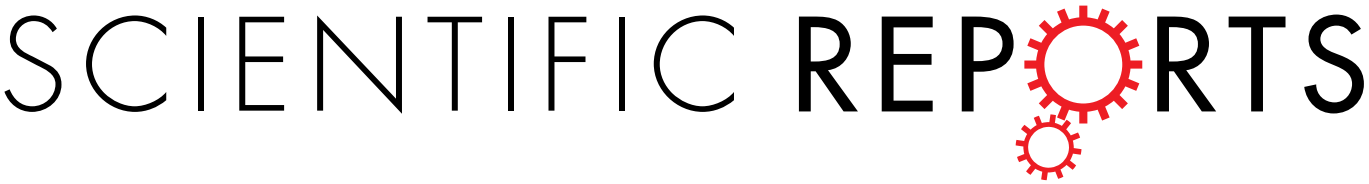

OPEN

Inferring Centrality from Network Snapshots

Haibin Shao[1], Mehran Mesbahi[2], Dewei Li[1] & Yugeng Xi[1]

Received: 24 June 2016
Accepted: 09 December 2016
Published: 18 January 2017

The topology and dynamics of a complex network shape its functionality. However, the topologies of many large-scale networks are either unavailable or incomplete. Without the explicit knowledge of network topology, we show how the data generated from the network dynamics can be utilised to infer the tempo centrality, which is proposed to quantify the influence of nodes in a consensus network. We show that the tempo centrality can be used to construct an accurate estimate of both the propagation rate of influence exerted on consensus networks and the Kirchhoff index of the underlying graph. Moreover, the tempo centrality also encodes the disturbance rejection of nodes in a consensus network. Our findings provide an approach to infer the performance of a consensus network from its temporal data.

Centrality is designed to identify the most important nodes in a network and has been examined for several decades[1–3]. A node with a larger centrality value is considered more influential in a network[4]. The network topology plays a paramount role in centrality metrics, for instance, through its degree (the number of connections incident upon a node), eigenvector (based on the idea that connections to high-scored nodes contribute more to the score of the node than connections to low-scored nodes), betweenness (quantifies the number of times a node acts as a bridge along the shortest path between two other nodes), closeness (the reciprocal of the sum of its distances from all other nodes), and k-shell (or k-core of a graph is its maximal connected subgraph in which all nodes have degree at least k), etc. However, the topology of many large-scale networks could be unavailable, incomplete or unreliable. Even when the network topology is known, it might be expensive to compute its centrality index using global information. To this end, we examine whether we can compute the network centrality without the explicit knowledge of network topology and whether each individual node can infer its role from local observations. In addition to the network topology, the dynamics provides an alternative ingredient for identifying the role of nodes in a complex network. Recently, more links between centrality metrics and the network dynamics have been explored, such as in the context of percolation centrality[5], control centrality[6], and grounding centrality[7], etc.

In this paper, we focus on consensus networks, a type of complex network whose components are diffusively coupled, and provide an elegant prototype for collective behaviors such as flocking of birds, schooling of fish, and phase transition in self-propelled agents[8]. We introduce a novel metric tempo centrality to quantify the influence of nodes in a consensus network, and show that the proposed metric can be computed from the temporal data of the network, even when the network topology is unavailable. An algorithm is subsequently given for the computation of tempo centrality. Moreover, the tempo centrality is shown to have a close correlation with both convergence rate and functional robustness of consensus networks.

Results

Tempo Centrality. Consider a networked system with $n \in \mathbb{N}$ nodes depending on the application. Each agent $i \in \mathcal{V}$ has a state at time t denoted by $x_i(t)$ that could represent position, temperature, or opinion. The interaction topology amongst agents is dictated by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with a node set $\mathcal{V} = \{1, 2, \cdots, n\}$, an edge set $\mathcal{E} = \mathcal{V} \times \mathcal{V}$, and an adjacency matrix $\mathcal{A}$, respectively. The adjacency matrix $\mathcal{A} \in \mathbb{R}^{n \times n}$ is such that $\mathcal{A}_{ij} = 1$ when $(i, j) \in \mathcal{E}$ and $\mathcal{A}_{ij} = 0$ otherwise. The networks discussed in this paper are assumed to be simple, undirected and connected[9].

[1]Department of Automation, Shanghai Jiao Tong University and the Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China. [2]Department of Aeronautics and Astronautics, University of Washington, Seattle, WA, 98195-2400, USA. Correspondence and requests for materials should be addressed to M.M. (email: mesbahi@aa.washington.edu) or D.L. (email: dwli@sjtu.edu.cn)

We discuss the consensus dynamics

$$\dot{x}_i(t) = \sum_{j=1}^{n} \mathcal{A}_{ij}(x_j(t) - x_i(t)) \tag{1}$$

that has been extensively examined in distributed coordination of multi-agent systems[10–13], opinion formation in social networks[14,15] and synchronization amongst oscillators[16]. We shall refer to network adopting consensus dynamics (1) as the consensus network. We say consensus is achieved for a consensus network when $\lim_{t\to\infty} x_i(t) = \lim_{t\to\infty} x_j(t)$ for all $i \neq j \in \mathcal{V}$[10]. The consensus dynamics (1) provides an elegant prototype for the emergence of synchronization[17].

Controlling a consensus network (steering the states of all the agents to a desired value) can be achieved by influencing a group of agents $\mathcal{W} \subset \mathcal{V}$ and the effect of control depends on the selection of the set $\mathcal{W}$. In this paper, we are interested in understanding the role of agents in a consensus network when acting as anchors through which the external inputs exert influence on the network.

Denote the set of external inputs as $\mathcal{U} = \{u_1, u_2, \cdots, u_m\}$, where $u_i \in \mathbb{R}$ and $i \in \{1, 2, \ldots, m\}$. Those agents who are directly influenced by the external inputs are referred to as the leader agents (or informed agents). The set of leader agents is denoted by $\mathcal{V}_l$. Denote by $\mathcal{B} \in \mathbb{R}^{n\times m}$ as the influence matrix such that $\mathcal{B}_{il} = 1$ if and only if there exists $u_l \in \mathcal{U}$ such that the agent $i \in \mathcal{V}_l$ is directly influenced by u_l, and $\mathcal{B}_{il} = 0$ otherwise. It is assumed that each agent can be directly influenced by at most one external input. Thus the sum of each row of $\mathcal{B}$ is less than 1 and there exists a one-to-one correspondence between $\mathcal{V}_l$ and $\mathcal{B}$. Denote the influence matrix determined by a node set $\mathcal{W} \subset \mathcal{V}$ as $\mathcal{B}(\mathcal{W})$. The individual behaviour is thus shifted to the following influenced consensus dynamics as a result of the external input,

$$\dot{x}_i(t) = \sum_{j=1}^{n} \mathcal{A}_{ij}(x_j(t) - x_i(t)) + \sum_{l=1}^{m} \mathcal{B}_{il}(u_l - x_i(t)). \tag{2}$$

The network state, denoted by $\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^{\mathrm{T}}$, characterises the behaviour of the overall system where $^{\mathrm{T}}$ represents the transpose operation. The dynamics of the network state with external inputs is then

$$\dot{\boldsymbol{x}}(t) = -\mathcal{L}_{\mathcal{B}}\boldsymbol{x}(t) + \mathcal{B}\mathbf{u}, \tag{3}$$

where $\boldsymbol{u} = [u_1, u_2, \ldots, u_m]^{\mathrm{T}}$, $\mathcal{L}_{\mathcal{B}} = \mathcal{L} + \mathcal{B}\mathcal{B}^T$, and $\mathcal{L} \in \mathbb{R}^{n\times n}$ represents the graph Laplacian or Kirchhoff matrix satisfying $\mathcal{L}_{ij} = \sum_{j=1}^{n} a_{ij}$ for all $i = j$ and $\mathcal{L}_{ij} = -a_{ij}$ for all $i \neq j$. The complex network with the influenced consensus dynamics (2) is referred to as the influenced consensus network.

An example of the structure of an influenced consensus network is shown in Supplementary Figure 1, where agent 1 is directly influenced by an external input $u \in \mathbb{R}$ and therefore is a leader agent. In this example, the influence matrix is

$$\mathcal{B} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

and the corresponding influenced consensus dynamics is

$$\begin{aligned} \dot{x}_1 &= (x_2 - x_1) + (u - x_1) \\ \dot{x}_2 &= (x_1 - x_2) + (x_3 - x_2) \\ \dot{x}_3 &= (x_2 - x_3). \end{aligned}$$

On the one hand, the evolution of the influenced consensus network (3) depends on the type of external inputs. The set of external inputs $\mathcal{U}$ is homogeneous if $u_i = u_j$ for all $u_i, u_j \in \mathcal{U}$ and heterogeneous otherwise. If the external inputs are homogeneous, then consensus is achieved on the value of the external input, namely, $\lim_{t\to\infty} x(t) = u_i\mathbf{1}$ where $\mathbf{1} = [1, 1, \cdots, 1]^T \in \mathbb{R}^n$ and $i \in \{1, 2, \ldots, m\}$[11]. Clusters (where the states of agents converge to several distinct values) would emerge when external inputs are heterogeneous[18]. The results in this paper are available for both homogeneous and heterogeneous inputs. We do assume however that the external inputs are time-invariant which can be considered as the belief of a zealot[19] or the opinion of stubborn individuals or political leaders[15].

On the other hand, the dynamics of the influenced consensus network is also shaped by the influence matrix $\mathcal{B}$. Note that $\mathcal{B}\mathcal{B}^T$ is a diagonal matrix with only ii-th entry 1 where $i \in \mathcal{V}_l$, and 0 for other entries. The matrix $\mathcal{L}_{\mathcal{B}}$ is then derived from the perturbation of the graph Laplacian by the diagonal matrix $\mathcal{B}\mathcal{B}^T$; it is therefore referred to as the perturbed Laplacian matrix. The perturbed graph Laplacian of an undirected graph is symmetric; hence we denote the eigenvalues of $\mathcal{L}_{\mathcal{B}}$ as $0 < \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_n$ and the corresponding normalised eigenvectors as $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$. In essence, the spectra of $\mathcal{L}_{\mathcal{B}}$ dictates the evolution of influenced consensus dynamics (3).

Controlling a consensus network via different leader agents impacts the network differently. The influence of an agent (or a group of agents) in a consensus network can be measured by the fluctuation of the network state $\boldsymbol{x}(t)$ triggered by this agent(s). We shall characterise the state fluctuation by a differentiation operator.

Taking the influenced consensus network in Supplementary Figure 1 as an example, we use quantity

$$r_{\{1\}}(t) = \frac{||\dot{x}_1(t)||_2}{||\dot{\boldsymbol{x}}(t)||_2}$$

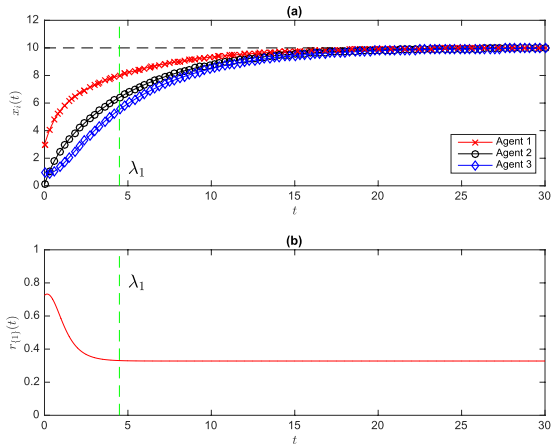

Figure 1. Coherence and tracking. (**a**) The trajectories of agents' states of the influenced consensus network in Supplementary Figure 1 and (**b**) the trajectory of $r_{\{1\}}(t)$.

to characterise the influence of leader agent 1, where $||\cdot||_2$ represents the Euclidean vector norm. The state evolution of each agent and $r_{\{1\}}(t)$ are shown in Fig. 1(a) and (b), respectively. The time window can be divided into two stages according to the convergence of $r_{\{1\}}(t)$, denoted by the coherence stage and the tracking stage, respectively. During the coherence stage, agents aggregate their states under the attractive force generated by diffusive couplings amongst agents. Since the consensus network is influenced by an external input $u = 10$, all agents subsequently track the external input via the leader agents. The two panels are both divided by the green dashed vertical lines at the time step $t \approx 4.5$, on which $r_{\{1\}}(t)$ has converged to a steady value of 0.328. The coherence stage is governed by all the eigenvalues of $\mathcal{L}_{\mathcal{B}}$ and the corresponding eigenvectors, but the tracking stage is dominated by λ_1 and $\boldsymbol{v}_1$. According to the Perron– Frobenius theorem, the entries of $\boldsymbol{v}_1$ are all nonzero and can be selected to be positive[20].

For a set of agents $\mathcal{W} = \{i_1, i_2, \cdots, i_s\} \subset \mathcal{V}$, we define its *tempo centrality* (TC), denoted by $\alpha(\mathcal{W})$, as $\|h(\mathcal{W})\boldsymbol{v}_1\|_2$, where $\boldsymbol{v}_1$ is the normalised eigenvector corresponding to the smallest eigenvalue of $\mathcal{L}_{\mathcal{B}(\mathcal{W})}$, $h(\mathcal{W}) = [e_{i_1}^{\mathrm{T}}, e_{i_2}^{\mathrm{T}}, \cdots, e_{i_s}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{s\times n}$, and $e_{i_j} \in \mathbb{R}^n$ denotes the column vector with 1 for the i_j-th entry and 0 otherwise, where $j \in \{1, 2, \ldots, s\}$. It has been shown in ref. 21 that the long-term behaviour of

$$r_{\mathcal{W}}(t) \triangleq \frac{\|h(\mathcal{W})\dot{\boldsymbol{x}}(t)\|_2}{\|\dot{\boldsymbol{x}}(t)\|_2} \tag{4}$$

is determined by $\boldsymbol{v}_1$, namely,

$$\lim_{t\to\infty} r_{\mathcal{W}}(t) = \|h(\mathcal{W})\boldsymbol{v}_1\|_2 = \alpha(\mathcal{W}), \tag{5}$$

implying that $\lim_{t\to\infty} r_{\mathcal{W}}(t)$ is independent of both the initial state $\boldsymbol{x}(0)$ and the external input $\boldsymbol{u}$, as shown in Supplementary Figs 2 and 3. Note that TC takes values from entries of a positive normalised eigenvector, therefore $\alpha(\mathcal{W}) \in (0, 1]$. The TC characterises the tempo of agents with respect to that of the entire network in the tracking stage when they act as leaders.

Computation of TC based on Temporal Data. The evolution of the network state is dictated by the network topology and the individual node dynamics. According to (5), the network snapshots that are generated by the individual dynamics can provide an alternative way for the computation of TC even when the network topology is unknown. We shall utilize this property for the design of an algorithm for computing TC from the temporal data of a consensus network.

Consider a consensus network with the set of leader agent $\mathcal{V}_l = \{i\}$ and the corresponding perturbed Laplacian matrix $\mathcal{L}_{\mathcal{B}}$. Denote by h_i and h_j as selection matrices for the leader agent $\{i\}$ and a distinct agent $\{j\}$, respectively. The following result has been established linking the network state and the spectra of perturbed Laplacian matrix in ref. 21,

$$\lim_{t\to\infty} r_{\{i\},\{j\}}(t) \triangleq \lim_{t\to\infty} \frac{\|h_i\dot{\boldsymbol{x}}(t)\|_2}{\|h_j\dot{\boldsymbol{x}}(t)\|_2} = \frac{\|h_i\boldsymbol{v}_1\|_2}{\|h_j\boldsymbol{v}_1\|_2}, \tag{6}$$

where $\boldsymbol{v}_1$ is the normalized eigenvector corresponding to the smallest eigenvalue of $\mathcal{L}_{\mathcal{B}}$.

We now show how TC can be computed from the network snapshots. Without loss of generality, designate agent 1 as the leader agent in the network, i.e., $\mathcal{V}_l = \{1\}$. Discretizing (3) at the sampling points $t_0 \le \delta_1 < \delta_2 < \ldots < \delta_m$ yields an m-length snapshots of the consensus network, i.e., $\tilde{\boldsymbol{x}}(k) = \boldsymbol{x}(t_0 + \delta_k)$ where $t_0 \ge 0$ is the initial time step and $k = 0, 1, \ldots, m$. Denote $\boldsymbol{\beta}(k) = [\beta_1(k), \beta_2(k), \cdots, \beta_n(k)]^{\mathrm{T}} \in \mathbb{R}^n$ satisfying

$$\beta_1(k+1) = \beta_1(k),$$

$$\beta_j(k+1) = \frac{\beta_1(k+1)}{\tilde{r}_{\{1\},\{j\}}(k)},$$

where

$$\tilde{r}_{\{1\},\{j\}}(k) = \frac{\|h_1(\tilde{\boldsymbol{x}}(k+1) - \tilde{\boldsymbol{x}}(k))\|_2}{\|h_j(\tilde{\boldsymbol{x}}(k+1) - \tilde{\boldsymbol{x}}(k))\|_2}, j \in \{2, 3, \cdots, n\}.$$

Then, it follows that

$$\boldsymbol{v}_1 = \lim_{k\to\infty} \frac{\boldsymbol{\beta}(k)}{\|\boldsymbol{\beta}(k)\|_2}.$$

According to the definition, the first entry of $\frac{\boldsymbol{\beta}(k)}{\|\boldsymbol{\beta}(k)\|_2}$ approaches the TC of agent 1 when $k \to \infty$. A value threshold $\varepsilon > 0$ needs to be set such that the computation of $\tilde{r}_{\{1\},\{j\}}(k)$ terminates when reaching its steady state. The above discussions conclude the following algorithm that computes the TC of leader agent 1.

Algorithm 1 Computation of Tempo Centrality from Temporal Data
Require: $\tilde{\boldsymbol{x}}(k)$, $k = 1, 2, \cdots, m$; $\varepsilon > 0$; $\beta_1(0) \neq 0$.
$k = 1$
while $\exists\, j \in \{1, 2, \cdots, n\}\backslash\{1\}$ such that $\|\tilde{r}_{\{1\},\{j\}}(k) - \tilde{r}_{\{1\},\{j\}}(k-1)\| > \varepsilon$ **do**
for $s = 1$ **to** n **do**
$\beta_s(k) = \frac{\beta_1(k)}{\tilde{r}_{\{1\},\{j\}}(k)}$
end for
$k = k + 1$
end while
return $\boldsymbol{v}_1 = \frac{\boldsymbol{\beta}(k)}{\|\boldsymbol{\beta}(k)\|_2}$

The complexity of Algorithm 1 depends on the convergence of $r_{\{i\},\{j\}}(k)$ for all $i, j \in \mathcal{V}$, which is controlled by the ε. As a result, the computation complexity of Algorithm 1 is $\mathcal{O}(2n^2 \cdot \omega(\varepsilon))$ where n is the network size and $\omega(\varepsilon)$ denotes the total number of steps after which $r_{\{i\},\{j\}}(k)$ converges to its steady state with an accuracy of ε. Note that $\|\boldsymbol{\beta}(k)\|_2$ is a global information for a given influence structure and is usually not accessible to the agents in the network. However, according to (6), each agent can still infer the relative role of nodes via local measurements.

Here, we proceed to discuss the amount of data needed for the implementation of Algorithm 1. Theoretically, we have

$$\lim_{k\to\infty} \tilde{r}_{\{i\},\mathcal{V}}(k) = |h_i \boldsymbol{v}_1|, i \in \{1, 2, 3, \cdots n\}.$$

We thereby quantify the error of estimating $\boldsymbol{v}_1$ with an m-length network snapshots $\{\tilde{\boldsymbol{x}}(k)\}_{k=1}^m$ by

$$\psi = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{m}\sum_{k=1}^{m}|\tilde{r}_{\{i\},\mathcal{V}}(k) - |h_i \boldsymbol{v}_1||\right\} \tag{7}$$

where

$$\tilde{r}_{\{i\},\mathcal{V}}(k) = \frac{\|h_i(\tilde{\boldsymbol{x}}(k+1) - \tilde{\boldsymbol{x}}(k))\|_2}{\|\tilde{\boldsymbol{x}}(k+1) - \tilde{\boldsymbol{x}}(k)\|_2}, i \in \{1, 2, 3, \cdots, n\}. \tag{8}$$

The smaller the value of ψ the more accurate the estimation of $\boldsymbol{v}_1$ with network snapshots. Intuitively, more data will improve the estimation accuracy. Figure 2(a) and (b) show the estimation error ψ is decreasing when more snapshots are used in an ER random network and a scale-free network. In fact, we can start sampling the network state $x(t)$, over a time interval of $t \in [t_1, t_2]$, from t_1 in a forward sequence (forward sampling), or from t_2 in a reversed sequence (backward sampling). The backward sampling is more preferable since $r_{\{i\},\mathcal{V}}(t)$ is closer to its steady state. In Fig. 2(c) and (d), we apply the backward sampling and the estimation error ψ is shown to be dramatically decreased compared with the forward sampling. In the backward sampling, it turns out that the length of snapshots m has little influence on the estimation error ψ. The two methods mentioned above sample $x(t)$ with respect to time t uniformly, which is somewhat difficult to achieve due to the disturbance in

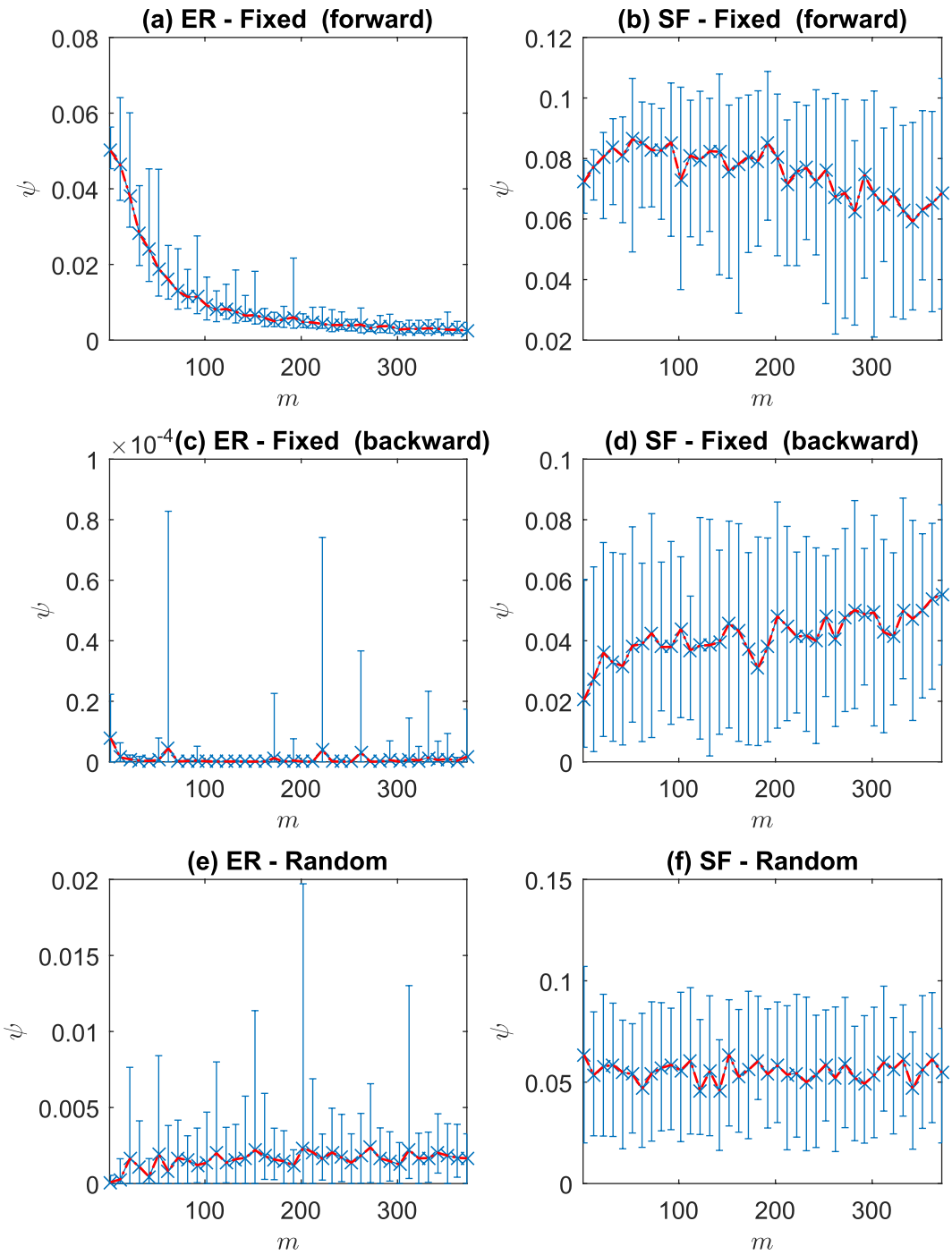

Figure 2. The error bar plot between the estimation error ψ and the length of snapshots m for forward sampling, backward sampling and random sampling in ER random networks (ER) and scale-free networks (SF) on 100 nodes. The error bar is plotted with 500 realizations for each network type and one leader is randomly selected for each such realizations. The edge occurrence probability for ER random networks is 0.05. In random sampling, at most m of snapshots are selected.

measurement. We subsequently examine the performance of random sampling of $x(t)$, namely, selecting at most m snapshots randomly. Figure 2(e) and (f) show that the performance of random sampling and the length of snapshots m also has little influence on the improvement of estimation accuracy in this scenario (for instance, in ER random networks, the estimation error obtained from $m = 5$ snapshots in random sampling is better than that from $m = 300$ in a forward sampling, as shown in Fig. 2(a) and (e)). We now show the influence of initial point of sampling δ_1 on ψ in Fig. 3. It turns out that the later the sampling process starts the more accurate the effect of estimation. Supplementary Figures 16, 17 and 18 show this trend in more networks. Therefore, the initial point of sampling is a critical factor. As a result, we can conclude that (a) the careful selection of sampling points is critical

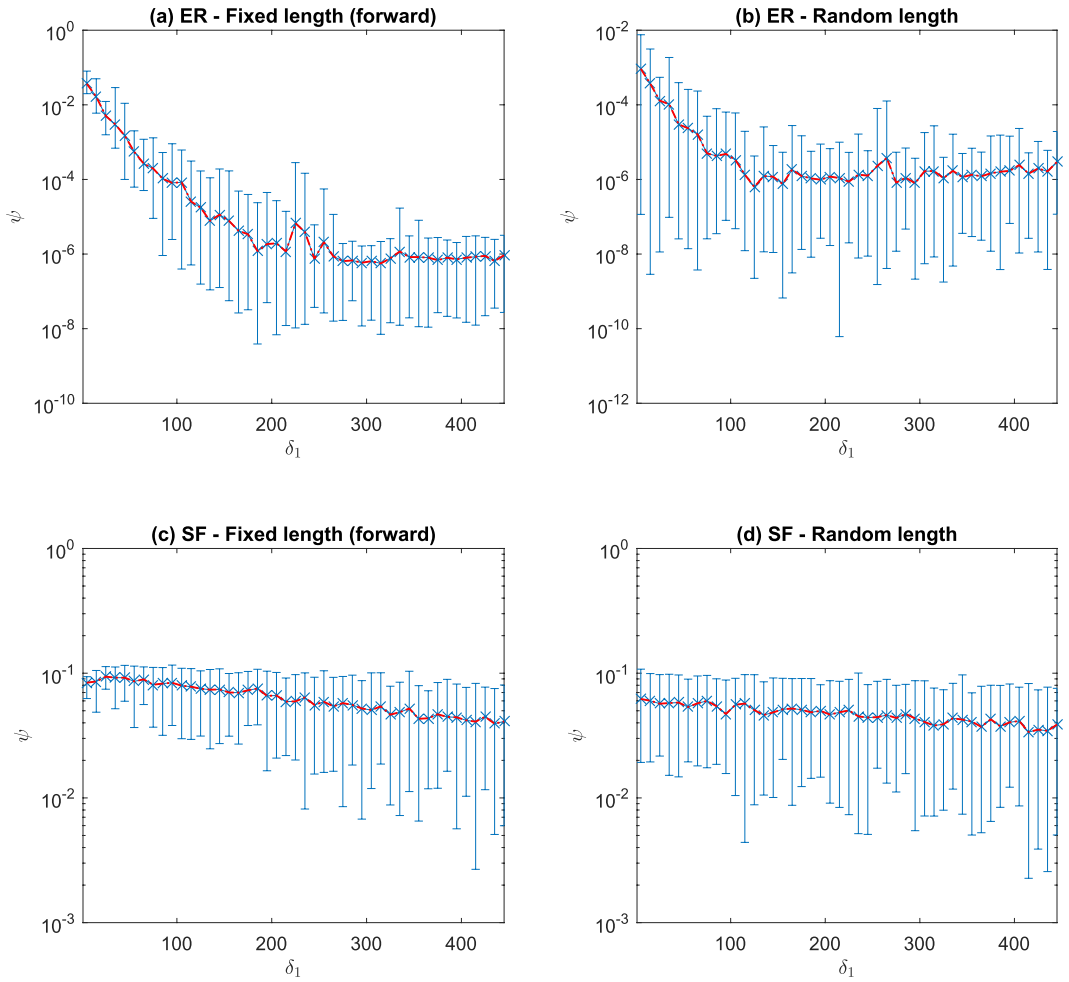

Figure 3. Error bar plot of the influence of initial point of sampling on ψ in 100-node ER random networks (ER) with forward sampling (**a**) and random sampling (**b**); and scale-free (SF) networks with forward sampling (**c**) and random sampling (**d**). A consecutive 20 snapshots from δ_1 have been selected. 500 realizations are used for each type of network.

in the efficiency of Algorithm 1, (b) the length of snapshots can be as short as $m = 2$ (in order to make the computation of (8) feasible) and (c) the sampling points can be heterogeneously distributed over the time interval.

Correlation with Network Performance. In this section, we proceed to explore the correlation between TC and two network performance metrics, namely, the convergence rate of consensus and the network robustness, both of which are closely related to the spectra of $\mathcal{L}_{\mathcal{B}}$[7,21]. Since TC is derived from the normalised eigenvector of the perturbed Laplacian matrix, there are analytic connections between TC and network performances that are determined by the spectra of $\mathcal{L}_{\mathcal{B}}$. However, one can compute TC based on network snapshots and thus the related network performances can be estimated with a data-driven approach.

Convergence Rate of Consensus. The smallest eigenvalue of $\mathcal{L}_{\mathcal{B}}$ (denoted by λ_1 in our discussion) provides us with the convergence rate of the consensus network influenced by the external input, on the other hand, λ_1 reflects how fast the diffusion of the external input (friendly or malicious, deterministic or stochastic) proceeds on the network[7,22]. Therefore, λ_1 can be regarded as a measure of spreading power of nodes in a consensus network from the perspective of the propagation efficiency of external inputs (refer to section 1 in Supplementary Information for details). Here, we provide some analytical facts between TC and λ_1. For an arbitrary agent $i \in \mathcal{V}$ with the corresponding selection matrix h and influence matrix $\mathcal{B}$, starting from

$$\mathcal{L}_{\mathcal{B}} \boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1, \tag{9}$$

and multiplying (9) by $\mathbf{1}^T$ from left on both sides yields

$$\mathbf{1}^T \mathcal{L}_{\mathcal{B}} \boldsymbol{v}_1 = \lambda_1 \mathbf{1}^T \boldsymbol{v}_1. \tag{10}$$

Since the entries of $\boldsymbol{v}_1$ can be chosen to have the same signs,

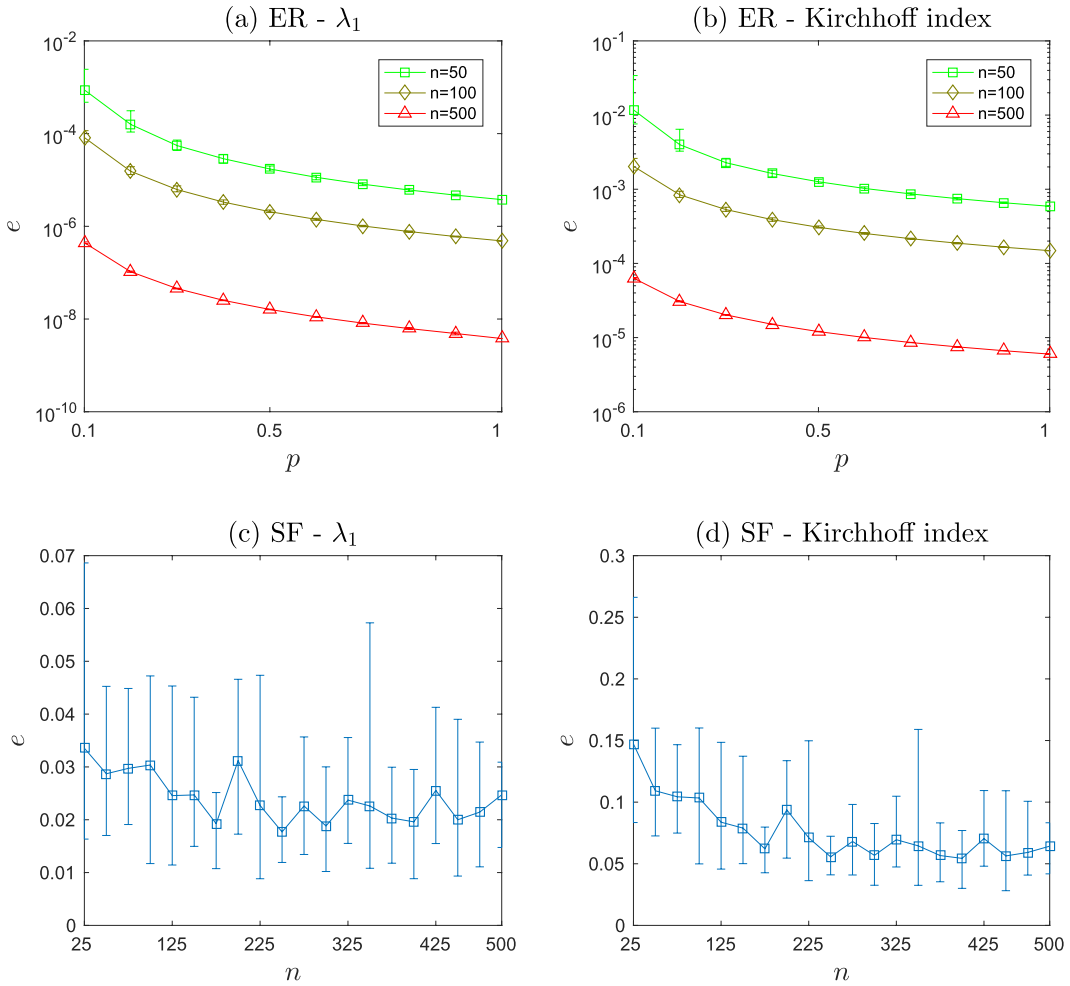

Figure 4. Error bar plot of approximation error in ER random networks and scale-free networks. The error of approximating λ_1 by $\frac{\alpha(\{i\})}{\sqrt{n}}$ in ER random networks (**a**) and scale-free networks (**c**); the error of approximating Kirchhoff index by $\frac{1}{2}\left(\sum_{i=1}^{n}\frac{\sqrt{n}}{\alpha(\{i\})} - n^2\right)$ in ER random networks (**b**) and scale-free networks (**d**). The possibility for edge occurrence p ranges from 0.1 to 1 at the regular interval 0.1 (the plot for $p=0.05$ to $p=0.1$ is shown in Supplementary Figure 19). For each realization of the network, we compute the TC for all nodes $i \in \mathcal{V}$ in the network and the corresponding $\lambda_1(\mathcal{L}_{\{i\}})$. Since the Kirchhoff index corresponds to the entire network, we only need to compute the Kirchhoff index for each realization of the network and its approximation by TC. 500 realizations are used for each type of network.

$$\lambda_1\|\boldsymbol{v}_1\|_1 = |h\boldsymbol{v}_1| = \gamma(\{i\}) \tag{11}$$

where $\|\cdot\|_1$ represents the 1-norm of vectors.

Note that $\boldsymbol{v}_1$ is normalised in terms of the 2-norm of vectors, i.e., $\|\boldsymbol{v}_1\|_2 = 1$. According to the equivalency of vector norms, we have

$$\|\mathbf{v}_1\|_1 \le \sqrt{n}\|\mathbf{v}_1\|_2 = \sqrt{n}, \tag{12}$$

which implies the propagation efficiency λ_1 is lower bounded by $\frac{\alpha(\{i\})}{\sqrt{n}}$. Figure 4(a,c) and Supplementary Table 3 show that λ_1 can be well-approximated by $\frac{\alpha(\{i\})}{\sqrt{n}}$ in ER random networks, scale-free networks and many empirical networks. The error bar of approximation errors in Fig. 4 reflects that the approximation effect is improved in large and dense networks. However, as shown in Supplementary Figures 19(a) and (c), the maximum relative errors corresponding to large sparse networks still imply an accurate estimation of λ_1 with TC. We proceed to use Pearson correlation analysis to show that there is a strong linear correlation between TC and λ_1. The Pearson coefficient between TC and λ_1 in an 500-node Erdős-Rényi (ER) random network and a 500-node scale-free network (SF) is shown in Fig. 5(a,b). The Pearson coefficients (PC) are 1 and 0.99943 for ER random network and scale-free network, respectively, implying the total positive correlation between TC and λ_1.

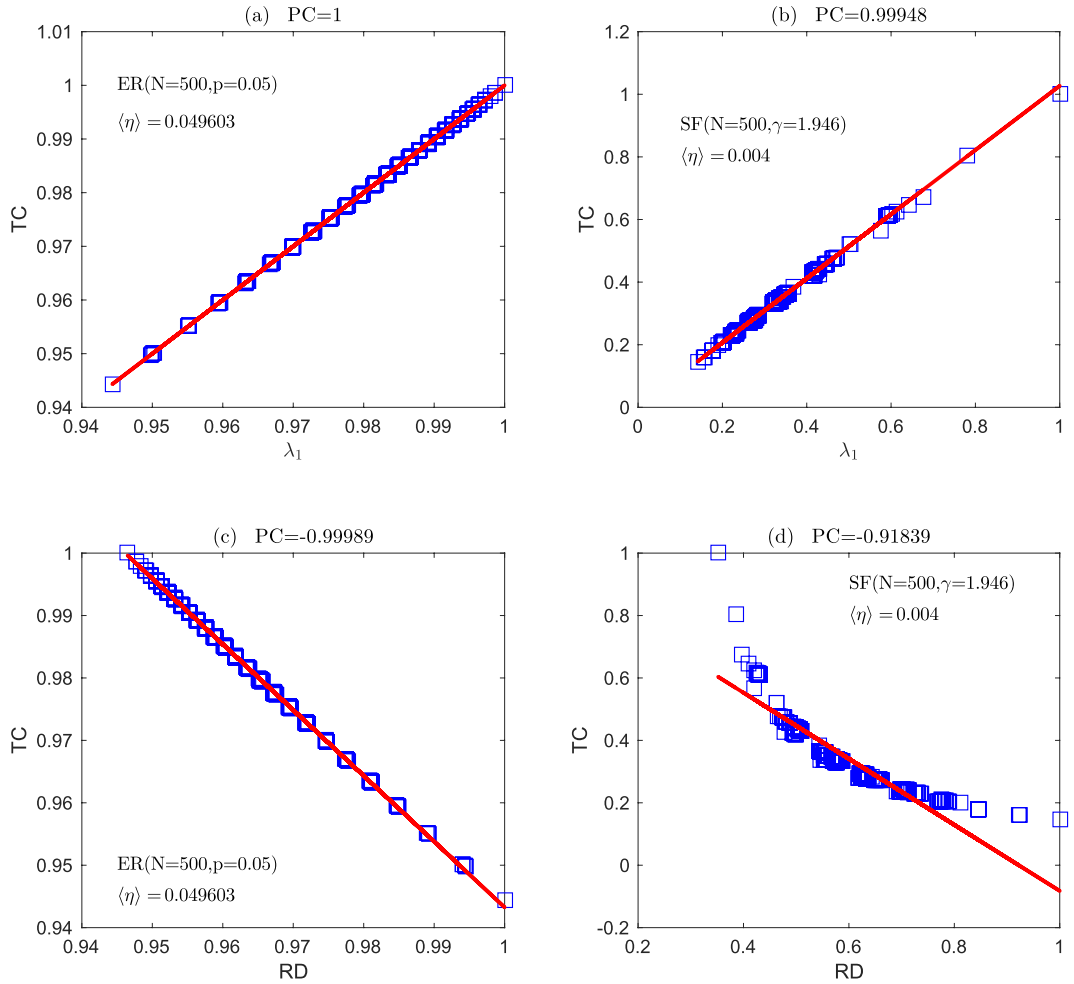

Figure 5. The TC as a function of λ_1 in an Erdős-Rényi (ER) random network (**a**) and a scale-free network (SF) (**b**). The Pearson coefficient between TC and λ_1 are 1 (ER) and 0.99948 (SF), respectively. The TC as a function of resistance distance (RD) in an Erdős-Rényi (ER) random network (**c**) and a scale-free network (SF) (**d**). The Pearson coefficient between TC and RD are −0.99989 (ER) and −0.91839 (SF), respectively.

Kirchhoff Index and Network Robustness. Viewing each edge in the network as a one-Ohm resistor, the effective resistance from the leader agents to all agents in the network is shown to be an index of network robustness from the perspective of both the $\mathcal{H}_2$ system norm (ability of disturbance rejection in the presence of additive noise in consensus networks)[23,24] and the vulnerability of the network connectivity to node or edge failures[23–25]. The perturbed Laplacian matrix turns out to be an important construct that is closely related to the effective resistance. Specifically, the $\mathbf{trace}(\mathcal{L}_{\mathcal{B}}^{-1})$ measures the functional robustness of the influenced consensus network in terms of the variance of deviation from consensus and the disturbance rejection properties of consensus networks, both of which can be characterised by the effective resistance of the network[22,26–29]. The structural robustness of complex networks has also been widely investigated in the context of percolation theory[30,31].

The resistance distance from agents i to j, denoted by d_{ij}, is the effective resistance between i and j, which can be quantified by the diagonal entries of $\mathcal{L}_{\mathcal{B}}^{-1}$ [24,32]. Suppose that the network is influenced by a single external input via agent $i \in \mathcal{V}$ and denote the resultant perturbed Laplacian matrix as $\mathcal{L}_{\{i\}}$, which is invertible if $\mathcal{G}$ is connected[32]. The resistance distance from input u to agent $j \in \mathcal{V}$ is $[\mathcal{L}_{\{i\}}^{-1}]_{jj}$ [22,32], and thus the resistance distance from the agent i to agent j is such that $d_{ij} = [\mathcal{L}_{\{i\}}^{-1}]_{jj} - 1$. Hence, the resistance distance from agent i to all agents in $\mathcal{G}$, denoted by q_i, is the summation of d_{ij} over all $j \in \mathcal{V}$, i.e.,

$$\begin{aligned} q_i &= \sum_{j=1}^{n} d_{ij} \\ &= \mathbf{trace}(\mathcal{L}_{\{i\}}^{-1}) - n \\ &= \left(\sum_{j=1}^{n} \frac{1}{\lambda_j(\mathcal{L}_{\{i\}})} \right) - n. \end{aligned} \tag{13}$$

In the following discussion, we shall refer to q_i as the resistance distance (RD) index of $i \in \mathcal{V}$. The Kirchhoff index of a network $\mathcal{G}$ (introduced by Klein and Randic′ in ref. 24) is half of the total effective resistance between all pair of nodes in $\mathcal{G}$ and can therefore be computed by

$$\begin{aligned}\kappa &= \frac{1}{2}\sum_{i=1}^{n} q_i \\ &= \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{\lambda_j(\mathcal{L}_{\{i\}})} - n^2\right).\end{aligned} \tag{14}$$

Kirchhoff index is a measure of connectivity and size of a network in terms of resistance distance[23]. A network with smaller value of Kirchhoff index is considered to be more robust to node or edge failures. Kirchhoff index is also related to the average power dissipation of the circuit with a random current excitation and the average commute time of a Markov chain on a graph. Algorithms have also been proposed for minimizing the Kirchhoff index a network[33].

The smallest eigenvalue of the perturbed Laplacian matrix can be considered as the perturbation of the smallest eigenvalue of the Laplacian matrix, which is always zero. The influence of this perturbation decreases when the network size grows, and consequently the term $\frac{1}{\lambda_1(\mathcal{L}_{\{i\}})}$ would dominate q_i in (13). Based on the estimation of $\lambda_1(\mathcal{L}_{\{i\}}) \approx \frac{\alpha(\{i\})}{\sqrt{n}}$, we can further approximate q_i by

$$q_i \approx \frac{\sqrt{n}}{\alpha(\{i\})} - n, \tag{15}$$

and the Kirchhoff index by

$$\kappa \approx \frac{1}{2}\left(\sum_{i=1}^{n}\frac{\sqrt{n}}{\alpha(\{i\})} - n^2\right). \tag{16}$$

Figure 4(b) and (d) and Supplementary Table 3 show the effect of this approximation in ER random networks, scale-free networks, and empirical networks. The maximum relative errors shown in Supplementary Figures 19(c) and 19(d) corresponding to large sparse networks still imply a good estimation of Kirchhoff index with TC. The Pearson coefficient between TC and resistance distance in 500-node Erdős-Rényi (ER) random network and 500-node scale-free network is shown in Fig. 5(c) and (d). The Pearson coefficients (PC) are -0.99988 and -0.91485 for the ER random network and the scale-free network (both implying the strong negative correlations). We have shown the results of Pearson correlation analysis and the corresponding regression coefficient in Supplementary Tables 1 and 2 for all 7 metrics mentioned in this paper under ER random networks, scale-free networks, and empirical networks. It has been shown that there exist a total positive correlation between TC and λ_1, and a strong negative correlation between TC and resistance distance for different classes of networks, enabling us to estimate these network performance metrics from network snapshots. The robustness of nodes in terms of centrality attack is shown in Fig. 6 and Supplementary Figures 21 and 22.

Zachary's Karate Club. As a case study, we have applied TC to quantify the role of members in the Zachary's Karate club[34]. In this case, TC highlights the president (node 1) and lesson instructor (node 34) in the club by assigning the top two largest TC values to them. The rank of each node in the network in terms of the value of TC, degree, eigenvector, betweenness, closeness and k-shell is shown in Table 1. We compute the Spearman's rank correlation coefficient (the Pearson correlation between the rank values of two variables) between TC and degree, eigenvector, betweenness, closeness and k-shell, respectively. TC is shown closer to the degree (with Spearman's rank correlation coefficient as 0.9422) in this example, followed by closeness (0.9251), betweenness (0.9013), eigenvector (0.8597) and k-shell (0.8481). We subsequently compared the TC with above five centrality metrics in the Karate club network by normalising the values of these centrality metrics as illustrated in Supplementary Figure 4. The distribution of the TC values in the heat map is shown in Fig. 7. According to the pattern of the normalised centrality values, TC identifies the distinct role of nodes compared with the other five centrality metrics in this example. The scatter plots and Pearson coefficient between TC and seven other metrics (above five centralities plus λ_1 and RD) in Karate club network are shown in Supplementary Figures 5 and 6 respectively. More scatter plots are shown in Supplementary Figures 7–15 for 9 empirical networks.

Discussion

Instead of quantifying the importance of a single agent in the network, the importance of a group of agents can also be characterised by TC, which is meaningful when the group is considered as in union or multiple leader agents are needed to exert influence on a consensus network. For a given group of leader agents, the larger the value of TC the faster the external influence propagates throughout the network and the less resistance the external influence has to overcome. In the meanwhile, the larger TC implies that the leader agent related to the influence matrix evolves towards the external input with a faster tempo than the rest of the network. Note that the influence of a single agent decreases when the network size grows; as a result, more leader agents are needed to exert a stronger influence.

Topology identification (through such methods as compressive sensing, linear system identification, power spectral analysis, convex programming etc.) is generally based on the minimization of the distance between the

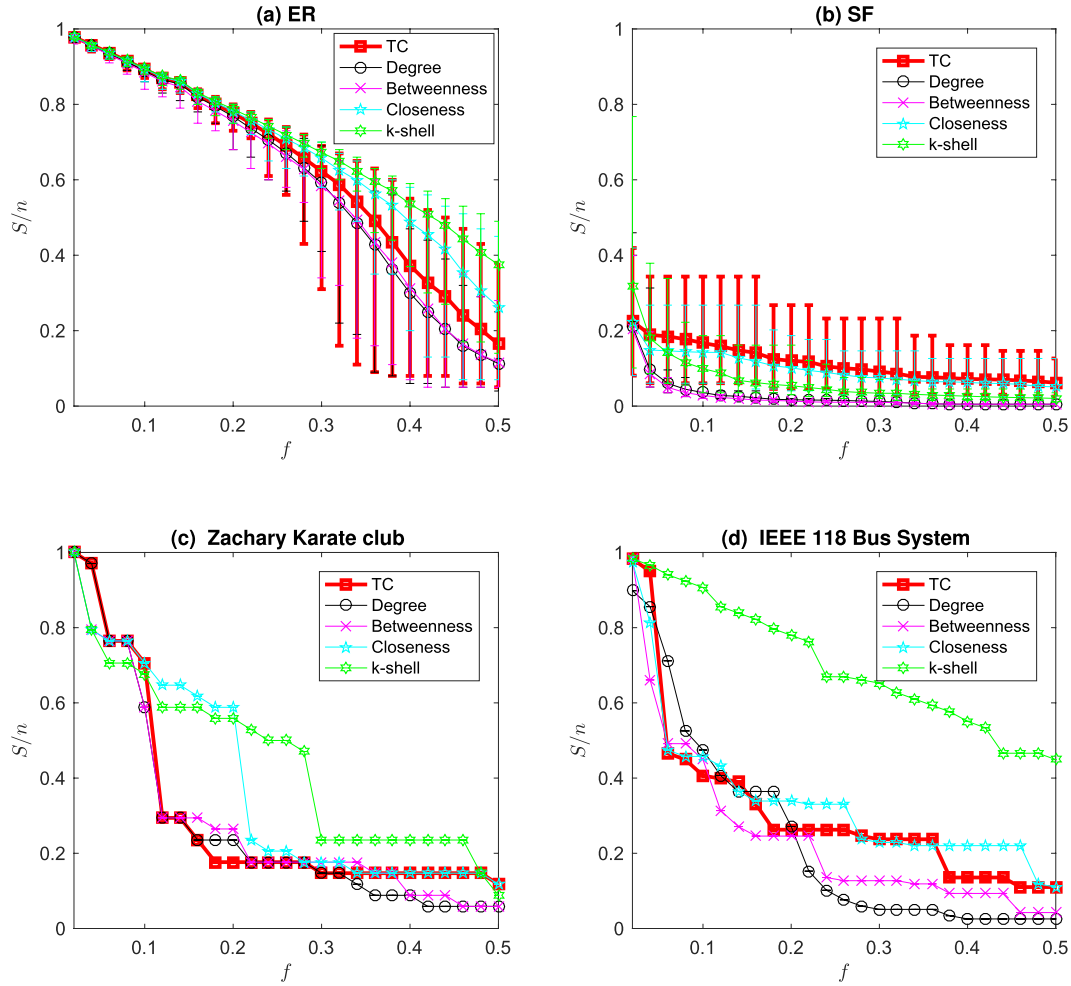

Figure 6. The effect of centrality attack $\frac{S}{n}$ as a function of the fraction of removed nodes f for centrality metrics including TC, degree, betweenness, closeness and k-shell. (**a**) The ER random (ER) network with $n = 500$ and $p = 0.05$, (**b**) the scale-free (SF) network with $n = 100$ and $\langle \eta \rangle = 0.02$, (**c**) the Karate club network with $n = 34$, and (**d**) the IEEE 118 bus system network with $n = 118$. The error bar plots for ER and SF are over 500 realizations for each network type.

estimated and the real network structures. Therefore, the entry-wise errors are acceptable for the optimization procedure. However, the role of an individual node can be different with respect to an entry-wise error, even when the two estimated network structures are equivalent from an optimization perspective. Our data-driven approach is robust to this ambiguity.

The second smallest eigenvalue of the graph Laplacian is known as the algebraic connectivity of a network, characterising the convergence rate of the consensus process[10,35]; and the corresponding eigenvector, called the Fiedler vector, has been widely applied in the network partitioning[35]. We can regard all agents in a network $\mathcal{G}$ along with the set of external input $\mathcal{U}$ as an expanded directed graph, in which the directed edges are present only from the external input to the leader agents. The vector $[\boldsymbol{v}_1^{\mathrm{T}}\ \mathbf{0}_{m\times 1}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n+m}$ can be considered as the Fiedler vector for this expanded digraph, where $\mathbf{0}_{m\times 1}$ denotes the $m \times 1$ vector with all zero entries. The impact of the external input on the network $\mathcal{G}$ can be characterised by the spectra of $\mathcal{L}_{\mathcal{B}}$. Using the spectral decomposition of the perturbed Laplacian matrix, i.e., $\mathcal{L}_{\mathcal{B}} = \sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathrm{T}}$, we observe that the eigenvector $\boldsymbol{v}_1$ shapes the behaviour of the network when the dynamics in directions of $\boldsymbol{v}_2, \boldsymbol{v}_3 \ldots, \boldsymbol{v}_n$ vanish. As one can see from Supplementary Figure 20, that the distribution of entries in $\boldsymbol{v}_1$ is shaped by the placement of the leader agent, that is, the farther a node is from the leader agent, the larger the absolute value of its corresponding entry in $\boldsymbol{v}_1$.

A summary of the utility of TC and our proposed approach is as follows. The TC is shown to be closely correlated to two performance metrics for consensus networks, namely, the convergence rate of the influenced consensus (characterising the propagation efficiency of the external influence over a consensus network) and the network functional robustness in terms of disturbance rejection; the TC can be computed from the network snapshots even when the network topology is unknown; the data-driven approach in this paper enables each agent to determine its relative role in the network by using the state of its neighbors; this relative role obtained by each agent via local information is a global information; since the linearization of the Kuramoto dynamics around the

Node	TC	D	E	B	C	K	Node	TC	D	E	B	C	K
1	2	2	2	1	1	1	**18**	30	28	24	29	22	28
2	5	5	5	7	9	2	**19**	24	29	18	30	29	29
3	3	4	3	4	2	3	**20**	15	19	13	9	8	15
4	9	6	8	15	10	4	**21**	28	30	19	31	30	30
5	21	17	28	21	20	11	**22**	29	31	23	32	23	31
6	17	11	26	10	17	12	**23**	25	32	20	33	31	32
7	18	12	27	11	18	13	**24**	10	10	12	13	16	16
8	12	13	11	23	14	5	**25**	20	20	32	18	24	17
9	7	8	6	6	5	6	**26**	19	21	31	16	25	18
10	23	23	17	20	15	23	**27**	32	33	30	34	33	33
11	22	18	29	22	21	14	**28**	13	14	15	12	11	19
12	34	34	33	24	32	34	**29**	16	22	16	19	13	20
13	31	24	25	25	26	24	**30**	14	15	14	17	19	21
14	8	9	7	8	6	7	**31**	11	16	10	14	12	8
15	27	25	21	26	27	25	**32**	6	7	9	5	4	22
16	26	26	22	27	28	26	**33**	4	3	4	3	7	9
17	33	27	34	28	34	27	**34**	1	1	1	2	3	10

Table 1. Rank of each node in Zachary's Karate club network in terms of TC, degree (D), eigenvector (E), betweenness (B), closeness (C) and k-shell (K). The rank is arranged in descending order according to the respective centrality value.

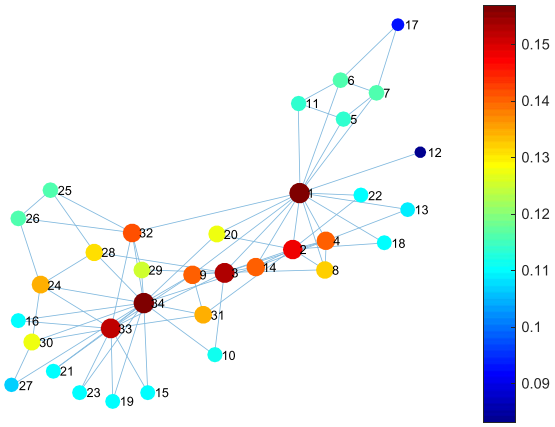

Figure 7. The TC distribution on Zachary's Karate club network. The size of dots are proportional to their TC value. The top two largest TC values are achieved for agent 34 and agent 1 (0.1569 and 0.1561, respectively).

origin is consensus dynamics the results discussed in this paper have a natural extension to the Kuramoto model, widely investigated for oscillator networks[36].

Methods

Normalization of Centrality. The centrality vector $\boldsymbol{c} = [c_1, c_2, \ldots, c_n]^{\mathrm{T}}$ of a network $\mathcal{G}$ is composed of the centrality value c_i of all nodes $i \in \mathcal{V}$ in $\mathcal{G}$. Since the value of centrality could differ in scale, we normalize c_i to the interval [0, 1] by the maximum value of a given centrality as follows

$$c_i' = \frac{c_i}{\max_i\{c_i\}}, \; i = 1, 2, \cdots, n.$$

Pearson Correlation Coefficient. The Pearson correlation coefficient for two vectors $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^{\mathrm{T}} \in \mathbb{R}^n$ and $\mathbf{y} = [y_1, y_2, \cdots, y_n]^{\mathrm{T}} \in \mathbb{R}^n$ is defined as

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \frac{n\sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{\sqrt{[n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2][n\sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2]}}.$$

Relative Error. The approximation error is measured by the relative error between vectors. Let $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ be an approximation of vector $\boldsymbol{x} \in \mathbb{R}^n$. The relative error used in this paper is,

$$e = \frac{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2}{\|\boldsymbol{x}\|_2}.$$

Edge Density. The edge density of a network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ is defined as the ratio between the number of edges and the maximum number of possible edges in a network, i.e.,

$$\eta = \frac{2|\mathcal{E}|}{|\mathcal{V}|(|\mathcal{V}| - 1)}.$$

Error Bar. The error bar plots in this paper represent the fluctuation of data around their respective mean. The marker in the centre of an error bar represents the mean of the data; the upper and lower intervals represent the deviation of the mean from the maximum and minimum value of the data.

Centrality Attack. In the centrality attack, the nodes in a network are sorted in descending order according to their centrality value, then the first $f \in (0, 1]$ fraction of nodes are removed from the network. The normalized size of largest connected component $\frac{S}{n} \in (0, 1]$ is used to quantify the effect of the attack.

Networks. An n-node Erdős-Rényi (ER) random network with the probability of edge realization $p \in [0, 1]$ is denoted by $\mathcal{G}(n, p)$ and is constructed by randomGraph.m in Octave routines for network analysis[37]. The scale-free network with degree distribution $p(k) \sim k^{-\gamma}$ is generated by SFNG.m in B-A Scale-Free Network Generation and Visualization[38]. The empirical networks are from KONECT[39]. The centrality value of the network degree, eigenvector, betweenness and closeness are computed using Octave routines for network analysis[37]. The k-shell is computed by corenums.m in Graph Algorithms In Matlab Code (gaimc)[40].

References

1. Freeman, L. C. Centrality in social networks conceptual clarification. *Social Networks* **1,** 215–239 (1979).
2. Borgatti, S. P. Centrality and network flow. *Social Networks* **27,** 55–71 (2005).
3. Borgatti, S. P. & Everett, M. G. A graph-theoretic perspective on centrality. *Social Networks* **28,** 466–484 (2006).
4. Sabidussi, G. The centrality index of a graph. *Psychometrika* **31,** 581–603 (1966).
5. Piraveenan, M., Prokopenko, M. & Hossain, L. Percolation centrality: Quantifying graph-theoretic impact of nodes during percolation in networks. *PloS ONE* **8,** e53095 (2013).
6. Liu, Y., Slotine, J. & Barabási, A. Control centrality and hierarchical structure in complex networks. *PloS ONE* **7,** e44459–e44459 (2011).
7. Pirani, M. & Sundaram, S. On the smallest eigenvalue of grounded laplacian matrices. *IEEE Transactions on Automatic Control* **61,** 509–514 (2016).
8. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. & Shochet, O. Novel type of phase transition in a system of self-driven particles. *Physical review letters* **75,** 1226 (1995).
9. Godsil, C. & Royle, G. F. *Algebraic graph theory* (Springer Science & Business Media, 2013).
10. Olfati-Saber, R. & Murray, R. M. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control* **49,** 1520–1533 (2004).
11. Jadbabaie, A., Lin, J. & Morse, A. S. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control* **48,** 988–1001 (2003).
12. DeGroot, M. H. Reaching a consensus. *Journal of the American Statistical Association* **69,** 118–121 (1974).
13. Mesbahi, M. & Egerstedt, M. *Graph theoretic methods in multiagent networks* (Princeton University Press, 2010).
14. Lorenz, J. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C* **18,** 1819–1838 (2007).
15. Acemoglu, D., Como, G., Fagnani, F. & Ozdaglar, A. Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research* **38,** 1–27 (2013).
16. Fell, J. & Axmacher, N. The role of phase synchronization in memory processes. *Nature Reviews Neuroscience* **12,** 105–118 (2011).
17. Strogatz, S. *SYNC: The emerging science of spontaneous order* (Hyperion, 2003).
18. Shao, H., Xi, Y. & Mesbahi, M. On the degree of synchrony. In *Proceedings of 54th IEEE Conference on Decision and Control (CDC)*, 986–991 (IEEE, 2015).
19. Mobilia, M. Does a single zealot affect an infinite group of voters? *Physical Review Letters* **91,** 028701 (2003).
20. Horn, R. A. & Johnson, C. R. *Matrix analysis* (Cambridge University Press, 2012).
21. Shao, H. & Mesbahi, M. Degree of relative influence for consensus-type networks. In *Proceedings of 2014 American Control Conference*, 2676–2681 (IEEE, 2014).
22. Chapman, A. & Mesbahi, M. Semi-autonomous consensus: Network measures and adaptive trees. *IEEE Transactions on Automatic Control* **58,** 19–31 (2013).
23. Ellens, W., Spieksma, F., Van Mieghem, P., Jamakovic, A. & Kooij, R. Effective graph resistance. *Linear Algebra and Its Applications* **435,** 2491–2506 (2011).
24. Klein, D. J. & Randić, M. Resistance distance. *Journal of Mathematical Chemistry* **12,** 81–95 (1993).
25. Abbas, W. & Egerstedt, M. Robust graph topologies for networked systems. *In Proceedings of the 3rd IFAC Workshop on Distributed Estimation and Control in Networked Systems* **3,** 85–90 (2012).
26. Lin, F., Fardad, M. & Jovanovic, M. R. Algorithms for leader selection in stochastically forced consensus networks. *IEEE Transactions on Automatic Control* **59,** 1789–1802 (2014).
27. Bamieh, B., Jovanovic, M., Mitra, P. & Patterson, S. Coherence in large-scale networks: Dimension-dependent limitations of local feedback. *IEEE Transactions on Automatic Control* **57,** 2235–2249 (2012).
28. Lovisari, E., Garin, F. & Zampieri, S. Resistance-based performance analysis of the consensus algorithm over geometric graphs. *SIAM Journal on Control and Optimization* **51,** 3918–3945 (2013).

29. Clark, A., Bushnell, L. & Poovendran, R. A supermodular optimization framework for leader selection under link noise in linear multi-agent systems. *IEEE Transactions on Automatic Control* **59,** 283–296 (2014).
30. Cohen, R. & Havlin, S. *Complex networks: structure, robustness and function* (Cambridge University Press, 2010).
31. Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406,** 378–382 (2000).
32. Barooah, P. & Hespanha, J. P. Graph effective resistance and distributed control: Spectral properties and applications. In *Proceedings of the 45th IEEE Conference on Decision and Control*, 3479–3485 (IEEE, 2006).
33. Ghosh, A., Boyd, S. & Saberi, A. Minimizing effective resistance of a graph. *SIAM Review* **50,** 37–66 (2008).
34. Zachary, W. W. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research* 452–473 (1977).
35. Fiedler, M. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal* **23,** 298–305 (1973).
36. Acebrón, J. A., Bonilla, L. L., Vicente, C. J. P., Ritort, F. & Spigler, R. The kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics* **77,** 137 (2005).
37. Bounova, G. Octave networks toolbox (2015). https://doi.org/10.5281/zenodo.22398.
38. George, M. Ba scale-free network generation and visualization. *Matlab Central* **12** (2006).
39. Kunegis, J. Konect: the koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*, 1343–1350 (ACM, 2013).
40. Gleich, D. gaimc: graph algorithms in matlab code. *Matlab Toolbox. Matlab* (2009).

Acknowledgements

The work of H.S., D.L., and Y.X. has been supported by the National Science Foundation of China (Grant Nos 61333009, 61433002, 61521063, 61590924, 71361130012) and National Program on Key Basic Research Project of China (973 Program, Grant No. 2014CB249200). The work of M.M. has been supported by the National Science Foundation of U.S.A. (Grant No. SES-1541025). We thank anonymous reviewers for the constructive suggestions on this paper.

Author Contributions

M.M. and Y.X. envisioned the research. M.M., Y.X. and H.S. discussed the main points of the manuscript. H.S. implemented the theoretical analysis. H.S. and D.L. designed the numerical simulation and analysed the result. H.S. wrote the manuscript. All authors reviewed the manuscript.

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: Shao, H. *et al.* Inferring Centrality from Network Snapshots. *Sci. Rep.* **7**, 40642; doi: 10.1038/srep40642 (2017).

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