $N$  **do**  
    | Find all neighbors of node  $v_i$  within the truncation radius  $R$ ;  
    | Compute  $MCGM(v_i)$  of node  $v_i$  using Eq. 18;  
**end**  
return  $Rank[v, MCGM(v)]$ ;

---

Firstly, calculate the degree value,  $k$ -shell value and eigenvector centrality value of each node in the toy network, the results are shown in Table 1.

Secondly, calculate  $k_{max} = 5, k_{smax} = 3, x_{max} = 0.1917, k_{mid} = 3, k_{smid} = 2$  and  $x_{mid} = 0.1256$ , furthermore, calculate  $\alpha = 0.9827$ .

Finally, the result of MCGM with  $R = 2$  of the toy network is shown in Table 2. Take node 3 as an example, the 1-order neighbors of node 3 are node 2, node 4 and node 7, the 2-order neighbors of node 3 are node 1, node 5 and node 6, so  $MCGM(3) = 16.9320$ .

**Data description.** In this paper, we apply ten real networks from six fields to test the performance of MCGM, including one transportation network (USAir<sup>35</sup>), one communication network (Email<sup>36</sup>), one infra-

Node	DC	KS	EC
1	1	1	0.0259
2	3	2	0.0943
3	3	2	0.1256
4	4	3	0.1714
5	4	3	0.1534
6	4	3	0.1534
7	5	3	0.1917
8	1	1	0.0421
9	1	1	0.0421

**Table 1.** The degree value,  $k$ -shell value and eigenvector centrality value of each node in the toy network.

Node	1-order neighbors	2-order neighbors	MCGM
1	2	3,7	1.9679
2	1,3,7	4,5,6	13.1293
3	2,4,7	1,5,6	16.9320
4	3,5,6,7	2,8,9	29.0955
5	4,6,7,8	2,3,9	26.0652
6	4,5,7,9	2,3,8	26.0652
7	2,3,4,5,6	1,8,9	35.9099
8	5	4,6,7	3.4704
9	6	4,5,7	3.4704

**Table 2.** The result of MCGM with  $R = 2$  of the toy network.

Networks	$N$	$M$	$\langle k \rangle$	$\langle d \rangle$	$C$	$r$	$H$	$\beta_c$
USAir	332	2126	12.8072	2.7381	0.7494	-0.2079	3.4639	0.0231
Email	1133	5451	9.6222	3.6060	0.2540	0.0782	1.9421	0.0565
Power	4941	6594	2.6691	18.9892	0.1065	0.0035	1.4504	0.3483
Router	5022	6258	2.4922	6.4488	0.0329	-0.1384	5.5031	0.0786
Jazz	198	2742	27.6970	2.2350	0.6334	0.0202	1.3951	0.0266
NS	379	914	4.8232	6.0419	0.7981	-0.0817	1.6630	0.1424
PB	1222	16714	27.3552	2.7375	0.3600	-0.2213	2.9707	0.0125
Facebook	4039	88234	43.6910	3.6925	0.6170	0.0636	2.4392	0.0095
WV	7066	100736	28.5129	3.2475	0.2090	-0.0833	5.0992	0.0069
Sex	15810	38540	4.8754	5.7846	0.0000	-0.1145	5.8276	0.0365

**Table 3.** The topological features of ten real networks.

structure network (Power<sup>37</sup>), one technological network (Router<sup>38</sup>), two collaboration networks (Jazz<sup>39</sup> and NS<sup>40</sup>) and four social networks (PB<sup>41</sup>, Facebook<sup>42</sup>, WV<sup>43</sup> and Sex<sup>44</sup>). Table 3 shows these networks' topological features, including the number of nodes, the number of links, the average degree, the average distance, the clustering coefficient<sup>37</sup>, denoted by  $C$ , the assortative coefficient<sup>45</sup>, denoted by  $r$ , the degree heterogeneity<sup>46</sup>, denoted by  $H$ , and the epidemic threshold<sup>32</sup> of SIR model<sup>31</sup>.

**Empirical results.** Based on the above real networks, the well-known SIR model<sup>31</sup> is used to compare the influential rankings produced by algorithms and simulations. Given the network and the transmission probability  $\beta$ , in order to guarantee the reliability of the results, 1000 independent realizations are executed and averaged to obtain the standard ranking of the influence of nodes (see details about SIR model in Preliminaries). In each realization, every node is selected once as the seed once. We apply the Kendall's Tau ( $\tau$ ) between the standard ranking and the ranking produced by the algorithm to measure the accuracy of an algorithm. Since  $\tau \in [-1, 1]$ , the closer the  $\tau$  is to 1, the better the performance of the algorithm. The benchmark algorithms include degree centrality<sup>10</sup> (DC), H-index<sup>11</sup>,  $k$ -shell decomposition method<sup>12</sup> (KS), eigenvector centrality<sup>13</sup> (EC), betweenness centrality<sup>14</sup> (BC), closeness centrality<sup>15</sup> (CC), DynamicRank<sup>47</sup> (DR), the extended version of gravity centrality<sup>18</sup>

Networks	DC	H-index	KS	EC	BC	CC	DR	GC+	IGC+	LGM	MCGM
USAir	0.7370	0.7568	0.7529	0.8946	0.5171	0.8027	0.9096	0.8985	0.9006	0.8875	<b>0.9145</b>
Email	0.7653	0.7883	0.7702	0.8832	0.6243	0.8163	0.8991	0.9119	<b>0.9133</b>	0.8697	0.9091
Power	0.4264	0.4009	0.3122	0.2818	0.3254	0.3838	0.7570	0.7906	<b>0.8387</b>	0.7442	0.7639
Router	0.3139	0.1928	0.1810	0.5924	0.3096	0.6383	0.8215	0.7896	0.7823	0.7894	<b>0.8324</b>
Jazz	0.8150	0.8513	0.7638	0.8854	0.4641	0.7008	0.8761	0.9158	0.9244	0.8666	<b>0.9333</b>
NS	0.5790	0.5610	0.5106	0.3660	0.3003	0.3397	0.7377	0.8511	0.8722	0.8372	<b>0.8736</b>
PB	0.8524	0.8694	0.8595	0.8738	0.6771	0.7852	0.9060	<b>0.9189</b>	0.9176	0.9030	0.9184
Facebook	0.6798	0.7066	0.7075	0.6226	0.4529	0.3940	0.7865	0.8414	0.8372	0.8275	<b>0.8639</b>
WV	0.7619	0.7662	0.7657	0.8334	0.6978	0.8127	0.8360	0.8298	0.8305	0.8276	<b>0.8379</b>
Sex	0.4664	0.4855	0.4925	0.7404	0.4118	0.7677	0.8139	0.8038	0.8076	0.7789	<b>0.8448</b>

**Table 4.** The algorithms' accuracies of MCGM and the benchmark algorithms measured by Kendall's Tau for  $\beta = \beta_c$ . The parameters in the related algorithms (i.e., LGM and MCGM) are adjusted to their optimal values subject to the largest  $\tau$ , that is, we need to search the optimal truncation radius which can maximize  $\tau$  by traversing the truncation radius. Obviously, searching the optimal truncation radius in this way is very time-consuming, fortunately, in subsequent experiments, we find that MCGM still performs very competitively even if the truncation radius is just set to 2. For each network, the best algorithm is emphasized by bold.

(GC+), the extended version of improved gravity centrality<sup>19</sup> (IGC+) and local gravity model<sup>20</sup> (LGM). Table 4 compares the accuracies of MCGM and the ten benchmark algorithms for  $\beta = \beta_c$ . Furthermore, the accuracies of different  $\beta$  values (not too far from  $\beta_c$ ) are shown in Fig. 2.

As shown in Table 4, the methods based on gravity law (GC+, IGC+, LGM and MCGM) show great advantages over the classic methods (DC, H-index, KS, EC, BC, CC), especially in Power, Router and NS, the advantage of the methods based on gravity law are extremely obvious. Notice that, except the above three networks, the performance of EC is significantly better than other classic methods, and even performs competitively in comparison with the methods based on gravity law, which indirectly shows that the stability of the method based on the gravity law is better and their performance will not decline precipitously due to the differences of networks. Furthermore, for the methods based on gravity law, MCGM generally performs best since it effectively considers more characteristics of nodes. As shown in Fig. 2, MCGM still performs very competitively compared with the ten benchmark algorithms for different  $\beta$  not too far from  $\beta_c$ , suggesting the robustness of our findings.

Figure 3 shows the optimal truncation radius of MCGM in the ten real networks. It is not difficult to find that the optimal truncation radius of most networks is concentrated at  $R = 2$ . Therefore, we may simply set  $R = 2$  to test the performance of MCGM. Table 5 compares the accuracies of MCGM with  $R = 2$  and the benchmark algorithms.

As shown in Table 5, MCGM with  $R = 2$  generally performs best in comparison with the benchmark algorithms, it still obtains the best results in six of the ten real networks. Since the optimal truncation radius approximately scales linearly with the average distance<sup>20</sup>, if the average distance of the network is relatively large, setting  $R = 2$  will have a significant impact on the performance of MCGM, such as Power whose average distance is 18.9892. Fortunately, most real networks have small-world property,  $R^*$  tends to be small in most cases.

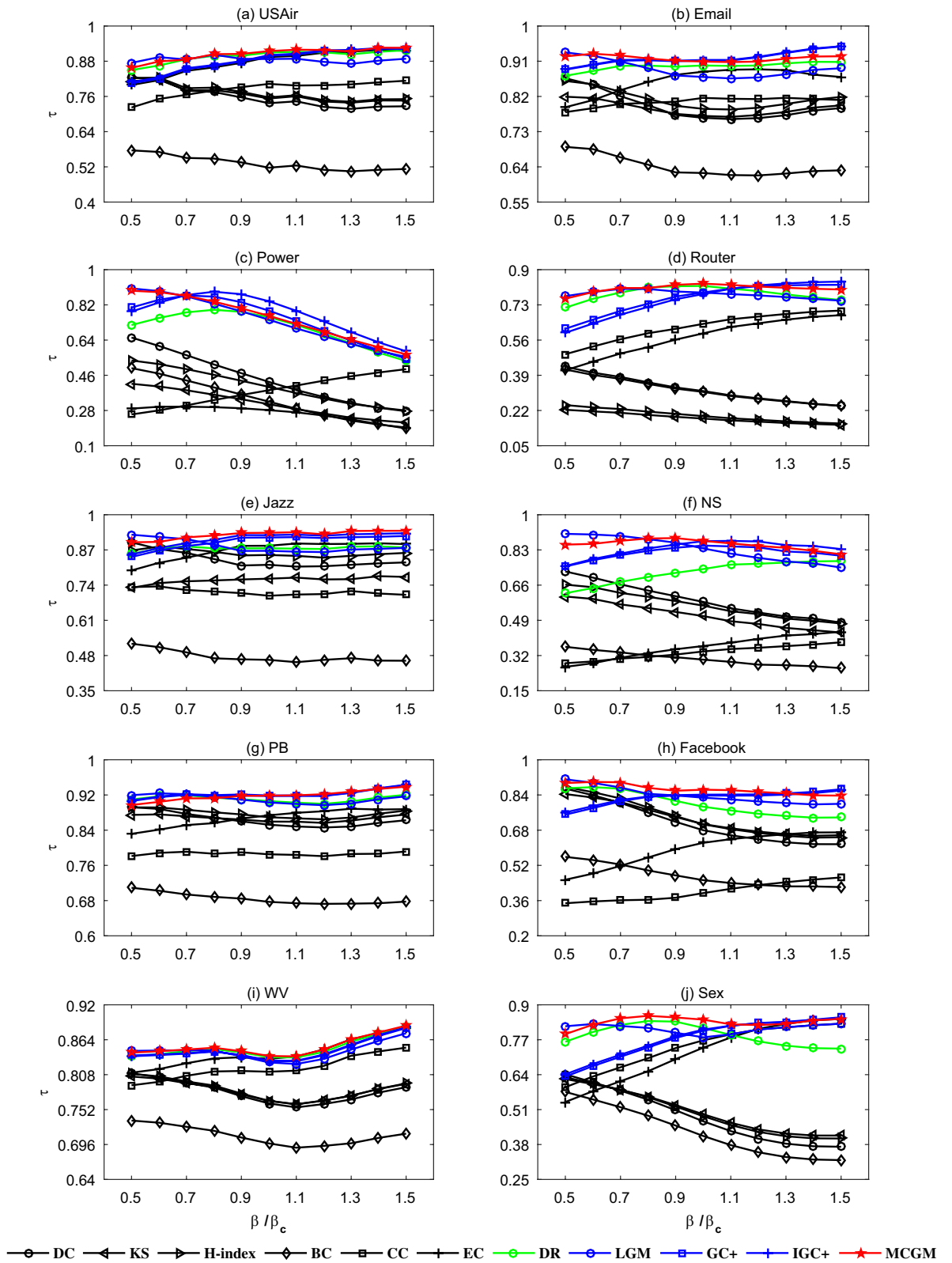
Furthermore, we need to compare MCGM and MCGM without normalization to illustrate the importance of normalization. Table 6 compares the accuracies of MCGM using Eq. (15), MCGM using Eq. (16) and MCGM using Eq. (18). As shown in Table 6, MCGM has been gradually improved by normalization, suggesting the importance of normalization and the effectiveness of our normalization strategy.

Finally, we apply the monotonicity<sup>34</sup> to measure the resolution of different algorithms. As shown in Table 7, MCGM generally performs best even if it only considers 1-order neighbors or 2-order neighbors in most cases. The results reported in Table 7 demonstrate MCGM is a remarkably high-resolution algorithm.

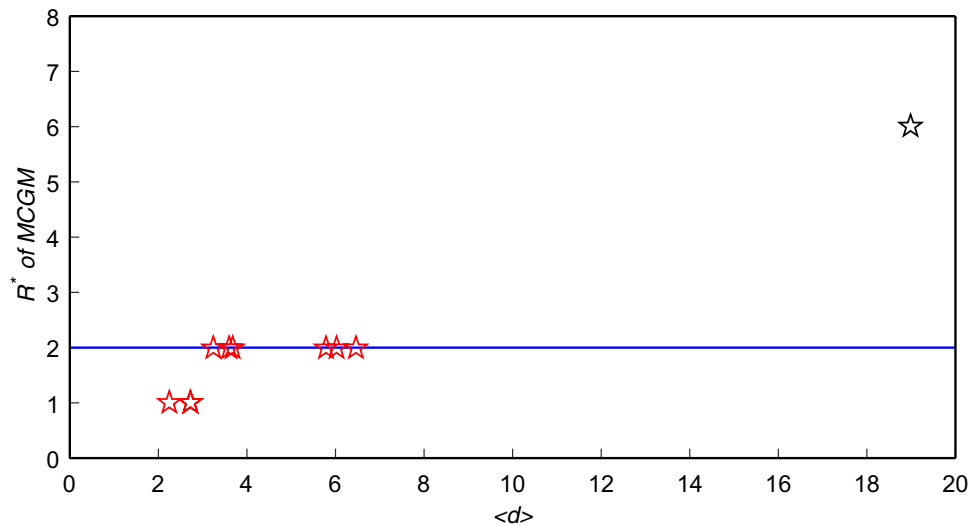
**Computational complexity.** The computational complexity of the methods used in this paper is shown in Table 8. The computational complexity of DC, KS and EC is  $O(N)$ ,  $O(M)$  and  $O(N + M)$ , respectively. Therefore, it is obvious that the part with the highest computational complexity of MCGM is computing the  $R$ -order neighbors of each node, it needs  $N \langle k \rangle^R$  times operations. Hence the computational complexity of MCGM is  $O(N \langle k \rangle^R)$ . Since most real networks have small-world property,  $R^* = 2$  in most cases (see Fig. 3), so the computational complexity of MCGM is generally not more than  $O(N \langle k \rangle^2)$ , where  $\langle k \rangle \ll N$ .

## Discussion

In summary, we propose a novel gravity model that effectively integrates multi-characteristics of nodes, named as multi-characteristics gravity model (MCGM). The number of neighbors, the influence of neighbors, the location of nodes, and the path information between nodes are all taken into consideration in our model. In addition, we propose a normalization strategy to solve the problem that different indices are not in the same order of magnitude, Table 6 suggests the importance of normalization and the effectiveness of our normalization strategy. Compared with well-known state-of-the-art methods, empirical analyses of the SIR spreading dynamics on ten real networks suggest that our model always performs very competitively, as shown in Table 4.



**Figure 2.** The algorithms' accuracies measured by Kendall's Tau for different  $\beta$ . The six classic algorithms (DC, H-index, KS, EC, BC and CC) are represented by black symbols, DR is represented by green symbols, the typical algorithms based on the gravity law (GC+, IGC+ and LGM) are represented by blue symbols, MCGM is represented by red symbols.



**Figure 3.** The  $R^*$  of MCGM for  $\beta = \beta_c$ . Ten pentagrams represent ten networks and the blue line is  $R = 2$ . The  $R^*$  of MCGM in USAir, Jazz and PB is 1, the  $R^*$  of MCGM in Email, Router, NS, Facebook, WV and Sex is 2, and the  $R^*$  of MCGM in Power is 6.

Networks	DC	H-index	KS	EC	BC	CC	DR	GC+	IGC+	LGM	MCGM ( $R = 2$ )
USAir	0.7370	0.7568	0.7529	0.8946	0.5171	0.8027	<b>0.9096</b>	0.8985	0.9006	0.8875	0.9092
Email	0.7653	0.7883	0.7702	0.8832	0.6243	0.8163	0.8991	0.9119	<b>0.9133</b>	0.8697	0.9091
Power	0.4264	0.4009	0.3122	0.2818	0.3254	0.3838	0.7570	0.7906	<b>0.8387</b>	0.7442	0.6616
Router	0.3139	0.1928	0.1810	0.5924	0.3096	0.6383	0.8215	0.7896	0.7823	0.7894	<b>0.8324</b>
Jazz	0.8150	0.8513	0.7638	0.8854	0.4641	0.7008	0.8761	0.9158	0.9244	0.8666	<b>0.9255</b>
NS	0.5790	0.5610	0.5106	0.3660	0.3003	0.3397	0.7377	0.8511	0.8722	0.8372	<b>0.8736</b>
PB	0.8524	0.8694	0.8595	0.8738	0.6771	0.7852	0.9060	<b>0.9189</b>	0.9176	0.9030	0.9123
Facebook	0.6798	0.7066	0.7075	0.6226	0.4529	0.3940	0.7865	0.8414	0.8372	0.8275	<b>0.8639</b>
WV	0.7619	0.7662	0.7657	0.8334	0.6978	0.8127	0.8360	0.8298	0.8305	0.8276	<b>0.8379</b>
Sex	0.4664	0.4855	0.4925	0.7404	0.4118	0.7677	0.8139	0.8038	0.8076	0.7789	<b>0.8448</b>

**Table 5.** The algorithms’ accuracies of MCGM with  $R = 2$  and the benchmark algorithms measured by Kendall’s Tau for  $\beta = \beta_c$ . For each network, the best algorithm is emphasized by bold.

Networks	MCGM (Eq. 15)	MCGM (Eq. 16)	MCGM (Eq. 18)
USAir	0.8946	0.9060	<b>0.9145</b>
Email	0.8782	0.8986	<b>0.9091</b>
Power	0.7557	0.7569	<b>0.7639</b>
Router	0.7992	0.7983	<b>0.8324</b>
Jazz	0.8888	0.9212	<b>0.9333</b>
NS	0.8428	0.8710	<b>0.8736</b>
PB	0.9047	0.9110	<b>0.9184</b>
Facebook	0.8381	0.8547	<b>0.8639</b>
WV	0.8299	0.8341	<b>0.8379</b>
Sex	0.7877	0.7996	<b>0.8448</b>

**Table 6.** The algorithms’ accuracies of MCGM using Eq. (15), MCGM using Eq. (16) and MCGM using Eq. (18) measured by Kendall’s Tau for  $\beta = \beta_c$ . The parameters are adjusted to their optimal values subject to the largest  $\tau$ . For each network, the best algorithm is emphasized by bold.



Networks	DC	H-index	KS	EC	BC	CC	DR	GC+	IGC+	LGM	MCGM
USAir	0.8586	0.8355	0.8114	0.9951	0.6970	0.9892	0.9951	<b>0.9951</b>	0.9951	0.9933	0.9951
Email	0.8874	0.8583	0.8088	0.9999	0.9400	0.9988	0.9999	0.9999	<b>0.9999</b>	0.9998	0.9999
Power	0.5927	0.3930	0.2460	0.9999	0.8314	0.9998	0.9962	0.9996	0.9997	0.9999	<b>0.9999</b>
Router	0.2886	0.0876	0.0691	0.9964	0.2985	0.9961	0.9956	0.9965	0.9965	0.9964	<b>0.9966</b>
Jazz	0.9659	0.9383	0.7944	0.9994	0.9885	0.9878	0.9993	<b>0.9995</b>	0.9993	0.9991	0.9994
NS	0.7642	0.6825	0.6421	0.9955	0.3388	0.9928	0.9950	0.9954	<b>0.9956</b>	0.9933	0.9955
PB	0.9328	0.9268	0.9064	0.9993	0.9489	0.9980	0.9993	0.9993	0.9993	0.9991	<b>0.9993</b>
Facebook	0.9739	0.9665	0.9419	0.9999	0.9855	0.9967	0.9999	0.9999	0.9999	0.9999	<b>0.9999</b>
WV	0.7761	0.7732	0.7673	0.9996	0.7704	0.9994	0.9996	0.9996	0.9996	0.9996	<b>0.9996</b>
Sex	0.6002	0.5457	0.5288	0.9997	0.6757	0.9996	0.9996	0.9997	0.9997	0.9997	<b>0.9997</b>

**Table 7.** The monotonicity of different algorithms. The parameters in the related algorithms (i.e., LGM and MCGM) are adjusted to their optimal values subject to the largest  $\tau$ . For each network, the best algorithm is emphasized by bold.

Methods	Topology	Complexity
DC	Local	$O(N)$
H-index	Semi-local	$O(N + M)$
KS	Global	$O(M)$
EC	Global	$O(N + M)$
BC	Global	$O(NM + N^2 \log N)$
CC	Global	$O(NM + N^2 \log N)$
DR	Semi-local	$O(N(k)^3)$
GC+	Global	$O(N(k)^3)$
IGC+	Global	$O(N(k)^3)$
LGM	Semi-local	$O(N(k)^R)$
MCGM	Global	$O(N(k)^R)$

**Table 8.** The computational complexity of MCGM and the benchmark algorithms.

However, MCGM needs to find the optimal truncation radius by traversing the truncation radius and it is very time-consuming. Fortunately, the optimal truncation radius approximately scales linearly with the average distance<sup>20</sup>, and most real networks have small-world property<sup>37,48</sup>, so even if the truncation radius is just set to 2, MCGM still performs very competitively in most cases, as shown in Table 5. Therefore, without increasing the computational complexity, MCGM effectively considers more characteristics of nodes and obtains more accurate results.

Although the computational complexity of MCGM is not high, it needs the global topological structure, same as GC+ and IGC+. While LGM can work under the case where the global topology is not known. As a result, our suggestions for practical use are as follows: if the network's global topology is known, apply MCGM and set  $R$  to 2, otherwise, apply LGM and set  $R$  to 2 or 3.

Of course, there are still some potential problems in the future. First of all, the gravity law is symmetrical, but due to the different effects of different nodes or the inherent asymmetry of dynamics<sup>49,50</sup>, an asymmetric form of the gravity law may be relevant. Secondly, in weighted complex networks, the heterogeneity of links greatly changes nodes' importance<sup>31</sup>, a weighted form of the gravity law may be relevant. Finally, in order to establish a unified research framework, a unified gravity model is needed to be proposed. Although GC+, IGC+ and LGM are proposed from different perspectives, a unified form of expression exists. We propose a rough model which intends to start further discussion on this issue. The rough unified gravity model is described as

$$UGM(i) = \sum_{d(i,j) \leq R, j \neq i} \frac{(ak(i) + (1-a)k_s(i))(bk(j) + (1-b)k_s(j))}{d(i,j)^2}, \quad (19)$$

where  $a$  and  $b$  are adjustable parameters. If  $a = 1$  and  $b = 1$ , the unified gravity model degenerates to LGM. If  $a = 0$  and  $b = 0$ , the unified gravity model degenerates to GC (GC+ can be obtained by Eq. (6)). If  $a = 0$  and  $b = 1$ , the unified gravity model degenerates to IGC (IGC+ can be obtained by Eq. (8)).

### Data availability

All relevant data are available at <https://github.com/MLIF/Network-Data>.

## References

- Barabási, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
- Zhu, P., Zhi, Q., Guo, Y. & Wang, Z. Analysis of epidemic spreading process in adaptive networks. *IEEE Trans. Circuits Syst. II Express Briefs* **66**, 1252–1256 (2018).
- Borge-Holthoefer, J. & Moreno, Y. Absence of influential spreaders in rumor dynamics. *Phys. Rev. E* **85**, 026116 (2012).
- Xu, W. *et al.* Identifying structural hole spanners to maximally block information propagation. *Inf. Sci.* **505**, 100–126 (2019).
- Albert, R., Albert, I. & Nakarado, G. L. Structural vulnerability of the North American power grid. *Phys. Rev. E* **69**, 025103 (2004).
- Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G. & Nussinov, R. Structure and dynamics of molecular networks: A novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* **138**, 333–408 (2013).
- Bellingeri, M. & Bodini, A. Food web's backbones and energy delivery in ecosystems. *Sci. Rep.* **125**, 586–594 (2016).
- Bellingeri, M., Cassi, D. & Vincenzi, S. Increasing the extinction risk of highly connected species causes a sharp robust-to-fragile transition in empirical food webs. *Ecol. Model.* **251**, 1–8 (2013).
- Lü, L. *et al.* Vital nodes identification in complex networks. *Phys. Rep.* **650**, 1–63 (2016).
- Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *Math. Sociol.* **2**, 113–120 (1972).
- Lü, L., Zhou, T., Zhang, Q. M. & Stanley, H. E. The H-index of a network node and its relation to degree and coreness. *Nat. Commun.* **7**, 10168 (2016).
- Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010).
- Bonacich, P. Some unique properties of eigenvector centrality. *Soc. Netw.* **29**, 555–564 (2007).
- Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40**, 35–41 (1977).
- Freeman, L. C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1**, 215–239 (1979).
- Karemera, D., Oguledo, V. I. & Davis, B. A gravity model analysis of international migration to North America. *Appl. Econ.* **32**, 1745–1755 (2000).
- Porojan, A. Trade flows and spatial effects: The gravity model revisited. *Open Econ. Rev.* **12**, 265–280 (2001).
- Ma, L. L., Ma, C., Zhang, H. F. & Wang, B. H. Identifying influential spreaders in complex networks based on gravity formula. *Phys. A* **451**, 205–212 (2015).
- Wang, J., Li, C. & Xia, C. Improved centrality indicators to characterize the nodal spreading capability in complex networks. *Appl. Math. Comput.* **334**, 388–400 (2018).
- Li, Z. *et al.* Identifying influential spreaders by gravity model. *Sci. Rep.* **9**, 8387 (2019).
- Liu, F., Wang, Z. & Deng, Y. GMM: A generalized mechanics model for identifying the importance of nodes in complex networks. *Knowl. Based Syst.* **193**, 105464 (2020).
- Yang, X. & Xiao, F. An improved gravity model to identify influential nodes in complex networks based on k-shell method. *Knowl. Based Syst.* **227**, 107198 (2021).
- Shang, Q., Deng, Y. & Cheong, K. H. Identifying influential nodes in complex networks: Effective distance gravity model. *Inform. Sciences* **577**, 162–179 (2021).
- Ullah, A. *et al.* Identification of nodes influence based on global structure model in complex networks. *Sci. Rep.* **11**, 6173 (2021).
- Yan, X., Cui, Y. & Ni, S. Identifying influential spreaders in complex networks based on entropy weight method and gravity law. *Chinese Phys. B* **29**, 048902 (2020).
- Li, H., Shang, Q. & Deng, Y. A generalized gravity model for influential spreaders identification in complex networks. *Chaos Solitons Fract.* **143**, 110456 (2021).
- Li, Z. & Huang, X. Y. Identifying influential spreaders in complex networks by an improved gravity model. *Sci. Rep.* **11**, 22194 (2021).
- Huang, X., Chen, D., Wang, D. & Ren, T. Identifying influencers in social networks. *Entropy* **22**, 450 (2020).
- Maji, G., Namtirtha, A., Dutta, A. & Malta, M. C. Influential spreaders identification in complex networks with improved k-shell hybrid method. *Expert Syst. Appl.* **144**, 113092 (2020).
- Wang, X., Yang, Q., Liu, M. & Ma, X. Comprehensive influence of topological location and neighbor information on identifying influential nodes in complex networks. *PLoS ONE* **16**, e0251208 (2021).
- Hethcote, H. W. The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653 (2009).
- Castellano, C. & Pastor-Satorras, R. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.* **105**, 218701 (2010).
- Kendall, M. A new measure of rank correlation. *Biometrika* **30**, 81–89 (1938).
- Bae, J. & Kim, S. Identifying and ranking influential spreaders in complex networks by neighborhood coreness. *Phys. A* **395**, 549–559 (2014).
- Batageli, V. & Mrvar, A. Pajek Datasets. Available at <http://vlado.fmf.uni-lj.si/pub/networks/data/>. (2007).
- Guimerà, R., Danon, L., Díaz-Guilera, A., Giralt, F. & Arenas, A. Self-similar community structure in a network of human interactions. *Phys. Rev. E* **68**, 065103 (2003).
- Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
- Spring, N., Mahajan, R., Wetherall, D. & Anderson, T. Measuring ISP topologies with rocketfuel. *IEEE/ACM Trans. Networking* **12**, 2–16 (2004).
- Gleiser, P. & Danon, L. Community structure in Jazz. *Adv. Complex Syst.* **6**, 565 (2003).
- Newman, M. E. J. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **74**, 036104 (2006).
- Adamic, L. A. & Glance, N. The political blogosphere and the 2004 U.S. election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*. 36–43 (ACM Press, 2005).
- Mauley, J. J. & Leskovec, J. Learning to discover social circles in ego networks. *Adv. Neural. Inf. Process. Syst.* **25**, 548–556 (2012).
- Leskovec, J., Huttenlocher, D. & Kleinberg, J. Predicting positive and negative links in online social networks. In *Proceedings of the 19th international conference on World Wide Web*. 641–650 (ACM Press, 2010).
- Rocha, L. E., Liljeros, F. & Holme, P. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Comput. Biol.* **7**, e1001109 (2011).
- Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).
- Hu, H. B. & Wang, X. F. Unified index to quantifying heterogeneity of complex networks. *Phys. A* **387**, 3769–3780 (2008).
- Chen, D. B., Sun, H. L., Tang, Q., Tian, S. Z. & Xie, M. Identifying influential spreaders in complex networks by propagation probability dynamics. *Chaos* **29**, 033120 (2019).
- Amaral, L. A. N., Scala, A., Barthélemy, M. & Stanley, H. E. Classes of small-world networks. *PNAS* **97**, 11149–11152 (2000).
- Yan, G., Fu, Z. Q. & Chen, G. Epidemic threshold and phase transition in scale-free networks with asymmetric infection. *Eur. Phys. J. B* **65**, 591–594 (2008).
- Wang, W. *et al.* Asymmetrically interacting spreading dynamics on complex layered networks. *Sci. Rep.* **4**, 5097 (2014).
- Bellingeri, M., Bevacqua, D., Scotognella, F. & Cassi, D. The heterogeneity in link weights may decrease the robustness of real-world complex weighted networks. *Sci. Rep.* **9**, 10692 (2019).

## Acknowledgements

The authors greatly appreciate the reviews' suggestions and the editor's encouragement.

## Author contributions

Z.L. devised the research project. Z.L. performed the research. Z.L. and X.H. analyzed the data. Z.L. and X.H. wrote the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Z.L. or X.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022