

# SCIENTIFIC REPORTS



OPEN

## Locating influential nodes via dynamics-sensitive centrality

Jian-Guo Liu<sup>1,2</sup>, Jian-Hong Lin<sup>2</sup>, Qiang Guo<sup>2</sup> & Tao Zhou<sup>3</sup>

Received: 10 May 2015

Accepted: 22 January 2016

Published: 24 February 2016

**With great theoretical and practical significance, locating influential nodes of complex networks is a promising issue. In this paper, we present a dynamics-sensitive (DS) centrality by integrating topological features and dynamical properties. The DS centrality can be directly applied in locating influential spreaders. According to the empirical results on four real networks for both susceptible-infected-recovered (SIR) and susceptible-infected (SI) spreading models, the DS centrality is more accurate than degree,  $k$ -shell index and eigenvector centrality.**

Spreading dynamics represents many important processes in nature and society<sup>1,2</sup>, such as the propagation of computer viruses<sup>3</sup> and traffic congestion<sup>4</sup>, reaction diffusion<sup>5</sup>, spreading of infectious diseases<sup>6</sup> and cascading failures<sup>7</sup>. The estimation of nodes' spreading influences can help in hindering epidemics or accelerating innovation<sup>8</sup>, and similar methods can be further applied in identifying influential spreaders in social networks<sup>9</sup>, quantifying the influence of scientists and their publications<sup>10</sup>, evaluating the impacts of injection points in the diffusion of microfinance<sup>11</sup>, finding drug targets in directed pathway networks<sup>12</sup>, predicting essential proteins in protein interaction networks<sup>13</sup>, and so on.

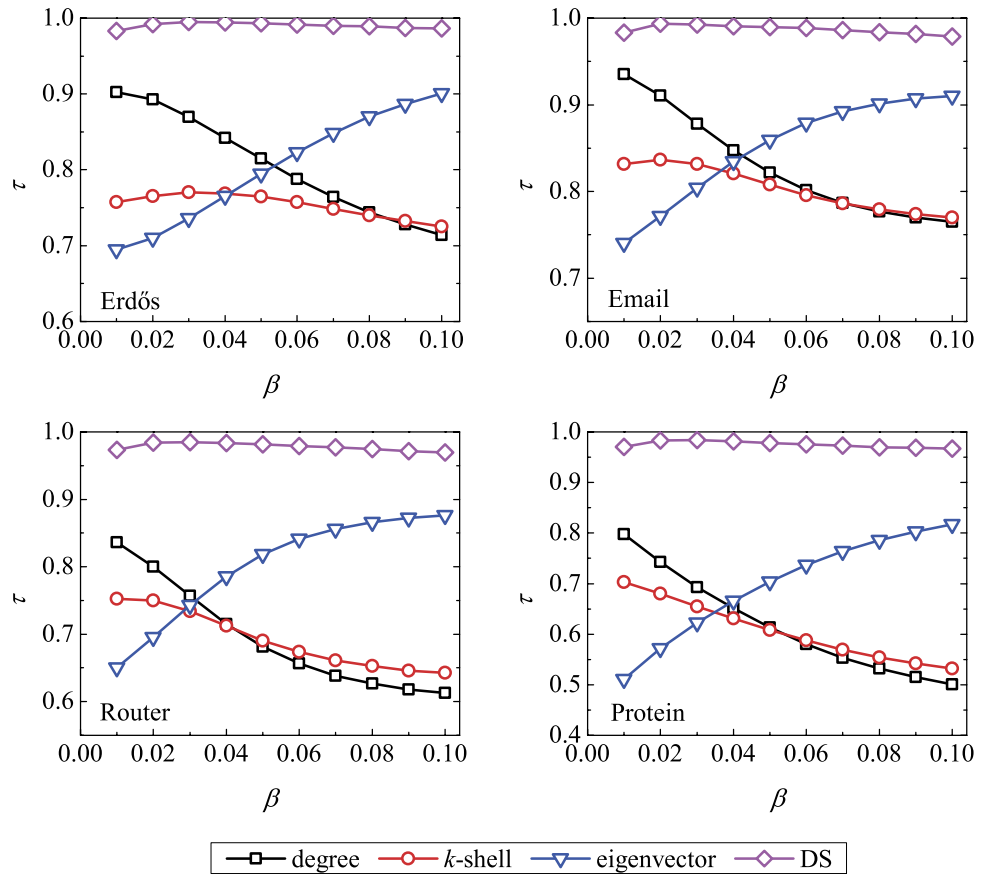
The significance of this issue triggers a variety of novel approaches in identifying influential spreaders in networks, which can be roughly categorized into three classes. Firstly, some scientists argued that the location of a node is more important than its immediate neighbors, and thus proposed  $k$ -shell index<sup>14,15</sup> and its variants<sup>16–19</sup> as indicators of spreading influences. Secondly, some scientists quantified a node's influence only accounting for its local surroundings<sup>20–22</sup>. Thirdly, some scientists evaluated nodes' influences according to the steady states of some introduced dynamical processes, such as random walk<sup>23,24</sup> and iterative refinement<sup>25</sup>.

The above-mentioned approaches only take into account the topological features, while recent experiments indicate that the performance of structural indices is very sensitive to the specific dynamics on networks<sup>26–28</sup>. For example, when the spreading rate is very small, the degree usually performs better than the eigenvector centrality<sup>29</sup> and  $k$ -shell index<sup>14</sup>, while when the infectivity is very high, the eigenvector centrality is the best one among the three (see Figs 1 and 2, with details shown later). To the best of our knowledge, there are few works taking into account the properties of the underlying spreading dynamics<sup>30–32</sup>. Encoding the interplay between topology and dynamics, Klemm *et al.*<sup>30</sup> suggested that the eigenvector centrality can be used in estimating nodes' dynamical influences in the susceptible-infected-recovered (SIR) spreading model (also called susceptible-infected-removed model)<sup>33</sup>. Li *et al.*<sup>31</sup> provided complementary explanation of the suitability of eigenvector centrality based on perturbation around the equilibrium of the epidemic dynamics and discussed the limitations of eigenvector centrality for homogeneous community networks. Both the above two works did not pay enough attention to the specific parameters in the spreading models, and thus their suggested index only works well in a limited range of the parameter space. Bauer and Lizier<sup>32</sup> proposed an approach based on counting the number of possible infection walks with different lengths to approximate the number of infected nodes generated by a given initially infected node. Their method is an effective one to predict the outbreak size over a wide range of spreading rate but less efficient due to the considerable computational cost. In addition, for the fundamental complexity in counting the number of paths connecting two nodes, their method could not be formulated in a compact analytical form.

In this paper, we describe the infectious probabilities of nodes by a matrix differential function that accounts both topological features and dynamical properties. Accordingly, we propose a dynamics-sensitive (DS) centrality to predict the outbreak size at given time step, which can be directly applied in quantifying the spreading influences of nodes. According to the empirical results on four real networks, for both the SIR model<sup>33</sup> and the susceptible-infected (SI) model<sup>34,35</sup>, the DS centrality can locate influential nodes accurately and performs

<sup>1</sup>Data Science and Cloud Service Centre, Shanghai University of Finance and Economics, Shanghai 200433, PR China.

<sup>2</sup>Research Center of Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, PR China. <sup>3</sup>Complex Lab, Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, PR China. Correspondence and requests for materials should be addressed to J.-G.L. (email: liujg004@ustc.edu.cn) or T.Z. (email: zhutou@ustc.edu)



**Figure 1.** The accuracy of four centrality measures in evaluating nodes' spreading influences according to the SIR model ( $\mu = 1$ ) in the four real networks, quantified by the Kendall's Tau. The spreading rate  $\beta$  varies from 0.01 to 0.10, and the time step is set as  $t = 5$ . Each data point is obtained by averaging over  $10^4$  independent runs.

particularly well in the early stage of spreading. The method proposed in this paper can be extended to other Markov processes on networks.

### Dynamics-Sensitive Centrality

A simple undirected connected 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 with node  $j$ , and  $a_{ij} = 0$  otherwise.  $\mathbf{A}$  is binary and symmetric with zeros along the main diagonal, and thus its eigenvalues are real and can be arrayed in a descending order as  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ . Since  $\mathbf{A}$  is a symmetric and real-valued matrix, it can be factorized as  $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ , where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ ,  $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n]$  and  $\mathbf{q}_i$  is the eigenvector of the eigenvalue  $\lambda_i$ .

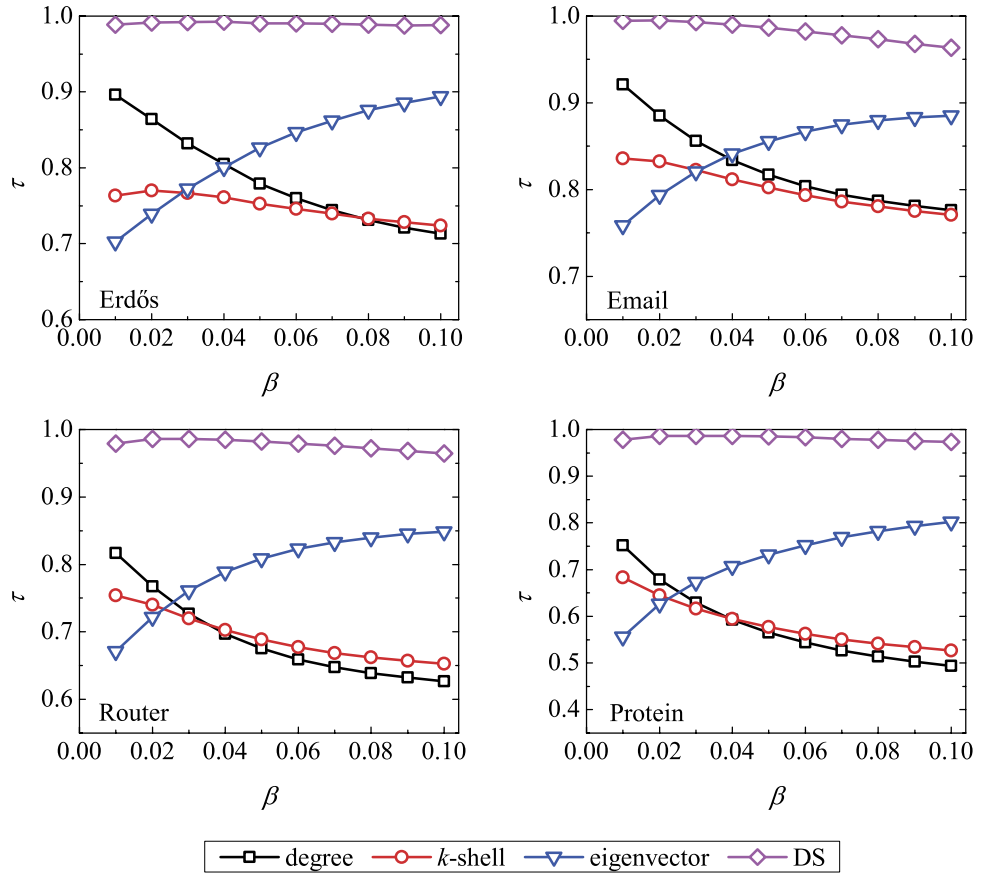
We consider a discrete-time spreading model where an infected node would infect its neighbors with spreading rate  $\beta$  and recover with recovering rate  $\mu$  (see **Materials and Methods** for details). Denote  $\mathbf{x}(t)$  ( $t > 0$ ) as an approximation of the cumulative probabilities that nodes are excited between time step 1 and  $t$ , and then  $\mathbf{x}(t) - \mathbf{x}(t - 1)$  ( $t > 1$ ) is approximated as the probabilities of nodes to be infected at time step  $t$ . If  $i$  is the only initially infected node, then  $x_i(0) = 1$  and  $x_{j \neq i}(0) = 0$ . In the first time step,  $\mathbf{x}(1) = \beta\mathbf{A}\mathbf{x}(0)$ , and for  $t > 1$ , we have (see the derivation in **Materials and Methods**)

$$\mathbf{x}(t) - \mathbf{x}(t - 1) = \beta\mathbf{A}[\beta\mathbf{A} + (1 - \mu)\mathbf{I}]^{t-1}\mathbf{x}(0), \tag{1}$$

where  $\mathbf{I}$  is the identity matrix. Denoting  $\mathbf{H} = \beta\mathbf{A} + (1 - \mu)\mathbf{I}$ , then  $\beta\mathbf{A}\mathbf{H}^{t-1}\mathbf{x}(0)$  represents the probabilities of nodes to be infected at time step  $t$ , and thus the cumulative probabilities of nodes to have been infected between time step 1 and  $t$  can be approximated as

$$\mathbf{x}(t) = \sum_{r=2}^t [\mathbf{x}(r) - \mathbf{x}(r - 1)] + \mathbf{x}(1) = \sum_{r=0}^{t-1} \beta\mathbf{A}\mathbf{H}^r\mathbf{x}(0). \tag{2}$$

We define  $S_i(t)$  to be the spreading influence of node  $i$  at time step  $t$ , which can be quantified by the sum of infected probabilities of all nodes, given  $i$  the initially infected seed. According to Eq. (2), the infected probabilities can be written as



**Figure 2.** The accuracy of four centrality measures in evaluating nodes' spreading influences according to the SI model ( $\mu = 0$ ) in the four real networks, quantified by the Kendall's Tau. The spreading rate  $\beta$  varies from 0.01 to 0.10, and the time step is set as  $t = 5$ . Each data point is obtained by averaging over  $10^4$  independent runs.

$$\mathbf{x}(t) = \sum_{r=0}^{t-1} \beta \mathbf{A} \mathbf{H}^r \mathbf{e}_i, \tag{3}$$

where  $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^T$  is an  $n \times 1$  vector with only the  $i$ th element being 1. As all elements other than the  $i$ th one of  $\mathbf{e}_i$  are zero,  $\mathbf{x}(t)$  is indeed the sum of all the  $i$ th columns of  $\beta \mathbf{A}$ ,  $\beta \mathbf{A} \mathbf{H}$ ,  $\dots$ ,  $\beta \mathbf{A} \mathbf{H}^{t-1}$ . Given  $\mathbf{x}(0) = \mathbf{e}_p$ ,  $S_i(t)$  is defined as the sum of all elements of  $\mathbf{x}(t)$ , which is equal to the sum of all elements in the  $i$ th columns of  $\beta \mathbf{A}$ ,  $\beta \mathbf{A} \mathbf{H}$ ,  $\dots$ ,  $\beta \mathbf{A} \mathbf{H}^{t-1}$ , as

$$S_i(t) = [(\beta \mathbf{A} + \beta \mathbf{A} \mathbf{H} + \dots + \beta \mathbf{A} \mathbf{H}^{t-1})^T \mathbf{L}]_i, \tag{4}$$

where  $\mathbf{L} = (1, 1, \dots, 1)^T$  is an  $n \times 1$  vector whose components are all 1. Obviously,  $\mathbf{A}^T = \mathbf{A}$ ,  $\mathbf{H}^T = \mathbf{H}$  and  $\mathbf{A} \mathbf{H} = \mathbf{H} \mathbf{A}$ , so the spreading influence of all nodes can be described by the vector

$$\mathbf{S}(t) = \sum_{r=0}^{t-1} \beta \mathbf{A} \mathbf{H}^r \mathbf{L}. \tag{5}$$

Notice that,  $\sum_{r=0}^{t-1} \beta \mathbf{A} \mathbf{H}^r \mathbf{L} = \sum_{r=0}^{t-1} \beta \mathbf{A} \mathbf{H}^r (\sum_{i=1}^n \mathbf{e}_i) = \sum_{i=1}^n \sum_{r=0}^{t-1} \beta \mathbf{A} \mathbf{H}^r \mathbf{e}_i$ , and  $\sum_{r=0}^{t-1} \beta \mathbf{A} \mathbf{H}^r \mathbf{e}_i$  is the infected probabilities of all nodes given node  $i$  the only initially infected seed according to Eq. (2), so  $\mathbf{S}(t)$  can also be roughly explained as the sum of infected probabilities over the  $n$  cases with every node being the infected seed once. This relationship shows an underlying symmetry, that is, in an undirected network, the node having higher influence is also the one apt to be infected. The readers are warned that such conclusion is not mathematically rigorous since we have ignored the complicated entanglement by allowing the elements of  $\mathbf{x}(t)$  being larger than 1.

The eigenvectors of  $\mathbf{H}$  are the same to the ones of  $\mathbf{A}$  and  $\beta \lambda_i + 1 - \mu$  is the  $i$ th eigenvalue of  $\mathbf{H}$ , corresponding to  $\mathbf{q}_i$ . When  $\beta \lambda_1 + 1 - \mu < 1$ , i.e.  $\beta / \mu < 1 / \lambda_1$  (for the case  $\mu \neq 0$ ),  $\mathbf{H}^t \mathbf{L}$  could converge to null vector when  $t \rightarrow \infty$  and  $\mathbf{S}(t)$  could be written by the following way

$$\mathbf{S}(t) = \beta \mathbf{A}(\mathbf{I} - \mathbf{H})^{-1} \mathbf{L} = [(\beta/\mu) \mathbf{A} + (\beta/\mu)^2 \mathbf{A}^2 + \dots + (\beta/\mu)^t \mathbf{A}^t] \mathbf{L}. \quad (6)$$

For simplicity, in the SIR model, we set  $\mu = 1$ , and then

$$\mathbf{S}(t) = (\beta \mathbf{A} + \beta^2 \mathbf{A}^2 + \dots + \beta^t \mathbf{A}^t) \mathbf{L}, \quad (7)$$

where  $(\mathbf{A}^t \mathbf{L})_i$  counts the total number of walks of length  $t$  from node  $i$  to all nodes in the network, weighted by  $\beta^t$  that decays as the increase of the length  $t$ . As  $\mathbf{S}(t)$  quantifies nodes' spreading influences, we call it dynamics-sensitive (DS) centrality, where the term *dynamics-sensitive* emphasizes the fact that  $\mathbf{S}(t)$  is determined not only by the network structure (i.e.,  $\mathbf{A}$ ), but also the dynamical parameters (i.e.,  $\beta$  and  $t$ ). In particular, when  $t = 1$ , the initially infected node only has the chance to infect its neighbors and  $S_i(1) = (\beta \mathbf{A} \mathbf{L})_i$  with  $(\mathbf{A} \mathbf{L})_i$  being exactly the degree of node  $i$ . When  $\mu = 0$  (corresponding to the SI model) or  $\beta \geq 1/\lambda_1$ ,  $\mathbf{S}(t)$  would be infinite when  $t \rightarrow \infty$ , which could not reflect the spreading influences. In fact, there are two underlying assumptions. Firstly, the DS centrality calculates the probabilities of the nodes to be infected at time step  $t$  by linear coupling instead of nonlinear coupling. For example, if a susceptible node has  $m$  infected neighbors, the probability of the node to be infected is approximated as  $m\beta$  by the DS centrality, instead of  $1 - (1 - \beta)^m$ , which is the exact probability of the node to be infected. Secondly, the DS centrality calculates the probabilities of nodes to be infected under the assumption that the nodes are susceptible at time steps  $t$ . Notice that, our main goal is to find out the ranking of spreading influences of nodes, namely to identify influential nodes. As every node's infected probability is overestimated, the ranking may be less affected. Thus far, we are still not aware of the impacts on the ranking, while fortunately, as later indicated by the extensive numerical results, the DS centrality performs much better than other well-known indices for a very broad ranges of  $\beta$  and  $t$  that cover most practical scenarios.

## Results

We test the performance of DS centrality in evaluating the nodes' spreading influences according to the SIR model and SI model, with varying spreading rate  $\beta$ . Four real networks, including a scientific collaboration network, an email communication network, the Internet at the router level and a protein-protein interaction network, are used for the empirical analysis (see data description in **Materials and Methods**), and three well-known indices, including degree,  $k$ -shell index and eigenvector centrality, are used as benchmark methods for comparison (see **Materials and Methods** for the definitions of those indices). Given the time step  $t$ , the spreading influence of an arbitrary node  $i$  is quantified by the number of infected nodes (for SI model) or the number of infected and recovered nodes (for SIR model) at  $t$ , where the spreading process starts with only node  $i$  being initially infected. Here we use Kendall's Tau  $\tau^{36}$  to measure the correlation between nodes' spreading influences and the considered centrality measure, where  $\tau$  is in the range  $[-1, 1]$  and the larger  $\tau$  corresponds to the better performance (see **Materials and Methods** for the definition of  $\tau$ ).

As shown in Fig. 1, the Kendall's Tau  $\tau$  for the DS centrality is between 0.968 and 0.995 for  $\beta \in [0.01, 0.1]$ , indicating that the ranking lists generated by the DS centrality and the real SIR spreading process are highly identical to each other. In comparison, the DS centrality performs better than degree,  $k$ -shell index and eigenvector centrality since the DS centrality takes into account one more parameter  $\beta$ . As shown in Fig. 2, similar results are also observed for the SI model where the DS centrality performs better than others. The results for larger  $\beta$  and  $t$  are respectively shown in Figs S1 and S2 of Supplementary Information, where the DS centrality still could locate influential nodes accurately. Meanwhile, Fig. S3 shows that the DS centrality also performs much better than other compared indices in evaluating the nodes' spreading influences for  $\mu = 0.3$  and  $0.7$  in the SIR model.

Since  $\mathbf{A}$  is a symmetric, real-valued matrix, the DS centrality  $\mathbf{S}(t)$  can be written in the following way by decomposing  $\mathbf{A}$

$$S_i(t) = m_1 q_{1i} \sum_{j=1}^n q_{1j} + \sum_{r=2}^n m_r q_{ri} \sum_{j=1}^n q_{rj}, \quad (8)$$

where  $m_r = \beta \lambda_r [1 - (\beta \lambda_r + 1 - \mu)^t] (\mu - \beta \lambda_r)^{-1}$  for  $1 \leq r \leq n$ . Rewriting Eq. (8) into

$$\frac{S_i(t)}{m_1} = q_{1i} \sum_{j=1}^n q_{1j} + \sum_{r=2}^n \frac{m_r}{m_1} q_{ri} \sum_{j=1}^n q_{rj}. \quad (9)$$

With the increase of  $t$  and  $\beta$ ,  $\frac{m_r}{m_1}$  will converge to 0, and thus the ranking lists generated by  $\mathbf{S}(t)$  will be identical to  $\mathbf{q}_1$ , which is exactly the same to the eigenvector centrality. This relationship is in accordance with the results presented in Figs S1 and S2, where the difference between the eigenvector centrality and DS centrality gets smaller as the increase of  $\beta$  and  $t$ .

## Conclusion and Discussions

Estimating the spreading influences and then identifying influential nodes are fundamental task before any regulation on the spreading process. For such task, most known works only took into account the topological information<sup>8</sup>. Recently, Aral and Walker<sup>37</sup> showed that the attributes of nodes are highly correlated with nodes' influences and tendencies to be influenced. In this paper, in addition to the topological information, we get down to the underlying spreading dynamics and propose a dynamics-sensitive (DS) centrality, which is a kind of weighted sum of walks ending at the target node, where both the spreading rate and spreading time are accounted in the weighting function. The DS centrality can be directly applied in quantifying the spreading influences of nodes.

Network	$n$	$e$	$\langle k \rangle$	$1/\lambda_1$
Erdős	454	1313	5.784	0.079
Email	1133	5451	9.622	0.048
Router	2114	6632	6.274	0.036
Protein	2783	6007	4.317	0.063

**Table 1.** Basic statistical features of Erdős, Email, Router and Protein networks, including the number of nodes  $n$ , the number of the edges  $e$ , the average degree  $\langle k \rangle$  and the reciprocal of the largest eigenvalue  $1/\lambda_1$ .

According to the empirical analyses of the SIR model and SI model at given time step on four real networks, the DS centrality can locate influential nodes much more accurately than degree,  $k$ -shell index and eigenvector centrality. In fact, it is unfair to directly compare the DS centrality with the above structure-based indices since the DS centrality takes into account one more parameter  $\beta$ , while the present result is still significant as it gives an elegant way to account the specific dynamics and shows the remarkable improvement by introducing the parameter  $\beta$ , which may be expected but not yet well demonstrated before this paper.

The DS centrality performs particularly well in the early stage of spreading, which provides a powerful tool in early detection of potential super-spreaders for epidemic control. Notice that, we did not consider the asymptotical case with  $t \rightarrow \infty$ , since after sufficiently long spreading time, all nodes in the network would be infected under large spreading rate, especially in the SI model. In such case, different nodes spreading influences are not distinguishable.

The DS centrality tells us an often ignored fact that the most influential nodes are dependent not only on the network topology but also on the spreading dynamics. Given different models and parameters, the relative influences of nodes are also different. Roughly speaking, if the spreading rate is small, we can focus on the close neighborhood of a node since it is not easy to form a long spreading pathway (i.e.,  $\beta^t$  decays very fast as the increase of  $t$  when  $\beta$  is small) while if the spreading rate is high, the global topology should be considered. A clear limitation of this work is that before calculating the DS centrality, we have to know the spreading rate that is usually a hidden parameter. This parameter can be effectively estimated according to the early spreading process<sup>38</sup> and then we can calculate the DS centralities by varying the spreading rates over the estimated range and see which nodes are the most influential ones in average. In addition to the identification of the most influential node, a related but more challenging problem is to find the most influential sets of nodes<sup>39</sup>. The DS centrality could not be directly extended to solve this problem, however, similar to the present idea, the dynamical information is very necessary in dealing with this problem<sup>40</sup>.

Some other centralities related to specific dynamical processes have also been proposed recently, including routing centrality<sup>41</sup>, epidemic centrality<sup>42</sup>, diffusion centrality<sup>43</sup>, percolation centrality<sup>44</sup> and game centrality<sup>45</sup>. Comparing with these centralities, similar to the works by Klemm *et al.*<sup>30,46,47</sup>, this paper provides a more general framework that could deal with other Markov processes and thus can be extended and applied in many other important dynamics, such as the Ising model<sup>48</sup>, Boolean dynamics<sup>49</sup>, voter model<sup>50</sup>, synchronization<sup>51</sup>, and so on. For example, Zhou *et al.*<sup>35</sup> considered a variant SIR spreading model where an infected node can only contact one neighbor in one time step. In this case, the adjacency matrix  $\mathbf{A}$  in the DS centrality could be replaced by  $\mathbf{D}^{-1}\mathbf{A}$ , where  $\mathbf{D}$  is an  $n \times n$  matrix where  $i$ th main diagonal element is the degree of node  $i$  and other elements are 0. Then the DS centrality can be extended to this spreading model. Furthermore, the DS centrality can also be directly extended to asymmetrical networks and weighted networks. We hope this work could highlight the significant role of underlying dynamics in quantifying the individual nodes' importance, and then the difference between lists of critical nodes for different dynamical processes could be considered as their distinguished properties.

## Materials and Methods

**Derivation of Eq. (1).** The probabilities of nodes to be infected at time step  $t = 2$  can be approximated as

$$\mathbf{x}(2) - \mathbf{x}(1) = \beta\mathbf{A}[\mathbf{x}(1) + (1 - \mu)\mathbf{x}(0)] = \beta\mathbf{A}[\beta\mathbf{A} + (1 - \mu)\mathbf{I}]\mathbf{x}(0), \quad (10)$$

where  $\mathbf{x}(1) + (1 - \mu)\mathbf{x}(0)$  is the probabilities that nodes still have the ability to infect the susceptible nodes at time step 2. We assume that when  $t \leq p$ ,  $\mathbf{x}(p) - \mathbf{x}(p - 1) = \beta\mathbf{A}[\beta\mathbf{A} + (1 - \mu)\mathbf{I}]^{p-1}\mathbf{x}(0)$ , then for  $t = p + 1$ , we have

$$\mathbf{x}(p + 1) - \mathbf{x}(p) = \beta\mathbf{A} \left\{ \sum_{r=0}^{p-2} (1 - \mu)^r [\mathbf{x}(p - r) - \mathbf{x}(p - r - 1)] + (1 - \mu)^{p-1}\mathbf{x}(1) + (1 - \mu)^p\mathbf{x}(0) \right\}, \quad (11)$$

where  $\sum_{r=0}^{p-2} (1 - \mu)^r [\mathbf{x}(p - r) - \mathbf{x}(p - r - 1)] + (1 - \mu)^{p-1}\mathbf{x}(1) + (1 - \mu)^p\mathbf{x}(0)$  is the probabilities that nodes still have the ability to infect the susceptible nodes at time step  $p + 1$ . Therefore, we can get that

$$\begin{aligned}
 \mathbf{x}(p+1) - \mathbf{x}(p) &= \beta \mathbf{A} \left\{ \sum_{r=0}^{p-2} (1-\mu)^r \beta \mathbf{A} [\beta \mathbf{A} + (1-\mu) \mathbf{I}]^{p-r-1} \right. \\
 &\quad \left. + (1-\mu)^{p-1} [\beta \mathbf{A} + (1-\mu) \mathbf{I}] \right\} \mathbf{x}(0) \\
 &= \beta \mathbf{A} \left\{ \sum_{r=0}^{p-3} (1-\mu)^r \beta \mathbf{A} [\beta \mathbf{A} + (1-\mu) \mathbf{I}]^{p-r-1} \right. \\
 &\quad \left. + (1-\mu)^{p-2} \beta \mathbf{A} [\beta \mathbf{A} + (1-\mu) \mathbf{I}] \right. \\
 &\quad \left. + (1-\mu)^{p-1} [\beta \mathbf{A} + (1-\mu) \mathbf{I}] \right\} \mathbf{x}(0) \\
 &= \beta \mathbf{A} \left\{ \sum_{r=0}^{p-3} (1-\mu)^r \beta \mathbf{A} [\beta \mathbf{A} + (1-\mu) \mathbf{I}]^{p-r-1} \right. \\
 &\quad \left. + (1-\mu)^{p-2} [\beta \mathbf{A} + (1-\mu) \mathbf{I}]^2 \right\} \mathbf{x}(0) \tag{12} \\
 &\dots \\
 &= \beta \mathbf{A} \{ \beta \mathbf{A} [\beta \mathbf{A} + (1-\mu) \mathbf{I}]^{p-1} + (1-\mu) [\beta \mathbf{A} + (1-\mu) \mathbf{I}]^{p-1} \} \mathbf{x}(0) \\
 &= \beta \mathbf{A} [\beta \mathbf{A} + (1-\mu) \mathbf{I}]^p \mathbf{x}(0).
 \end{aligned}$$

Therefore, according to the mathematical induction, Eq. (1) is established.

**Spreading Model.** Here we apply the susceptible-infected-recovered (SIR) model (also called the susceptible-infected-removed model)<sup>33</sup>. In the SIR model, there are three kinds of individuals: (i) susceptible individuals that could be infected, (ii) infected individuals having been infected and being able to infect susceptible individuals, and (iii) recovered individuals that have been recovered and will never be infected again. In this paper, the spreading process starts with only one seed node being infected initially, and all other nodes are initially susceptible. At each time step, each infected node makes contact with its neighbors and each susceptible neighbor is infected with a probability  $\beta$ . Then each infected node enters the recovered state with a probability  $\mu$ . In the main text we set  $\mu = 1$ , and for other values of  $\mu$ , the results are very similar (see SI). In the standard SI model, nodes can only be susceptible or infected, corresponding to the case with  $\mu = 0$ . In this paper, the model, analysis and simulations are all based on the discrete-time dynamics.

**Benchmark Methods.** The degree of an arbitrary node  $i$  is defined as the number of its neighbors, namely

$$k_i = \sum_{j=1}^n a_{ij}, \tag{13}$$

where  $a_{ij}$  is the element of matrix  $\mathbf{A}$ . Degree centrality is widely applied for its simplicity and low computational cost, which works especially well in evaluating nodes' spreading influences when the spreading rate is small.

The main idea of eigenvector centrality is that a node's importance is not only determined by itself, but also affected by its neighbors' importance<sup>29</sup>. Accordingly, eigenvector centrality of node  $i$ ,  $v_i$ , is defined as

$$v_i = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} v_j, \tag{14}$$

where  $\lambda$  is a constant. Obviously, Eq. (13) can be written in a compact form as

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \tag{15}$$

where  $\mathbf{v} = (v_1, v_2, \dots, v_n)^T$ . That is to say,  $\mathbf{v}$  is the eigenvector of the adjacency matrix  $\mathbf{A}$  and  $\lambda$  is the corresponding eigenvalue. Since the considered network is a simple undirected connected network and thus the adjacency matrix  $\mathbf{A}$  is irreducible. According to Perron-Frobenius Theorem<sup>52</sup>, the elements in the leading eigenvector are strictly positive. Since the influences of nodes should be strictly positive,  $\mathbf{v}$  must be the leading eigenvector corresponding to the largest eigenvalue of  $\mathbf{A}$ , say  $\mathbf{v} = \mathbf{q}_1$ .

Kitsak *et al.*<sup>14</sup> argued that  $k$ -shell index (i.e., coreness) is a better index than degree to locate the influential nodes. The  $k$ -shell can be obtained by the so-called  $k$ -core decomposition<sup>53</sup>. The  $k$ -core decomposition process is initiated by removing all nodes with degree  $k = 1$ . This causes new nodes with degree  $k \leq 1$  to appear. These are also removed and the process is continued until all remaining nodes are of degree  $k > 1$ . The removed nodes (together with associated links) form the 1-shell, and their  $k$ -shell indices are all one. We next repeat this pruning process for the nodes of degree  $k = 2$  to extract the 2-shell, that is, in each step the nodes with degree  $k \leq 2$  are removed. We continue with the process until we have identified all higher-layer shells and all network nodes have been removed. Then each node  $i$  is assigned a  $k$ -shell index  $c_i$ .

**Kendall's Tau.** For each node  $i$ , we denote  $y_i$  as its spreading influence and  $z_i$  the target centrality measure (e.g., degree,  $k$ -shell index, eigenvector centrality and DS centrality), the accuracy of the target centrality in evaluating nodes' spreading influences can be quantified by the Kendall's Tau<sup>36</sup>, as

$$\tau = \frac{2}{n(n-1)} \sum_{i < j} \text{sgn}[(y_i - y_j)(z_i - z_j)], \quad (16)$$

where  $\text{sgn}(y)$  is a piecewise function, when  $y > 0$ ,  $\text{sgn}(y) = +1$ ;  $y < 0$ ,  $\text{sgn}(y) = -1$ ; when  $y = 0$ ,  $\text{sgn}(y) = 0$ .  $\tau$  measures the correlation between two ranking lists, whose value is in the range  $[-1, 1]$  and the larger  $\tau$  corresponds to the better performance.

## Data Description

Four real networks are studied in this paper as follows. (i) Erdős, a scientific collaboration network, where nodes are scientists and edges represent the co-authorships. The data set can be freely downloaded from the web site <http://wwwp.oakland.edu/enp/thedata/>. (ii) Email<sup>54</sup>, which is the email communication network of University Rovira i Virgili (URV) of Spain, involving faculty members, researchers, technicians, managers, administrators, and graduate students. (iii) Router<sup>55</sup>, the Internet at the router level, where each node represents a router and an edge represents a connection between two routers. (iv) Protein<sup>56</sup>, an initial version of a proteome-scale map of human binary protein-protein interaction. Basic statistical properties of the above four networks are presented in Table 1.

## References

- Zhou, T., Fu, Z. Q. & Wang, B. H. Epidemic dynamics on complex networks. *Porg. Nat. Sci.* **16**, 452 (2006).
- Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).
- Kephart, J. O., Sorkin, G. B., Chess, D. M. & White, S. R. Fighting computer viruses. *Sci. Am.* **277**, 56 (1997).
- Li, D. *et al.* Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proc. Nat. Acad. Sci. USA* **112**, 69 (2015).
- Colizza, V., Pastor-Satorras, R. & Vespignani, A. Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.* **3**, 276 (2007).
- Keeling, M. J. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton (2008).
- Motter, A. E. Cascade control and defense in complex networks. *Phys. Rev. Lett.* **93**, 098701 (2004).
- Pei, S. & Makse, H. A. Spreading dynamics in complex networks. *Journal of Statistical Mechanics: Theory and Experiment* **2013**, P12002 (2013).
- González-Bailón, S., Borge-Holthoefer, J., Rivero, A. & Moreno, Y. The dynamics of protest recruitment through an online network. *Sci. Rep.* **1**, 197 (2011).
- Zhou, Y. B., Lü, L. & Li, M. Quantifying the influence of scientists and their publications: distinguishing between prestige and popularity. *New J. Phys.* **14**, 033033 (2012).
- Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. The diffusion of microfinance. *Science* **341**, 1236498 (2013).
- Csermely, P., Korcsmáros, T., Kiss, H. J., London, G. & Nussinov, R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* **138**, 333 (2013).
- Li, M., Zhang, H., Wang, J. X. & Pan, Y. A new essential protein discovery method based on the integration of protein-protein interaction and gene expression data. *BMC Syst. Biol.* **6**, 15 (2012).
- Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888 (2010).
- Castellano, C. & Pastor-Satorras, R. Competing activation mechanisms in epidemics on networks. *Sci. Rep.* **2**, 371 (2012).
- Zeng, A. & Zhang, C. J. Ranking spreaders by decomposing complex networks. *Phys. Lett. A* **377**, 1031 (2013).
- Liu, J. G., Ren, Z. M. & Guo, Q. Ranking the spreading influence in complex networks. *Physica A* **392**, 4154 (2013).
- Lin, J. H., Guo, Q., Dong, W. Z., Tang, L. Y. & Liu, J. G. Identifying the node spreading influence with largest  $k$ -core values. *Phys. Lett. A* **378**, 3279 (2014).
- Liu, Y., Tang, M., Zhou, T. & Do, Y. Improving the accuracy of the  $k$ -shell method by removing redundant links: From a perspective of spreading dynamics. *Sci. Rep.* **5**, 13172 (2015).
- Chen, D. B., Lü, L., Shang, M. S., Zhang, Y. C. & Zhou, T. Identifying influential nodes in complex networks. *Physica A* **391**, 1777 (2012).
- Chen, D. B., Gao, H., Lü, L. & Zhou, T. Identifying influential nodes in large-scale directed networks: the role of clustering. *PLoS ONE* **8**, e77455 (2013).
- Pei, S., Muchnik, L., Andrade, J. S., Jr., Zheng, Z. & Makse, H. A. Searching for superspreaders of information in real-world social media. *Sci. Rep.* **4**, 5547 (2014).
- Lü, L., Yeung, C. H., Zhang, Y. C. & Zhou, T. Leaders in social networks, the delicious case. *PLoS ONE* **6**, e21202 (2011).
- Li, Q., Zhou, T., Lü, L. & Chen, D. B. Identifying influential spreaders by weighted LeaderRank. *Physica A* **404**, 47 (2014).
- Ren, Z. M., Zeng, A., Chen, D. B., Liao, H. & Liu, J. G. Iterative resource allocation for ranking spreaders in complex networks. *EPL* **106**, 48005 (2014).
- Borge-Holthoefer, J., Rivero, A. & Moreno, Y. Locating privileged spreaders on an online social network. *Phys. Rev. E* **85**, 066123 (2012).
- Borge-Holthoefer, J. & Moreno, Y. Absence of influential spreaders in rumor dynamics. *Phys. Rev. E* **85**, 026116 (2012).
- Liu, Y., Tang, M., Zhou, T. & Do, Y. Core-like groups resulting in invalidation of  $k$ -shell decomposition analysis. *Sci. Rep.* **5**, 9602 (2015).
- Borgatti, S. P. Centrality and network flow. *Soc. Netw.* **27**, 55 (2005).
- Klemm, K., Serrano, M. A., Eguiluz, V. M. & San Miguel, M. A measure of individual role in collective dynamics. *Sci. Rep.* **2**, 292 (2012).
- Li, P., Zhang, J., Xu, X. K. & Small, M. Dynamical influence of node revisited: a markov chain analysis of epidemic process on networks. *Chin. Phys. Lett.* **29**, 048903 (2012).
- Bauer, F. & Lizier, J. T. Identifying influential spreaders and efficiently estimating infection numbers in epidemic models: a walk counting approach. *EPL* **99**, 68007 (2012).
- Hethcote, H. W. The mathematics of infectious diseases. *SIAM Rev.* **42**, 599 (2000).
- Barthélemy, M., Barrat, A., Pastor-Satorras, R. & Vespignani, A. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.* **92**, 178701 (2004).

35. Zhou, T., Liu, J. G., Bai, W. J., Chen, G. & Wang, B. H. Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity. *Phys. Rev. E* **74**, 056109 (2006).
36. Kendall, M. G. A new measure of rank correlation. *Biometrika* **30**, 81 (1938).
37. Aral, S. & Walker, D. Identifying influential and susceptible members of social networks. *Science* **68**, 337 (2012).
38. Chen, D. B., Xiao, R. & Zeng, A. Predicting the evolution of spreading on complex networks. *Sci. Rep.* **4**, 6108 (2014).
39. Kempe, D., Kleinberg, J. & Tardos, É. Maximizing the spread of influence through a social network. *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 137–146, AAAI Press, 2015).
40. Morone, F. & Makse, H. A. Influence maximization in complex networks through optimal percolation. *Nature* **524**, 65 (2015).
41. Dolev, S., Elovici, Y. & Puzis, R. Routing betweenness centrality. *J. ACM* **57**(4), 25 (2010).
42. Šikić, M., Lančić, A., Antulov-Fantulin, N. & Štefančić, H. Epidemic centrality – is there an underestimated epidemic impact of network peripheral nodes? *Eur. Phys. J. B* **86**, 440 (2013).
43. Ide, K., Zamami, R. & Namatame, A. Diffusion Centrality in Interconnected Networks. *Proc. Comput. Sci.* **24**, 227 (2013).
44. Piraveenan, M., Prokopenko, M. & Hossain, L. Percolation centrality: Quantifying graph-theoretic impact of nodes during percolation in networks. *PLoS One* **8**, e53095 (2013).
45. Simko, G. I. & Csermely, P. Nodes having a major influence to break cooperation define a novel centrality measure: game centrality. *PLoS One* **8**, e67159 (2013).
46. Ghanbarnejad, F. & Klemm, K. Impact of individual nodes in Boolean network dynamics. *EPL* **99**, 58006 (2012).
47. Klemm, K. Searchability of Central Nodes in Networks. *J. Stat. Phys.* **151**, 707 (2013).
48. Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. Critical phenomena in complex networks. *Rev. Mod. Phys.* **80**, 1275 (2008).
49. Kauffman, S. A. *The Origins of Order*. Oxford University Press, New York, 1993.
50. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591 (2009).
51. Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y. & Zhou, C. Synchronization in complex networks. *Phys. Rep.* **469** 93 (2008).
52. Horn, R. A. & Johnson, C. R. *Matrix Analysis*. Cambridge University Press, Cambridge (1985).
53. Seidman, S. B. Network structure and minimum degree. *Soc. Net.* **5**, 269 (1983).
54. Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, F. & Arenas, A. Self-similar community structure in a network of human interactions. *Phys. Rev. E* **68**, 065103 (2003).
55. Rossi, R. A. & Ahmed, N. K. The Network Data Repository with Interactive Graph Analytics and Visualization. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (AAAI Press, 2015).
56. Rual, J. F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173 (2005).

## Acknowledgements

The authors acknowledge the valuable discussion with Wen-Xu Wang and Liming Pan, as well as the Erdős Number Project in Oakland University for the data set. This work is supported by the National Natural Science Foundation of China (Nos 71171136, 61374177, 11222543 and 61433014), JGL is supported by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning and the Shuguang Program Project of Shanghai Educational Committee (Grant No. 14SG42).

## Author Contributions

J.-G.L., J.-H.L. and T.Z. designed research, J.-G.L., J.-H.L. and Q.G. perform research, J.-G.L., J.-H.L. and T.Z. analysed data, and J.-G.L., J.-H.L., Q.G. and T.Z. wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Liu, J.-G. *et al.* Locating influential nodes via dynamics-sensitive centrality. *Sci. Rep.* **6**, 21380; doi: 10.1038/srep21380 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>