Rank the spreading influence of nodes using dynamic Markov process

Lin, Jianhong ; Chen, Bo-Lun ; Yang, Zhao ; Liu, Jian-Guo ; Tessone, Claudio J

DOI: https://doi.org/10.1088/1367-2630/acb590

PAPER • OPEN ACCESS

Rank the spreading influence of nodes using dynamic Markov process

To cite this article: Jianhong Lin *et al* 2023 *New J. Phys.* **25** 023014

IOP Publishing New J. Phys. 25 (2023) 023014 https://doi.org/10.1088/1367-2630/acb590

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
22 June 2022

REVISED
12 October 2022

ACCEPTED FOR PUBLICATION
9 January 2023

PUBLISHED
10 February 2023

PAPER

Rank the spreading influence of nodes using dynamic Markov process

Jianhong Lin[1,2,3,4,5,*], **Bo-Lun Chen**[1,6,*], **Zhao Yang**[7], **Jian-Guo Liu**[8,*] **and Claudio J. Tessone**[3,4,*]

[1] Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 233003, People's Republic of China
[2] Department of Mathematics, College of Information Science and Technology, Jinan University, Guangzhou 510632, China
[3] Blockchain and Distributed Ledger Technologies, Institute of Informatics, University of Zürich, Andreasstrasse 15, CH-8050 Zürich, Switzerland
[4] UZH Blockchain Center, University of Zürich, Andreasstrasse 15, CH-8050 Zürich, Switzerland
[5] ETH Zürich, Department of Management, Technology and Economics, Scheuchzerstrasse 7, CH-8092 Zürich, Switzerland
[6] Swiss Center for Data and Network Sciences, University of Zurich, CH-8050 Zürich, Switzerland
[7] Global People Analytics, Talent Rewards & Insights, Roche, CH-4070 Basel, Switzerland
[8] Institute of Accounting and Finance, Shanghai University of Finance and Economics, Shanghai 200433, People's Republic of China
[*] Authors to whom any correspondence should be addressed.

E-mail: jialin@ethz.ch, chenbolun@hyit.edu.cn, liujg004@ustc.edu.cn and tessone@ifi.uzh.ch

Keywords: influence of nodes, complex networks, Markov process

Abstract

Ranking the spreading influence of nodes is of great importance in practice and research. The key to ranking a node's spreading ability is to evaluate the fraction of susceptible nodes being infected by the target node during the outbreak, i.e. the outbreak size. In this paper, we present a dynamic Markov process (DMP) method by integrating the Markov chain and the spreading process to evaluate the outbreak size of the initial spreader. Following the idea of the Markov process, this method solves the problem of nonlinear coupling by adjusting the state transition matrix and evaluating the probability of the susceptible node being infected by its infected neighbors. We have employed the susceptible-infected-recovered and susceptible-infected-susceptible models to test this method on real-world static and temporal networks. Our results indicate that the DMP method could evaluate the nodes' outbreak sizes more accurately than previous methods for both single and multi-spreaders. Besides, it can also be employed to rank the influence of nodes accurately during the spreading process.

1. Introduction

Complex networks are widely used to represent interactions between people, technology, and various entities. Among all the studies within the area of network theory, understanding the dynamics of spreading processes is of particular interest. Although the spreading dynamics on networks are not a new phenomenon, studies in this field lead to better understandings of many important social and natural processes [1], such as the spreading of infectious diseases [2–4], the propagation of computer virus [5], the cascading process [6], traffic congestion [7], the centralization in Bitcoin system [8–10], and so on. One important approach in studying the spreading dynamics is to estimate and rank nodes' spreading abilities. Through this approach, one might first locate influential nodes of complex networks and later on control the outbreak of epidemics [11–13], target the opinion leaders in social networks [14–16], quantify the scientific impact [17, 18], and accelerate the adoption of innovation [19], etc.

Classical centrality measures have been developed to identify the nodes' spreading influence. The degree centrality [20] is probably the most straightforward one. Nodes with a larger degree centrality are considered to have better spreading abilities than the other nodes within a graph. The betweenness centrality [21], which calculates the number of shortest paths crossing through a certain node, represents the controllability of information flow over the networks. The closeness centrality [22] measures the inverse of the mean geodesic distance from a certain node to all other nodes. The more central a node is, the closer it is to all the other

nodes. The eigenvector centrality [23] assigns relative scores to all nodes in the network based on the concept that connections to influential nodes, i.e. high-scoring nodes, would be more important than that to low-scoring nodes. The k-shell decomposition method [24] assigns nodes to different shells and considers those located within the core of the network are the most efficient spreaders. Furthermore, a lot of methods for identifying the node spreading influence have been developed from different perspectives [25–31].

The classical centrality measures are based on the network topological structure solely. However, recent studies have shown that the nodes' spreading influence is determined not only by the network structure but also by the parameters of the dynamical processes [32–35]. Therefore, various structural-based centralities cannot properly identify nodes' influences since the rankings remain the same under different dynamical parameters. Šikić *et al* [32] argued that for a given susceptible-infected-recovered (SIR) model [36], the rank of nodes' influence largely depends on the spreading rate and recovering rate. Klemn *et al* [33] suggested that the eigenvector centrality could only identify the nodes' spreading influence accurately when the spreading rate is close to the inverse of the largest eigenvalue of the network [36]. Considering the susceptible-infected-susceptible (SIS) model [37], Ide *et al* [38] have proposed a numerical framework that uses the importance of the centrality type to determine how the vulnerable nodes change along the diffusion phases. Besides, Liu *et al* [34] have described the infectious probabilities of nodes by a matrix differential function and have developed the dynamics-sensitive (DS) centrality to predict the outbreak size for ranking nodes' spreading influence.

The centrality measures proposed in [32–34] are all linear methods based on discrete Markov process. However, it is important to note that the spreading process in SIR and SIS models is usually a non-linear couple process. Therefore, without taking the non-linear couple process into consideration, all the nodes' influence would be overestimated. For instance, if a susceptible node has n' infected nodes, the probability of this node being infected is $1-(1-\beta)^{n'}$ instead of $n'\beta$ approximated by the linear methods, where β is the spreading rate in the SIR or SIS model. In this paper, we present a dynamic Markov process (DMP) to evaluate the outbreak size of the nodes at given time steps. This method can be directly applied in ranking nodes' spreading influence. It overcomes the problem of nonlinear coupling by calculating the susceptible node to be infected by its neighbors sequentially and adjusting the state transition matrix during the spreading process. Our simulation results on SIR model show that the DMP method has comparable accuracy to the linear methods [32–34] for both single spreader and multi-spreaders [39]. Furthermore, we have employed the SIR and SIS models to test the DMP method on real static and dynamic networks [36, 37, 40, 41]. The simulation results show that the DMP method can rank the spreading influence of nodes accurately.

2. Methods

2.1. Centrality measures

A network $G=(V,E)$ with $n=|V|$ nodes and $e=|E|$ links could be described by an adjacency matrix $\mathbf{A}=\{a_{ij}\}$ where $a_{ij}=1$ if node i is connected to node j, and $a_{ij}=0$ otherwise. For directed network, if only node i is pointing to node j, then $a_{ij}=1$ and $a_{ji}=0$.

The degree of node i is defined as the number of its neighbors, namely

$$k_i=\sum_{j=1}^{n}a_{ij}, \tag{1}$$

where a_{ij} is the element of matrix $\mathbf{A}$.

The main idea of eigenvector centrality is that a node's importance is not only determined by itself, but also by its neighbors' importance [23]. Accordingly, eigenvector centrality of node i, i.e. v_i, is defined as

$$v_i=\frac{1}{\lambda}\sum_{j=1}^{n}a_{ij}v_j, \tag{2}$$

where λ is a constant. Obviously, equation (2) can be written in a compact form as

$$\mathbf{Av}=\lambda\mathbf{v}, \tag{3}$$

where $\mathbf{v}=(v_1,v_2,\ldots,v_n)^T$. That is to say, $\mathbf{v}$ is the eigenvector of the adjacency matrix $\mathbf{A}$ and λ is the corresponding eigenvalue. According to Perron–Frobenius Theorem [42], the elements of the leading

eigenvector are positive. Since the influences of nodes should be positive, $\mathbf{v}$ must be the leading eigenvector corresponding to the largest eigenvalue of $\mathbf{A}$, therefore we have $\mathbf{v} = \mathbf{q}_1$.

2.2. DMP method

In order to calculate the probabilities of nodes being infected, one needs to solve the problem of the nonlinear couple during the spreading process. Consider the union of a finite number of event $A'_1, \ldots, A'_{n'}$, the probability of the event $\cup_{i=1}^{n'} A'_i$ could be written as [43]:

$$P(\cup_{i=1}^{n'} A'_i) = 1 - \prod_{i=1}^{n'} (1 - P(A'_i)) = P(A'_1) + (1 - P(A'_1))P(A'_2) + \cdots + \prod_{i=1}^{n'-1} (1 - P(A'_i))P(A'_{n'}). \quad (4)$$

The equation (4) means that the probability of the event $\cup_{i=1}^{n'} A'_i$ equals the probability of the event A'_1 plus the probability of the event A'_2 while event A'_1 does not happen plus the probability of the event A'_3 while both event A'_1 and A'_2 do not happen, so on and so forth. Based on the above-described probability theorem, for a susceptible node with n' infected neighbors, the probability of being infected during the spreading process can be represented in the non-linear format as $1 - (1 - \beta)^{n'}$, or as $\beta + (1 - \beta)\beta + (1 - \beta)^2\beta + \cdots + (1 - \beta)^{n'-1}\beta$, where the first term β represents the probability that this node has been infected by its first infected neighbor, the second term $(1 - \beta)\beta$ represents the probability that it has not been infected by its first infected neighbor but has been infected by its second infected neighbor, and the third one $(1 - \beta)^2\beta$ represents the probability that the node has not been infected by its first infected neighbor nor its second infected neighbor but has been infected by its third infected neighbor, etc.

By combining the above-described process with the standard SIR model where an infected node would infect its susceptible neighbors with a spreading rate β and recover immediately, we propose a dynamics Markov process method as follows: define $\mathbf{x}(t)$ $(t \geqslant 0)$ as an $n \times 1$ vector whose components are approximated as the probabilities of nodes to be infected at time step t. Especially, if node i is the initially infected node, then $x_i(0) = 1$ and $x_{j \neq i}(0) = 0$. In the dynamics Markov process, the initial Markov state transition matrix $\mathbf{M} = \mathbf{A}^T$, where $\mathbf{A}^T$ is the transpose of $\mathbf{A}$. If $m_{ij} = 0$, node j could not be infected by node i anymore. Otherwise, m_{ij} is the probability of node i to be susceptible. When $t = 0$, if node i is the initially infected node, it could not be susceptible anymore. Therefore we have $m_{ij} = 0$, where $j = 1, 2, \ldots n$. When $t \geqslant 1$, we denote $\mathbf{C}(t)$ as an $n \times n$ matrix, where $c_{ji}(t)$ is the probability of node j to be infected by node i at time step t.

The updating rules are described below. We first calculate the influence of node 1 on all its susceptible neighbors. If $x_1(t-1) > 0$, node 1 would infect its susceptible neighbors with a probability β at time step t. The probability of node j being infected by node 1 is then

$$c_{j1}(t) = \beta m_{j1} x_1(t-1), \quad (5)$$

where $j = 1, 2, \ldots n$. After that, for all the nodes j, if $m_{jl} > 0$, where $l = 1, 2, \ldots n$, we calculate the probability of node j to be susceptible, i.e. the probability of node j that has not been infected by node 1. It equals to $m_{jl} - c_{j1}(t)$. After that, for all the nodes j, we update the element of the state transition matrix as follows:

$$m_{jl} := m_{jl} - c_{j1}(t), \quad (6)$$

where $l = 1, 2, \ldots n$. Once the state transition matrix has been updated, we continue to calculate the impact of node $2, 3, 4, \ldots n$ sequentially in the same way. In the end, the probability of node i being infected at time step t is

$$x_i(t) = \sum_{j=1}^{n} c_{ij}, \quad (7)$$

The spreading influence of the target node within a certain time T^* is $\sum_{t=1}^{T^*} \sum_{j=1}^{n} x_j(t)$. During the spreading process, the DMP method solves the problem of nonlinear coupling by calculating the probability of the node being infected by its infected neighbors sequentially via adjusting the state transition matrix $\mathbf{M}$. As shown in figure 1, node 1 and node 3 are the initial spreaders. According to the DMP method, we first calculate the probability of node 2 being infected by node 1, which is β. Then we update the state transition matrix $\mathbf{M}$. After that, we calculate the probability of node 2 to be infected by node 3, which is $\beta(1 - \beta)$. And in the end, we get the probability of node 2 to be infected by its both infected neighbors, which equals $\beta + \beta(1 - \beta) = 1 - (1 - \beta)^2$. We note that this is the exact probability of the node 2 being infected.

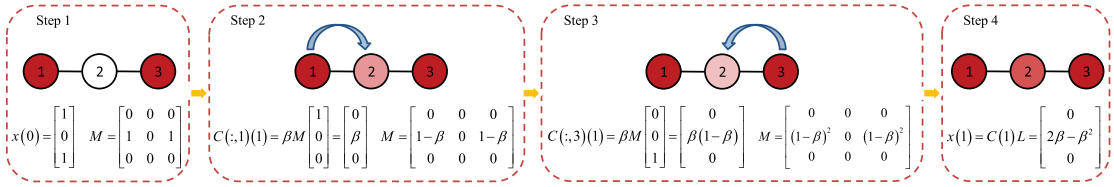

Figure 1. An example network with three nodes, where node 1 and node 3 is the initially infected nodes. The probability of node 2 $x_2(1)$ to be infected at time step 1 is $2\beta-\beta^2$ generated by the DMP method.

The DMP method could be extended to the SIS model with $\gamma=1$, where γ is the probability of the infected nodes entering the susceptible state (see the details in the Data Analysis section). In SIS model, the difference is that at each time step t, the transition matrix $\mathbf{M}$ could be updated by $m_{ij}=a_{ji}(1-x_i(t-1))$. For the temporal network, the network could be described by $\mathbf{A}(t)$ at each time step t. Thus, at time t the transition matrix $\mathbf{M}$ could be updated by $m_{ij}=a_{ji}(t)(1-\sum_{r=0}^{t-1}x_i(r))$.

2.3. Kendall's Tau

In this paper, we use Kendall's tau to measure the correlation between the nodes' spreading influence and centrality measures (e.g. degree, eigenvector centrality and DMP method). For each node i, we denote y_i as its spreading influence and z_i as the target centrality measure, the accuracy of the target centrality in evaluating nodes' spreading influences can be quantified by the Kendall's Tau [44], as

$$\tau=\frac{2}{\sqrt{(n(n-1)/2-n_1)(n(n-1)/2-n_2)}}\sum_{i<j}\mathrm{sgn}[(y_i-y_j)(z_i-z_j)], \tag{8}$$

where $n_1=\sum_i v_i(v_i-1)/2$, v_i is the number of the ith group of ties for the first quantity and $n_2=\sum_j u_j(u_j-1)/2$, u_j is the number of the jth group of ties for the second quantity and $\mathrm{sgn}(y)$ is a piecewise function: when $y>0$, $\mathrm{sgn}(y)=+1$; $y<0$, $\mathrm{sgn}(y)=-1$; when $y=0$, $\mathrm{sgn}(y)=0$. τ measures the correlation between two ranking lists, whose value is between $[-1,1]$ and a larger τ corresponds to better performance.

3. Data and models

3.1. Data description

We test the performance of the DMP method in estimating the nodes' spreading influence according to the SIR and SIS models on four real networks. The first network is 'C. elegans', a directed network representing the neural network of Caenorhabditis elegans [45]. The data is available at https://snap.stanford.edu/data/C-elegans-frontal.html. The second network is a scientific collaboration network, 'Erdós', where nodes are scientists and edges represent the co-authorships. The data can be freely downloaded from the website http://wwwp.oakland.edu/enp/thedata/. The third one is an email communication network of the University Rovira i Virgili of Spain, involving faculty members, researchers, technicians, managers, administrators, and graduate students [46]. The data can be found at http://konect.cc/networks/arenas-email/. The last network is a directed network based on the ODLIS dictionary network. This a hypertext reference resource for library and information science professionals, university students and faculty, and users of all types of libraries. The node represents website of Odlis and the edge represents the connection between two websites. This data is available at http://networkdata.ics.uci.edu/netdata/html/ODLIS.html. The basic statistical properties of these four networks are presented in table 1.

Besides, we have also analyzed four real-world dynamic networks in order to evaluate the effectiveness of the DMP method. The first temporal network is *Contacts in a workplace* (CW) network. This data includes contacts between individuals measured in an office building in France, from June 24 to July 3, 2013 [47]. The second one is the *Primary school* (PS) temporal network, where nodes are the children and teachers, and edges represent the contacts between them [48, 49]. The CW and PS network could be downloaded at www.sociopatterns.org/datasets/. The *email-Eu-core-temporal-Dept1* (EM01) and *email-Eu-core-temporal-Dept2* (EM02) [50] temporal network are generated by using email data from a large European research institution, where edges present email between members of the research institution. These two datasets are available at http://snap.stanford.edu/data/index.html. In table 2, we provide the detailed statistical properties of the above temporal networks.

Table 1. Basic statistical features of C. elegans, Erdös, Email, and Odlis networks, including the number of nodes n, the number of the edges e, the average degree $\langle k\rangle$ or $\langle k_{\text{out}}\rangle$ (for directed networks) and the reciprocal of the largest eigenvalue $1/\lambda_1$.

Network	n	e	$\langle k\rangle, \langle k\rangle_{\text{out}}$	$1/\lambda_1$
C. elegans	297	2345	7.896	0.109
Erdös	454	1313	5.784	0.079
Email	1133	5451	9.622	0.048
Odlis	2900	18241	6.290	0.077

Table 2. Basic statistical features of CW, PS, EM01, and EM02 temporal networks, including the number of nodes n and the number of the edges e respectively.

Network	n	e
CW	92	1492
PS	242	21295
EM01	309	11106
EM02	162	7758

3.2. The SIR and SIS Model

We apply the SIR model [36] and the SIS [37] model to simulate the spreading process and record the nodes' spreading influence at each time step. In the SIR model, there are three kinds of individuals: (a) susceptible individuals that could be infected, (b) infected individuals who can infect their susceptible neighbors, and (c) recovered individuals that will never be infected again. At each time step, every infected node will contact its neighbors and each of its susceptible neighbors will be infected with a probability β. Then the infected nodes enter the recovered state with a probability μ. While in the SIS model, there are only two kinds of individuals, i.e. the susceptible individuals and the infected ones. The infected nodes would infect their susceptible neighbors with the probability β and enter the susceptible state with a probability γ. For single-node spreading, only one seed node is infected at the beginning, and all the other nodes are susceptible. While for multiple-nodes spreading, a set of nodes are infected and the rests are initially susceptible. At each time step t, the number of nodes that switch from the susceptible state to the infected state represents the node's spreading influence. In this paper, all the analyses are based on discrete-time dynamics.

4. Results

4.1. Comparison between the DMP method and DS centrality

Figure 2 shows the comparison between the DMP method and DS centrality for evaluating the spreading influence of nodes on regular network with $N = 1000$ and $\langle k\rangle = 20$. The results suggest that the DMP method could evaluate the outbreak size for both single and multi spreaders more accurately than the DS centrality. One could easily observe that the theoretical results generated by the DMP method are over-estimated. The main reason for this overestimation is that the nodes infected by the initial node would infect themselves when time steps $t \geqslant 3$. One could then expect that a network without any loops would diminish this bias. As shown in figure 3, in a network without any loops, the DMP method could evaluate the spreading scope of nodes accurately compared with the simulation result on networks. We also test the performance of the DMP method for evaluating the outbreak size on empirical networks including the C. elegans, Erdős and Twitter network [51], where in the Twitter network there are $n = 116408$ nodes and $e = 150818$ edges. As shown in figures 4 and 5, the DMP method performs better than the DS centrality for evaluating the outbreak size generated by single and multiple spreaders on empirical networks, with a quality that remains for larger times.

4.2. Ranking the spreading influence of nodes in SIR model

We test the performance of the DMP method in ranking nodes' spreading influence during the spreading process on the SIR model with different spreading rates β. The spreading influence of an arbitrary node i is quantified by the number of infected nodes and recovered nodes at t, where the spreading process starts with only node i being initially infected. Here the Kendall's τ is used to evaluate the correlation between the nodes' spreading influence and the centrality measures (DMP method, degree and eigenvector centrality), where τ is in the range $[-1, 1]$, and a larger value of τ indicates a better performance. As shown in figure 6, in all the

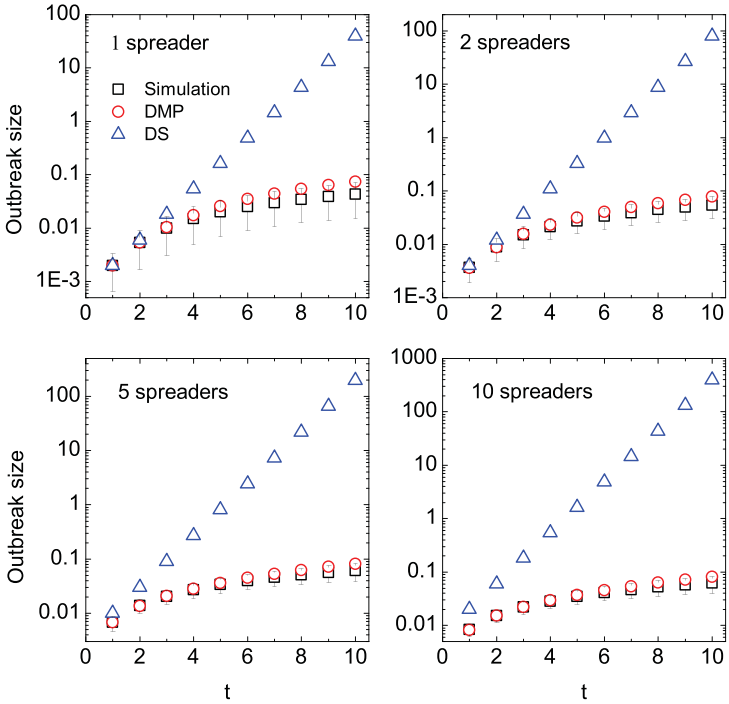

Figure 2. The performance of the DMP method and the DS centrality for evaluating the outbreak generated by both single spreader and multiple spreaders on a regular network of $N = 1000$ and $\langle k \rangle = 20$ during the SIR spreading process with spreading rate $\beta = 0.1$. The symmetric bars indicate the fluctuations around the average value computed on 10^5 realizations of the stochastic process.

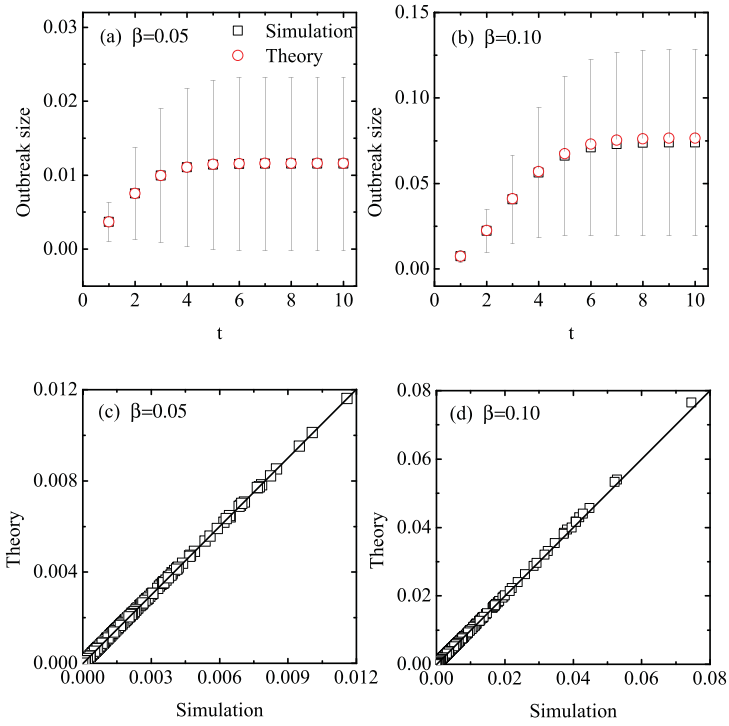

Figure 3. The accuracy of the DMP method for evaluating nodes' spreading influence in a model network without any loops of $N = 500$ and $\langle k \rangle = 8$ during the SIR spreading process. Subplots (a) and (b) show the outbreak size of node 1 in the model network generated by the DMP method at each time step when the spreading rate β is 0.05 and 0.1 respectively. Subplots (c) and (d) show the outbreak size of all nodes in the model network generated by the DMP method when the spreading rate β is 0.05 and 0.1 respectively. The symmetric bars indicate the fluctuations around the average value computed on 10^5 realizations of the stochastic process.

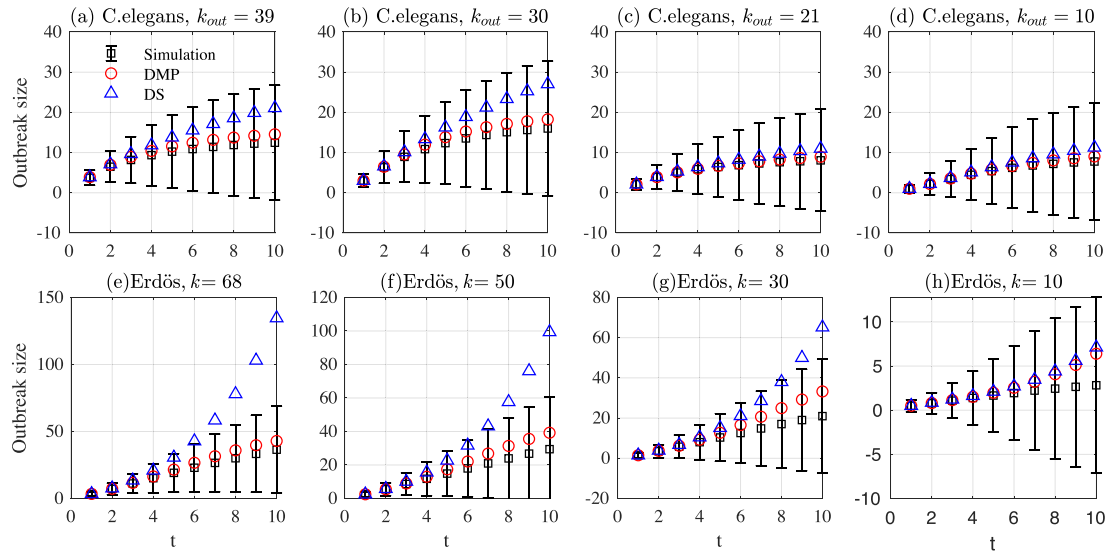

Figure 4. The performance of the DMP method and the DS centrality for evaluating the outbreak of different initial nodes on C. elegans and Erdős networks during the SIR spreading process with spreading rate $\beta = 0.1$. The symmetric bars indicate the fluctuations around the average value computed on 10^5 realizations of the stochastic process.

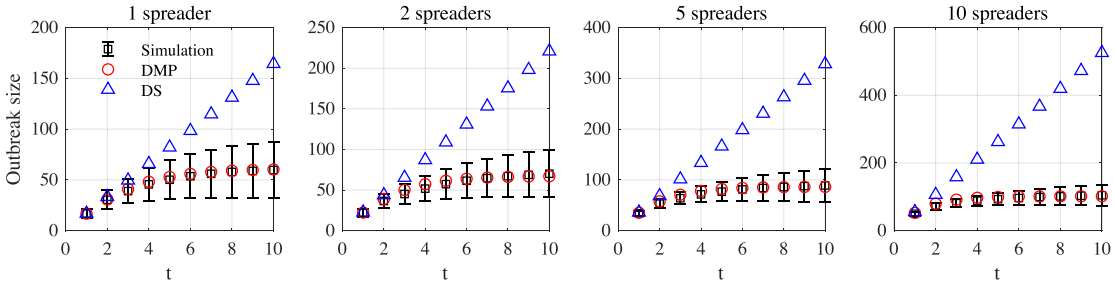

Figure 5. The performance of the DMP Method and the DS centrality for evaluating the outbreak generated by both single spreader and multiple spreaders on the Twitter network during the SIR spreading process with spreading rate $\beta = 0.1$. The symmetric bars indicate the fluctuations around the average value computed on 10^5 realizations of the stochastic process.

cases, the values of τ of the DMP method are always between 0.915 and 1.0, which suggests that the ranking lists generated by the DMP method are almost the same as the ones generated by the simulation result.

Furthermore, one can find that the accuracy of the DMP method for ranking nodes' spreading influence is affected by the network structure. As shown in figure 6, the descent speed of τ in the C. elegans network is smaller than the ones in the Email network. To get a deeper insight into how the network structure affects the performance of the DMP method, we analyze the correlation between the minimum Kendall's tau of the networks in figure 6 and the inverse of the largest eigenvalue of the network $1/\lambda_1$. The results are shown in figure 7. With an increasing value of $1/\lambda_1$, the minimum Kendall's tau increases. For instance, in the C. elegans network, the reciprocal of the largest eigenvalue $1/\lambda_1$ is 0.109, which is significantly larger than that of the Email network (0.048). The fact that the DMP method performs particularly well in the C. elegans network for ranking nodes' spreading influence indicates that the largest eigenvalue of the network is the main factor affecting the accuracy of the DMP method. A larger value of $1/\lambda_1$ would lead to a better performance of the DMP method.

4.3. Parameter sensitivity

In the above subsections, we set $\mu = 1$. To analyze the influence of the parameters on the performance of the DMP method, in figure 8 we set $\mu = 0.1$, 0.5 and fix $\beta = 0.1$, i.e. $\beta/\mu = 1$ and 0.2. We select different nodes as initially infected seeds. According to the simulation results, we still can see that the DMP method performs better than DS centrality for evaluating outbreak size generated by different initial nodes. Figure 9 shows the results with $\mu = 0$ and 1 on a directed tree network. We can see that when $\mu = 1$, both the DMP method and

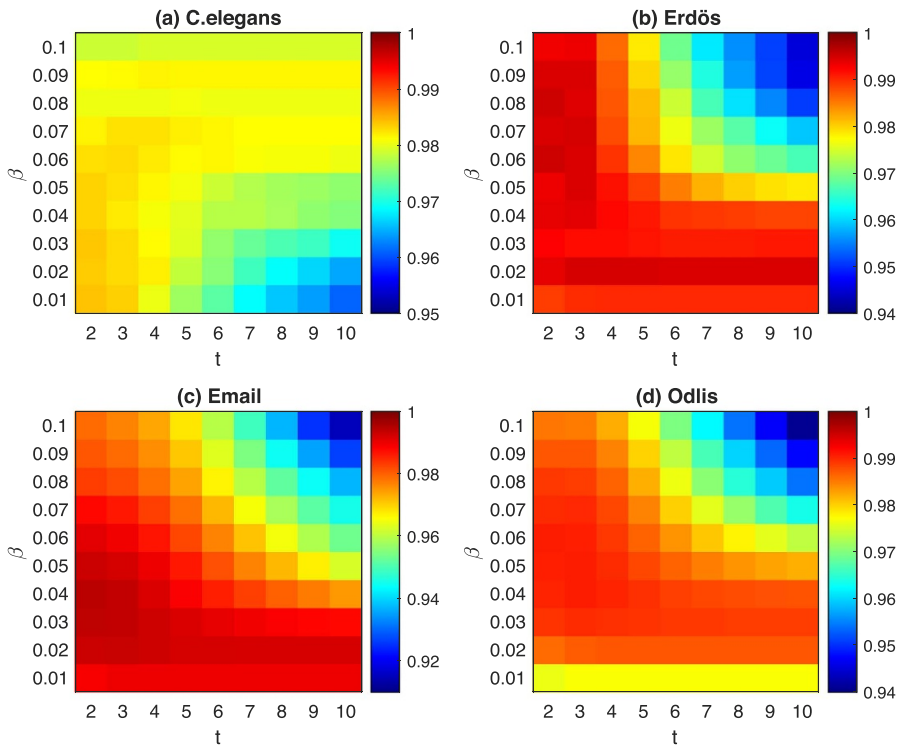

Figure 6. The accuracy of the DMP method in evaluating nodes' spreading influences according to the standard SIR model in the four real networks, quantified by Kendall's Tau. The spreading rate β varies from 0.01 to 0.10. Each data point is obtained by averaging over 10^5 independent runs.

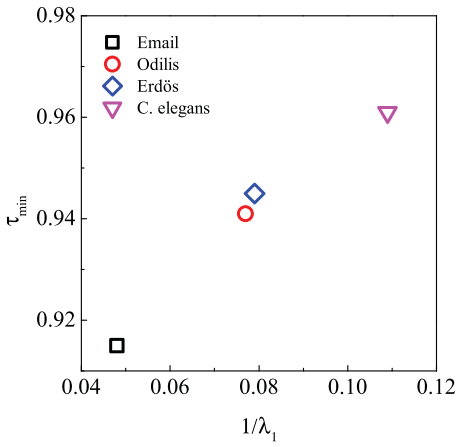

Figure 7. The correlation between the minimum Kendall's tau $\tau_{\min}$ and the inverse of the largest eigenvalue of the network $1/\lambda_1$.

DS centrality can evaluate outbreak size accurately on a directed tree network. Because when $\mu = 1$, in a directed tree network, a susceptible node would only be infected by one infected in-neighbors. In this special situation, the spreading process is a linear couple process. Thus, both the DMP method and the DS centrality can evaluate the outbreak size accurately.

In figure 4, we have shown the spreading influence of nodes during the whole process. In the current experiment, we rank the spreading influence of nodes in a special situation: the running time of the simulation is long enough such that there is not any infected node in the network. For each node, since we do not have the exact convergent time step in the simulation, here we set a fixed time step T^* as 5 in the DMP method. The results are shown in figure 10. The Kendall's tau τ of the DMP method is between 0.893 to 0.995, which indicates that the ranking lists generated by the DMP method and the SIR spreading process are

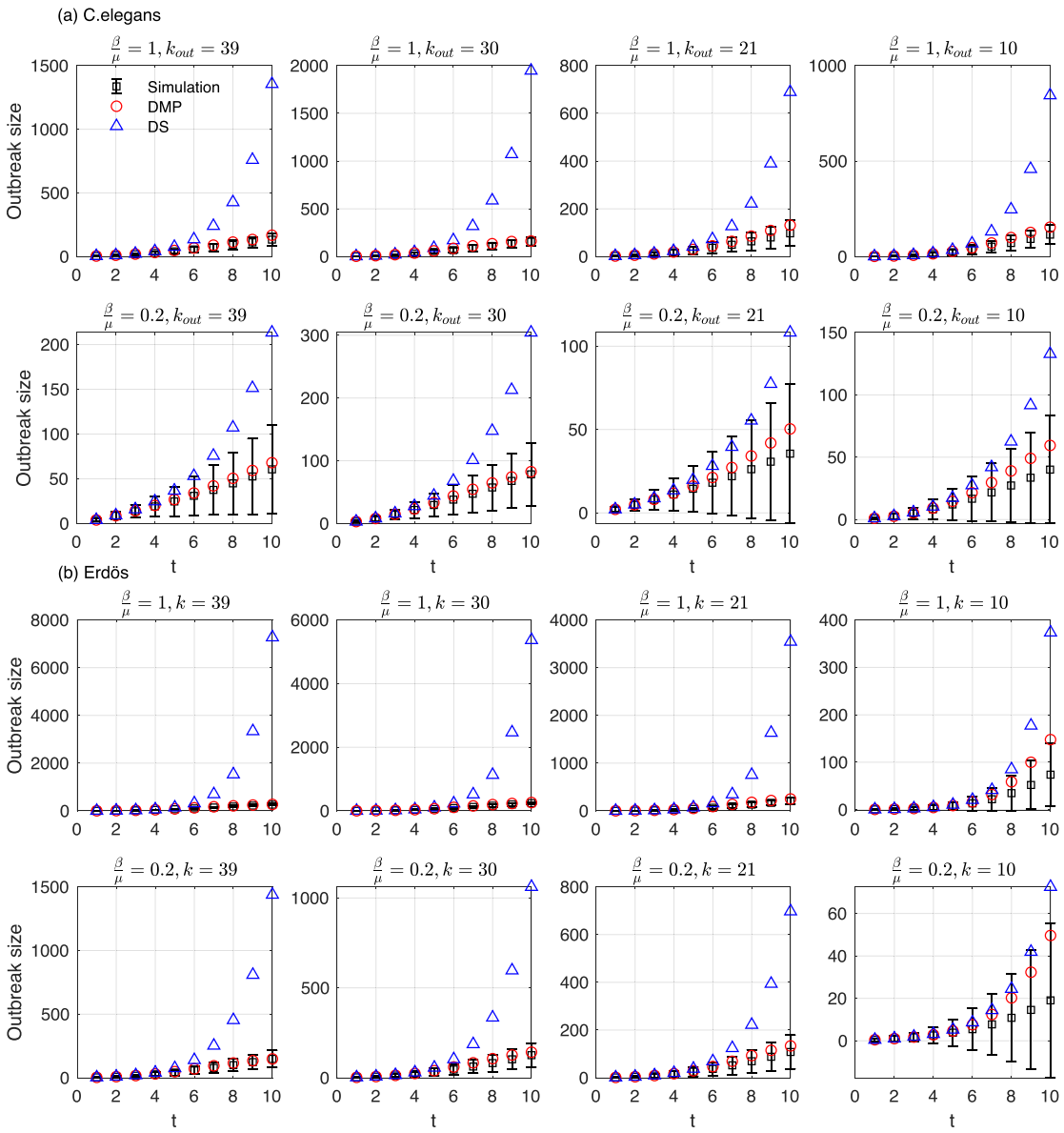

Figure 8. The performance of the DMP method and the DS centrality for evaluating the outbreak of different initial nodes on C. elegans and Erdős networks during the SIR spreading process with spreading rate $\beta = 0.1$. The symmetric bars indicate the fluctuations around the average value computed on 10^5 realizations of the stochastic process.

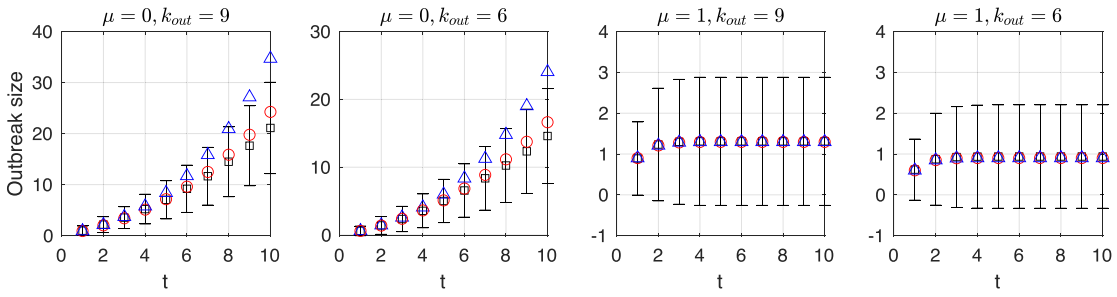

Figure 9. The performance of the DMP method and the DS centrality for evaluating the outbreak of different initial nodes on the directed tree network during the SIR spreading process with spreading rate $\beta = 0.1$. The symmetric bars indicate the fluctuations around the average value computed on 10^5 realizations of the stochastic process.

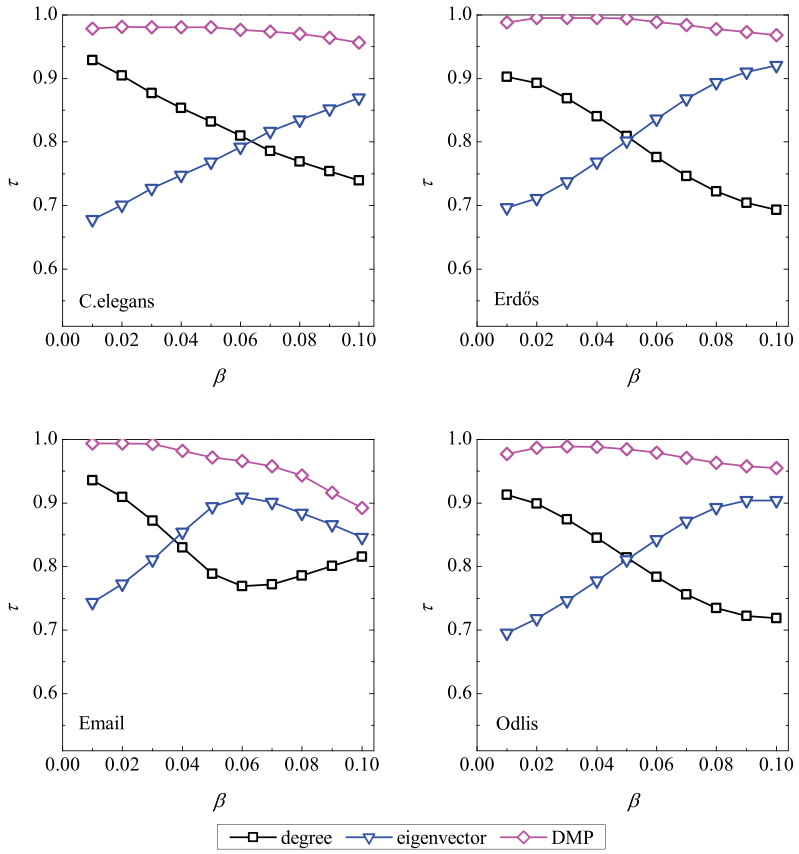

Figure 10. The comparison among the DMP method, degree and eigenvector in evaluating nodes' spreading influences according to the standard SIR model with enough time steps until there are not any infected nodes in the networks, quantified by Kendall's Tau. The spreading rate β varies from 0.01 to 0.10, each data point is obtained by averaging over 10^5 independent runs.

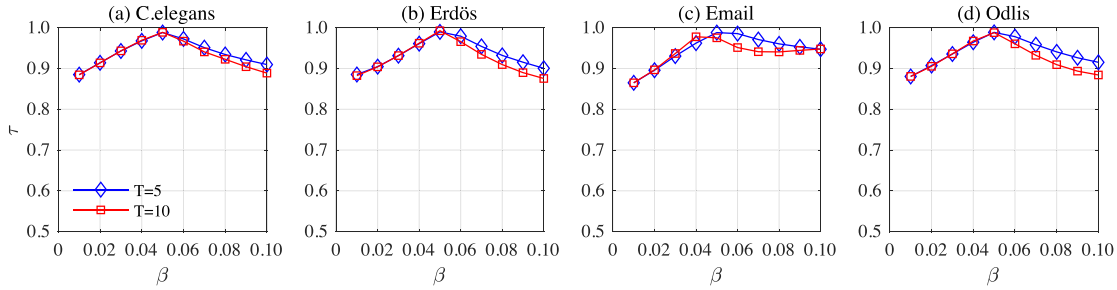

Figure 11. The accuracy of the DMP method with fixed $\beta = 0.05$ in evaluating the spreading influence of nodes in the SIR model with spreading rate β varying from 0.01 to 0.10 in four real networks, quantified by Kendall's Tau τ. In the simulation, the total time steps $T = 5$ and 10. Each data point is obtained by averaging over 10^5 independent runs.

almost the same. Compared with the degree and eigenvector centrality, the DMP method could locate the influential spreaders more accurately.

Another clear limitation of our method is that before applying the DMP method, we have to know the spreading rate, which is usually a hidden parameter. This parameter can be effectively estimated according to the early spreading process [52]. However, if we do not know the spreading rate, we can fix the spreading rate when calculating the DMP method. In figure 11, we fix $\beta = 0.05$ in the DMP method to rank the spreading influence of nodes with different spreading rates in the SIR model. We find that the values of Kendall's τ are between 0.86 and 1. It indicates the DMP method with fixed β can accurately evaluate the spreading influence of nodes in the SIR model with different spreading rates.

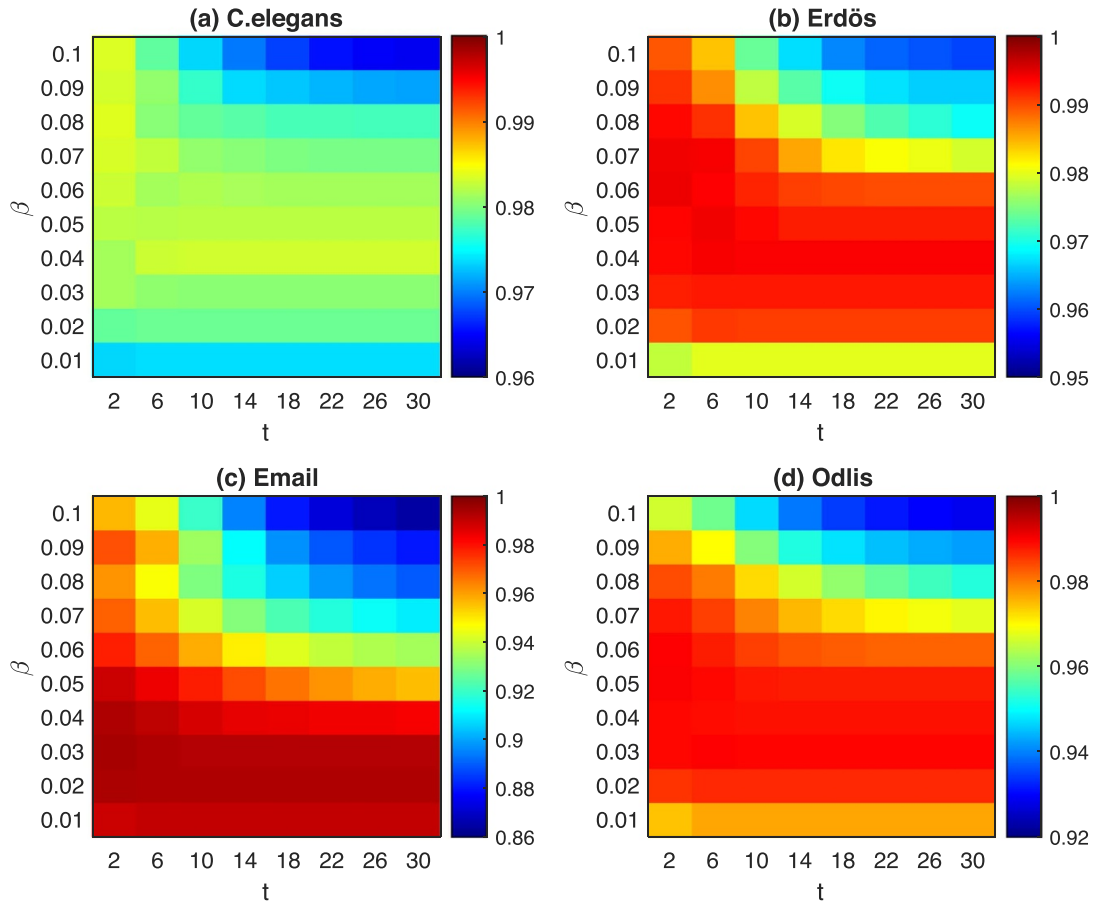

Figure 12. The accuracy of the DMP method in evaluating nodes' spreading influences according to the standard SIS model in four real networks, quantified by Kendall's Tau. The spreading rate β varies from 0.01 to 0.10. Each data point is obtained by averaging over 10^5 independent runs.

4.4. Ranking the spreading influence of nodes in the SIS model

The DMP method can also be used to evaluate the outbreak size in SIS model. For simplicity, we set $\gamma = 1$ in the SIS model. In this case, nodes in the network could be infected several times for high spreading rates β and long time step t. We set the final time step t^* to 30. The results are shown in figure 12. The results are very similar to the ones in the SIR model. The Kendall's tau τ of the DMP method is between 0.865 and 1.0. This indicates that the DMP method could also rank the nodes' spreading influence accurately in the SIS model.

4.5. Ranking the spreading influence of nodes in temporal networks

We have extended the DMP method to temporal networks. The results are shown in figure 13. The values of Kendall's τ are between 0.923 to 0.992, which indicates that the ranking lists generated by the DMP method and the real SIR model on temporal are highly identical to each other. Therefore, the DMP method could be used to detect the influential nodes in the temporal network accurately.

4.6. Pearson correlation coefficient

In figure 14 we check the performance of the DMP method by both Kendall's τ and Pearson correlation coefficient r. We can see that for most of the range, the values of r (Pearson coefficient) are larger than the ones of Kendall's τ. It indicates the ranking lists generated by the DMP are still the same as the simulation results evaluated by the Pearson correlation coefficient. In this way, we can conclude that DMP properly allows us to recover the characteristics of the dynamics on the network.

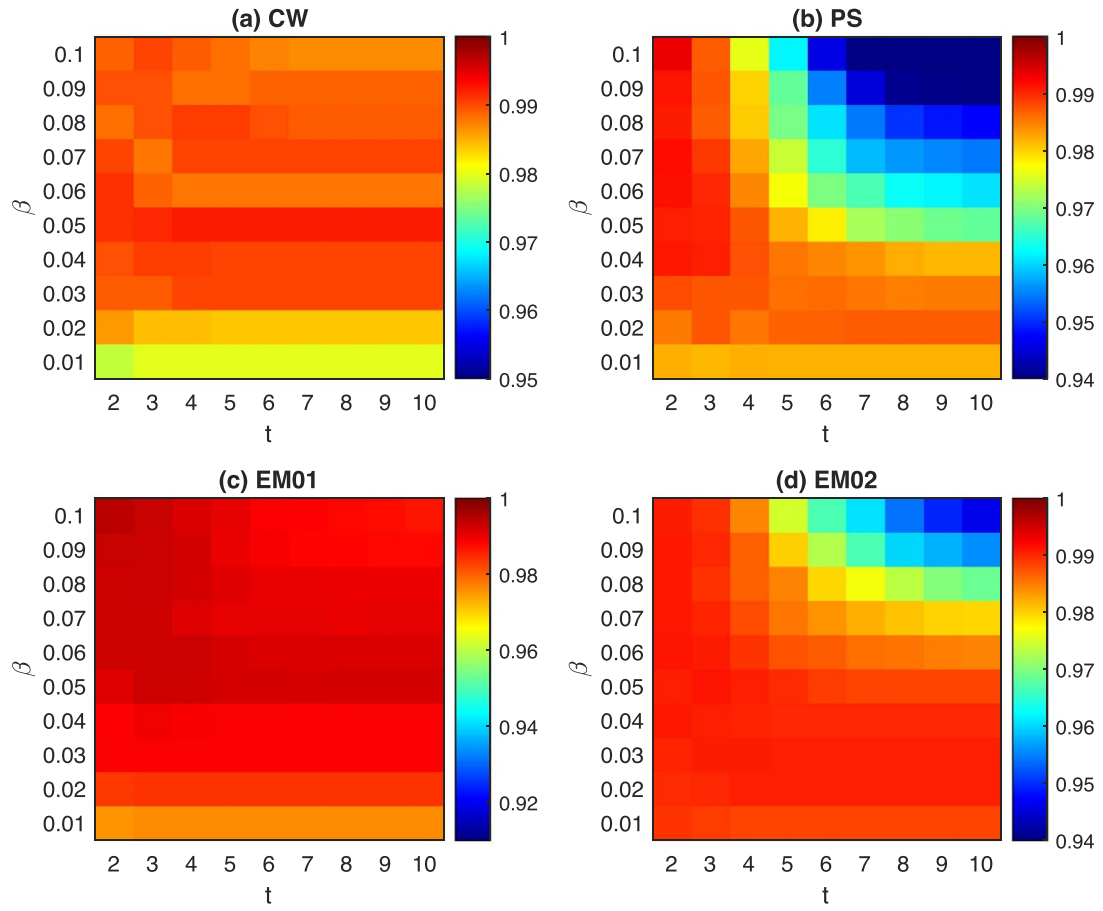

Figure 13. The accuracy of the DMP method in evaluating nodes' spreading influences according to the standard SIR model in four real temporal networks, quantified by Kendall's Tau. The spreading rate β varies from 0.01 to 0.10. Each data point is obtained by averaging over 10^5 independent runs.

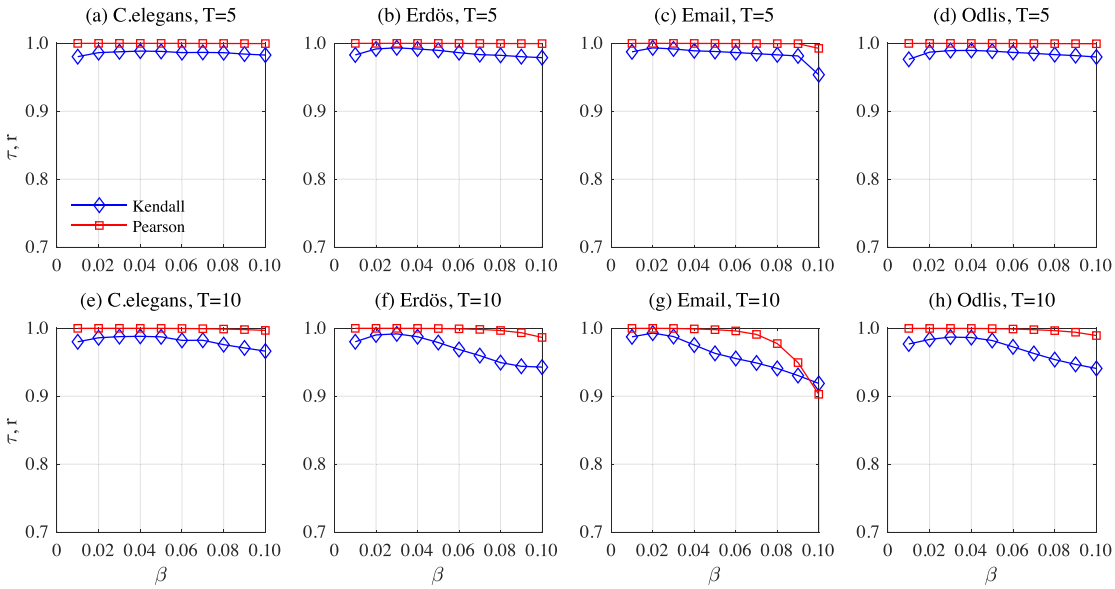

Figure 14. The accuracy of the DMP method in evaluating the spreading influence of nodes according to the standard SIR model in four real networks, quantified by Kendall's Tau τ and Pearson correlation coefficient r. The spreading rate β varies from 0.01 to 0.10 and the total time steps $T = 5$ and 10. Each data point is obtained by averaging over 10^5 independent runs.

5. Discussions

The essential question of ranking the node spreading influence is how to estimate the outbreak size of the initial spreader [53–56]. To answer this question, one needs to fix the nonlinear coupling issue during the spreading process. In this paper, we present a new method to evaluate the spreading scope from the perspective of the Markov chain process, namely the DMP. This method solves the problem of nonlinear coupling by adjusting the state transition matrix, in which the elements of the matrix are the probabilities of nodes in a susceptible state. The simulation results show that the DMP method could estimate the nodes' spreading scope at each time step accurately in a directed network without any loops. Compared with the DS centrality, the DMP method can evaluate the outbreak size more accurately on the model and empirical networks.

Furthermore, according to the empirical results on four real networks, for both the SIR and SIS model, the ranking list generated by the DMP method is very close to the ones of the simulation results, especially when the spreading rate and time step are small. It indicates that the DMP method can recover the characteristics of the spreading dynamics on the network.

The DMP method could also be used to evaluate the nodes' spreading scope generated by multi-spreaders. Our simulation results indicate that when there exist multiple spreaders, the DMP method significantly outperforms the DS [34] centrality with increasing values of spreaders and time steps. The key to identifying multiple influential spreaders is to solve the overlap problem [57], which is the non-linear couple problem during the spreading process. Given the fact that the DMP method is non-linear, it will be able to identify multiple influential spreaders by using the greedy approach [58–60]. Moreover, the DMP method is also suitable for detecting influential nodes in temporal networks.

Compared to the other methods in evaluating the outbreak size of the spreading dynamics, e.g. the Message-Passing Techniques [58] and the Percolation [61], the DMP method evaluates the spreading scope from the perspective of Markov process, and provides a general framework for ranking node spreading influence. Therefore, it can be extended and applied in modeling many other important dynamics such as Ising model [62, 63], Boolean dynamics [64], voter model [65], synchronization [66], and so on.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

J Lin and C J T acknowledge support from the Swiss National Science Foundation grant #200021_182659. This work is also partially supported by the China Scholarship Council, the Humanities and Social Sciences Project of the Ministry of Education of China, the Natural Science Foundation of Education Department of Jiangsu Province under Contract No. 20KJA520008, Six talent peaks project in Jiangsu Province (Grant No. XYDXX-034) and the National Natural Science Foundation of China (Grant No. 72171150, 72032003, and 6173248).

Authors contributions

Jianhong Lin, Bo-Lun Chen and Zhao Yang performed the analysis. Jianhong Lin and Claudio J Tessone designed the research. All authors wrote, reviewed and approved the manuscript.

References

[1] Tao Z, Zhongqian F and Binghong W 2006 Epidemic dynamics on complex networks *Prog. Nat. Sci.* **16** 452–7
[2] Keeling M J and Rohani P 2008 *Modeling Infectious Diseases in Humans and Animals* (Princeton, NJ: Princeton University Press)
[3] Meng X, Cai Z, Si S and Duan D 2021 Analysis of epidemic vaccination strategies on heterogeneous networks: based on SEIRV model and evolutionary game *Appl. Math. Comput.* **403** 126172
[4] Meng X, Han S, Wu L, Si S and Cai Z 2022 Analysis of epidemic vaccination strategies by node importance and evolutionary game on complex networks *Reliab. Eng. Syst. Saf.* **219** 108256
[5] Kephart J O, Sorkin G B, Chess D M and White S R 1997 Fighting computer viruses *Sci. Am.* **277** 88–93
[6] Motter A E 2004 Cascade control and defense in complex networks *Phys. Rev. Lett.* **93** 098701
[7] Li D, Fu B, Wang Y, Lu G, Berezin Y, Stanley H E and Havlin S 2015 Percolation transition in dynamical traffic network with evolving critical bottlenecks *Proc. Natl Acad. Sci.* **112** 669–72
[8] Lin J-H, Primicerio K, Squartini T, Decker C and Tessone C J 2020 Lightning network: a second path towards centralisation of the bitcoin economy *New J. Phys.* **22** 083022
[9] Lin J-H, Marchese E, Tessone C J and Squartini T 2022 The weighted bitcoin lightning network *Chaos Solitons Fractals* **164** 112620

[10] Campajola C, Cristodaro R, De Collibus F M, Yan T, Vallarano N, Tessone C J 2022 The evolution of centralisation on cryptocurrency platforms (arXiv:2206.05081)
[11] Pastor-Satorras R and Vespignani A 2002 Immunization of complex networks *Phys. Rev.* E **65** 036104
[12] Cohen R, Havlin S and Ben-Avraham D 2003 Efficient immunization strategies for computer networks and populations *Phys. Rev. Lett.* **91** 247901
[13] Chen B-L, Jiang W-X, Yu Y-T, Zhou L and Tessone C J 2022 Graph embedding based ant colony optimization for negative influence propagation suppression under cost constraints *Swarm Evol. Comput.* **72** 101102
[14] Bodendorf F and Kaiser C 2009 Detecting opinion leaders and trends in online social networks *Proc. 2nd ACM Workshop on Social web Search and Mining* (ACM) pp 65–68
[15] Uehara M and Tsugawa S 2019 Analysis of the evolution of the influence of central nodes in a twitter social network *2019 IEEE 43rd Annual Computer Software and Applications Conf. (COMPSAC)* vol 1 (IEEE) pp 892–5
[16] Tanase R, Tessone C J and Algesheimer R 2018 Identification of influencers through the wisdom of crowds *PLoS One* **13** e0200109
[17] Sinatra R, Wang D, Deville P, Song C and Barabási A-L 2016 Quantifying the evolution of individual scientific impact *Science* **354** aaf5239
[18] Wang D, Song C and Barabási A-L 2013 Quantifying long-term scientific impact *Science* **342** 127–32
[19] Pei S and Makse H A 2013 Spreading dynamics in complex networks *J. Stat. Mech.* **2013** P12002
[20] Freeman L C 1978 Centrality in social networks conceptual clarification *Soc. Netw.* **1** 215–39
[21] Freeman L C 1977 A set of measures of centrality based on betweenness *Sociometry* **40** 35–41
[22] Sabidussi G 1966 The centrality index of a graph *Psychometrika* **31** 581–603
[23] Bonacich P and Lloyd P 2001 Eigenvector-like measures of centrality for asymmetric relations *Soc. Netw.* **23** 191–201
[24] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E and Makse H A 2010 Identification of influential spreaders in complex networks *Nat. Phys.* **6** 888–93
[25] Liu J-G, Ren Z-M and Guo Q 2013 Ranking the spreading influence in complex networks *Physica* A **392** 4154–9
[26] Ren Z-M, Zeng A, Chen D-B, Liao H and Liu J-G 2014 Iterative resource allocation for ranking spreaders in complex networks *Europhys. Lett.* **106** 48005
[27] Lin J-H, Guo Q, Dong W-Z, Tang L-Y and Liu J-G 2014 Identifying the node spreading influence with largest *k*-core values *Phys. Lett.* A **378** 3279–84
[28] Lü L, Zhou T, Zhang Q-M and Stanley H E 2016 The H-index of a network node and its relation to degree and coreness *Nat. Commun.* **7** 10168
[29] Alvarez-Socorro A, Herrera-Almarza G and González-Díaz L 2015 Eigencentrality based on dissimilarity measures reveals central nodes in complex networks *Sci. Rep.* **5** 17095
[30] Lawyer G 2015 Understanding the influence of all nodes in a network *Sci. Rep.* **5** 8665
[31] Zhou M, Tan J, Liao H, Wang Z and Mao R 2020 Dismantling complex networks based on the principal eigenvalue of the adjacency matrix *Chaos* **30** 083118
[32] Šikić M, Lančić A, Antulov-Fantulin N and Štefančić H 2013 Epidemic centrality—is there an underestimated epidemic impact of network peripheral nodes? *Eur. Phys. J.* B **86** 440
[33] Klemm K, Serrano M A, Eguíluz V M and San Miguel M 2012 A measure of individual role in collective dynamics *Sci. Rep.* **2** 292
[34] Liu J-G, Lin J-H, Guo Q and Zhou T 2016 Locating influential nodes via dynamics-sensitive centrality *Sci. Rep.* **6** 21380
[35] Toral R, Tessone C J and Lopes J V 2007 Collective effects induced by diversity in extended systems *Eur. Phys. J. Spec. Top.* **143** 59–67
[36] Hethcote H W 2000 The mathematics of infectious diseases *SIAM Rev.* **42** 599–653
[37] Pastor-Satorras R and Vespignani A 2001 Epidemic spreading in scale-free networks *Phys. Rev. Lett.* **86** 3200
[38] Ide K, Zamami R and Namatame A 2013 Diffusion centrality in interconnected networks *Proc. Comput. Sci.* **24** 227–38
[39] Chen B-L, Jiang W-X, Chen Y-X, Chen L, Wang R-J, Han S, Lin J-H and Zhang Y-C 2022 Influence blocking maximization on networks: models, methods and applications *Phys. Rep.* **976** 1–54
[40] Scholtes I, Wider N, Pfitzner R and Tessone C J 2014 Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks *Nat. Commun.* **5** 5024
[41] Pfitzner R, Scholtes I, Garas A, Tessone C J and Schweitzer F 2013 Betweenness preference: quantifying correlations in the topological dynamics of temporal networks *Phys. Rev. Lett.* **110** 198701
[42] Hom R A and Johnson C R 1985 *Matrix Analysis* (Cambridge: Cambridge University Express)
[43] DeGroot M H and Schervish M J 2012 (London: Pearson Education)
[44] Kendall M G 1938 A new measure of rank correlation *Biometrika* **30** 81–93
[45] Watts D J and Strogatz S H 1998 Collective dynamics of 'small-world' networks *Nature* **393** 440–2
[46] Guimera R, Danon L, Diaz-Guilera A, Giralt F and Arenas A 2003 Self-similar community structure in a network of human interactions *Phys. Rev.* E **68** 065103
[47] Génois M, Vestergaard C L, Fournet J, Panisson A, Bonmarin I and Barrat A 2015 Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers *Netw. Sci.* **3** 326–47
[48] Gemmetto V, Barrat A and Ciro C 2014 Mitigation of infectious disease at school: targeted class closure vs school closure *BMC Infect. Dis.* **14** 695
[49] Stehlé J *et al* 2011 High-resolution measurements of face-to-face contact patterns in a primary school *PLoS One* **6** e23176
[50] Paranjape A, Benson R A and Leskovec J 2017 Motifs in temporal networks *Proc. 10th ACM Int. Conf. on Web Search and Data Mining* pp 601–10
[51] De Domenico M, Lima A, Mougel P and Musolesi M 2013 The anatomy of a scientific Rumor *Sci. Rep.* **3** 1–9
[52] Chen D-B, Xiao R and Zeng A 2014 Predicting the evolution of spreading on complex networks *Sci. Rep.* **4** 1–6
[53] Chen D-B, Sun H-L, Tang Q, Tian S-Z and Xie M 2019 Identifying influential spreaders in complex networks by propagation probability dynamics *Chaos* **29** 033120
[54] Bucur D and Holme P 2019 Beyond ranking nodes: predicting epidemic outbreak sizes by network centralities (arXiv:1909.10021)
[55] Poux-Médard G, Pastor-Satorras R and Castellano C 2020 Influential spreaders for recurrent epidemics on networks *Phys. Rev. Research* **2** 023332
[56] De Bellis A, Pastor-Satorras R and Castellano C 2021 Influence of individual nodes for continuous-time susceptible-infected-susceptible dynamics on synthetic and real-world networks. *Phys. Rev. E* **104** 014306
[57] Wang X, Zhang X, Yi D and Zhao C 2017 Identifying influential spreaders in complex networks through local effective spreading paths *J. Stat. Mech.* **2017** 053402

[58] Altarelli F, Braunstein A, Dall'Asta L, Wakeling J R and Zecchina R 2014 Containing epidemic outbreaks by message-passing techniques *Phys. Rev.* X **4** 021024
[59] Morone F and Makse H A 2015 Influence maximization in complex networks through optimal percolation *Nature* **524** 65–68
[60] Guo L, Lin J-H, Guo Q and Liu J-G 2016 Identifying multiple influential spreaders in term of the distance-based coloring *Phys. Lett.* A **380** 837–42
[61] Hu Y, Ji S, Feng L and Jin Y 2015 Quantify and maximise global viral influence through local network information (arXiv:1509.03484)
[62] Dorogovtsev S N, Goltsev A V and Mendes J F 2008 Critical phenomena in complex networks *Rev. Mod. Phys.* **80** 1275
[63] Didier S, Lera S, Lin J and Wu K 2022 Non-normal interactions create socio-economic bubbles (arXiv:2205.08661)
[64] Kaufmann S 1993 *The Origins of Order: Self-organization and Selection in Evolution* (Oxford: Oxford University Press)
[65] Castellano C, Fortunato S and Loreto V 2009 Statistical physics of social dynamics *Rev. Mod. Phys.* **81** 591
[66] Arenas A, Díaz-Guilera A, Kurths J, Moreno Y and Zhou C 2008 Synchronization in complex networks *Phys. Rep.* **469** 93–153