MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Analysis of Relative Influence of Nodes in Directed Networks

Naoki MASUDA, Yoji KAWAMURA and Hiroshi KORI

Analysis of Relative Influence of Nodes in Directed Networks

Naoki MASUDA, Yoji KAWAMURA and Hiroshi KORI

September 11, 2009

Abstract

Many complex networks are described by directed links; in such networks, a link represents, for example, the control of one node over the other node or unidirectional information flows. Some centrality measures are used to determine the relative importance of nodes specifically in directed networks. We analyze such a centrality measure called the influence. The influence represents the importance of nodes in various dynamics such as synchronization, evolutionary dynamics, random walk, and social dynamics. We analytically calculate the influence in various networks, including directed multipartite networks and a directed version of the Watts-Strogatz small-world network. The global properties of networks such as hierarchy and position of shortcuts, rather than local properties of the nodes, such as the degree, are shown to be the chief determinants of the influence of nodes in many cases. The developed method is also applicable to the calculation of the PageRank. We also numerically show that in a coupled oscillator system, the threshold for entrainment by a pacemaker is low when the pacemaker is placed on influential nodes. For a type of random network, the analytically derived threshold is approximately equal to the inverse of the influence. We numerically show that this relationship also holds true in a random scale-free network and a neural network.

1 Introduction

Networks abound in various fields; a network is a collection of nodes and links, where a link connects a pair of nodes. Most real-world networks are not entirely regular or random and have prominent properties as modeled by, for example, small-world, scale-free, hierarchical, and modular networks [1, 2]. In such networks, some nodes are considered to be more important than the others. Depending on the definition of importance, various centrality measures, which quantify the relative importance of different nodes, have been proposed. The most frequently used centrality measures are perhaps the degree (i.e., the number of links owned by a node) and the betweenness (i.e., the normalized number of shortest paths connecting any pair of nodes

passing through the node in question) [2, 3, 4]. New centrality measures have also been proposed in the field of complex networks [5, 6].

Although many centrality measures are available, very few of these describe the importance of nodes in collective behavior of nodes on networks (see [6]). In a previous study, we proposed a centrality measure called the influence [7]. The influence of a node denotes its importance in different types of dynamics. It represents the amplitude of the response of a synchronized network when an input is given to a certain node [8], the fixation probability for a newly introduced type (e.g., new information) at a node in voter-type evolutionary dynamics [9], the stationary density of a simple random walk in continuous time [9], the so-called reproductive value of a node [10], and the influence of a node in the DeGroot's model of consensus formation [11]. It makes sense to consider the influence only in directed networks; in undirected networks, the influences of all the nodes take an identical value. In principle, the influence as a centrality measure is close to the PageRank, which was originally developed for ranking websites [13].

To assess the influence (and also the PageRank) in real complex networks, it is not sufficient to take into account the local property of the node, such as the degree. The global structure of networks such as the small-world property, modular structure, and self-similarity [1, 2] generally affects the influence values.

In the present study, we analytically determine the influence of nodes in model networks such as weighted chain, directed multipartite networks with a hierarchical structure, and a directed version of the Watts-Strogatz small-world network [14]. For this purpose, we exploit the symmetry in networks and the relationship between the enumeration of directed spanning trees and the influence. We reveal the discrepancy between the actual influence and that predicted by the mean-field approximation (MA), which takes into account only the degree. In fact, the nodes that occupy globally important positions in terms of influence are generally different from those that are locally important. The globally important nodes govern the abovementioned dynamics on networks. Finally, to demonstrate the application of the influence as a centrality measure, we analyze a system of coupled oscillators and show that nodes with large influence values entrain other nodes relatively easily, i.e., with a relatively small coupling strength.

2   Influence

Consider a directed and weighted network having N nodes. The weight of the directed link from node i to node j is denoted by w_{ij}. We set $w_{ij} = 0$ when the link is absent. The influence of node i is denoted by v_i. We define

v_i as the solution for the following set of N linear equations:

$$v_i = \frac{\sum_{j=1}^{N} w_{ij} v_j}{k_i^{\text{in}}}, \quad (1 \leq i \leq N), \tag{1}$$

where $k_i^{\text{in}} \equiv \sum_{j=1}^{N} w_{ji}$ is the indegree of node i, and the normalization is given by $\sum_{i=1}^{N} v_i = 1$. When node i has many outgoing links, v_i can be large because there are many terms on the right-hand side of Eq. (1). When node i has many incoming links, node i is interpreted to be governed by many nodes. Then, v_i can be small because of the divisive factor k_i^{in} in Eq. (1). The rationale for the definition given in Eq. (1) is that v_i represents the importance of nodes in different types of dynamics on networks, as explained in Sec. 1. Values of v_i for two example networks are shown in Fig. 1.

Note that $v_i = 1/N$ for any network with $k_i^{\text{in}} = k_i^{\text{out}}$ $(1 \leq i \leq N)$, where $k_i^{\text{out}} \equiv \sum_{j=1}^{N} w_{ij}$ is the outdegree. The undirected network is included in this class of networks. Therefore, the influence has a nontrivial meaning only in directed networks. This situation also holds true in the case of the PageRank; in undirected networks, the PageRank is a linear function of the degree of node [13].

The MA of v_i is given by

$$v_i = \frac{\sum_{j=1}^{N} w_{ij} v_j}{k_i^{\text{in}}} \approx \frac{\sum_{j=1}^{N} w_{ij} \bar{v}}{k_i^{\text{in}}} \propto \frac{k_i^{\text{out}}}{k_i^{\text{in}}}, \tag{2}$$

where $\bar{v} \equiv \sum_{j=1}^{N} v_j / N = 1/N$. In Secs. 3 and 4, we argue that Eq. (2) does not satisfactorily describe v_i in certain practically important types of networks, including the Watts-Strogatz small-world network. For these networks, we calculate the exact v_i by using different methods.

The value of v_i can be associated with the number of directed spanning trees rooted at node i, as described below. Equation (1) implies that v_i is the left eigenvector of the Laplacian matrix L, whose (i, j) element is equal to $L_{ii} = \sum_{j=1, j \neq i}^{N} w_{ji}$ and $L_{ij} = -w_{ji}$ $(i \neq j)$. The corresponding eigenvalue is equal to 0; $\sum_{i=1}^{N} v_i L_{ij} = 0$ $(1 \leq j \leq N)$. The (i, j) cofactor of L is given as

$$D(i, j) \equiv (-1)^{i+j} \det L(i, j), \tag{3}$$

where $L(i, j)$ is an $(N - 1) \times (N - 1)$ matrix obtained by deleting the i-th row and the j-th column of L. Because $\sum_{j=1}^{N} L_{ij} = 0$ $(1 \leq i \leq N)$, $D(i, j)$ is independent of j. Therefore, considering the fact that L has eigenvalue 0, we obtain

$$\sum_{i=1}^{N} D(i, i) L_{ij} = \sum_{i=1}^{N} D(i, j) L_{ij}$$

$$= \det L = 0, \quad (1 \leq j \leq N). \tag{4}$$

3

Equation (4) indicates that $(D(1,1),\ldots,D(N,N))$ is the left eigenvector of L with eigenvalue 0. Therefore, we obtain

$$v_i \propto D(i,i) = \det L(i,i). \tag{5}$$

According to the matrix tree theorem [15], $\det L(i,i)$ is equal to the sum of the weights of all directed spanning trees of G rooted at node i. The weight of a spanning tree is defined as the product of the weights of the $N-1$ links used in the spanning tree. Therefore, we can calculate v_i by enumerating the spanning trees.

3   Calculation of influence in some model networks

In this section, we analytically calculate the influence for several networks. Through these calculations, we show that the influence extracts the globally important nodes, which is beyond the scope of the MA: $k_i^{\text{out}}/k_i^{\text{in}}$. Such nodes are located upstream in the hierarchy that is defined by the directionality of the links or around the source of a valuable directed shortcut.

3.1   Weighted chain

Consider a weighted chain of N nodes, as shown in Fig. 2(a). There is a link from node i to node $i+1$ with the weight $w_{i,i+1} > 0$ for each $1 \le i \le N-1$. There is a link from node i to node $i-1$ with the weight $w_{i,i-i} > 0$ for each $2 \le i \le N$. Generally, $w_{i,i+1}$ is not equal to $w_{i+1,i}$. There are no other links. There is only one spanning tree rooted at each node i, which is represented by $1 \leftarrow 2 \leftarrow \ldots \leftarrow i-1 \leftarrow i \rightarrow i+1 \rightarrow \ldots N-1 \rightarrow N$, where the arrow denotes either a directed link or a directed path without confusion. Therefore,

$$v_i = \frac{w_{2,1} w_{3,2} \ldots, w_{i,i-1} w_{i,i+1} w_{i+1,i+2} \ldots w_{N-1,N}}{\mathcal{N}}, \tag{6}$$

where the normalization constant is given by

$$\mathcal{N} = \sum_{i=1}^{N} w_{2,1} w_{3,2} \ldots, w_{i,i-1} w_{i,i+1} w_{i+1,i+2} \ldots w_{N-1,N}. \tag{7}$$

Note that the position of the node, whether located in the middle or the periphery of the chain, does not affect the value of v_i.

Consider a special case where $w_{1,2} = w_{2,3} = \ldots = w_{N-1,N} = 1$ and $w_{2,1} = w_{3,2} = \ldots = w_{N,N-1} = \epsilon$ (Fig. 2(b)). For this network, we obtain

$$v_i = \frac{\epsilon^{i-1}(1-\epsilon)}{(1-\epsilon^N)}. \tag{8}$$

4

When ϵ is small, node i having a small i is more influential. With the normalization constant neglected, the MA yields $k_1^{\mathrm{out}}/k_1^{\mathrm{in}} = 1/\epsilon$, $k_2^{\mathrm{out}}/k_2^{\mathrm{in}} = \ldots = k_{N-1}^{\mathrm{out}}/k_{N-1}^{\mathrm{in}} = 1$, and $k_N^{\mathrm{out}}/k_N^{\mathrm{in}} = \epsilon$. The MA is inconsistent with Eq. (8), except under the limit $\epsilon \to 0$, in which case $v_1 \approx 1$, $v_2, \ldots, v_N \approx 0$.

3.2 Weighted cycle

Consider a weighted cycle having N nodes, as depicted in Fig. 2(c). The weighted cycle is constructed by adding two links $N \to 1$ and $1 \to N$ with the weights $w_{N,1}$ and $w_{1,N}$, respectively, to the weighted chain.

In this network, there are N spanning trees rooted at node i, i.e., $j \leftarrow j+1 \leftarrow \ldots i-1 \leftarrow i \to i+1 \to \ldots \to j-1$, where $1 \leq j \leq N$, nodes $N+1$ and 0 are identified with nodes 1 and N, respectively. Therefore, we obtain

$$
\begin{aligned}
v_i \quad \propto \quad & w_{i,i+1}w_{i+1,i+2}\ldots w_{i-2,i-1} + w_{i,i-1}w_{i,i+1}\ldots w_{i-3,i-2} \\
+ \quad & w_{i-1,i-2}w_{i,i-1}w_{i,i+1}\ldots w_{i-4,i-3} + \ldots,
\end{aligned} \tag{9}
$$

where $w_{N,N+1} \equiv w_{N,1}$ and $w_{1,0} \equiv w_{1,N}$.

In the weighted chain, only the weights of the descending links, i.e., $w_{j,j+1}$ for $j \geq i$ and $w_{j+1,j}$ for $j+1 \leq i$, contribute to v_i. In contrast, in the weighted cycle, both $w_{j,j+1}$ and $w_{j+1,j}$ $(j, j+1 \neq i)$ contribute to v_i. Therefore, in the weighted cycle, the effect of each link weight on v_i is more blurred than that in the case of the weighted chain. This property comes from the fact that node i and node j $(i \neq j)$ are connected in two ways, i.e., clockwise and anticlockwise.

As a special case, consider a directed cycle in which $w_{2,1} = w_{3,2} = \ldots = w_{N,N-1} = w_{1,N} = 0$, $w_{1,2} = w_{2,3} = \ldots = w_{N-1,N} = 1$, and $w_{N,1} = \epsilon$ (Fig. 2(d)). In this network, the values of the influence are equal to

$$
v_1 \quad = \quad \frac{1}{1 + (N-1)\epsilon}, \tag{10}
$$

$$
v_2 = \ldots = v_N \quad = \quad \frac{\epsilon}{1 + (N-1)\epsilon}. \tag{11}
$$

The MA, which yields $k_1^{\mathrm{out}}/k_1^{\mathrm{in}} = 1/\epsilon$, $k_2^{\mathrm{out}}/k_2^{\mathrm{in}} = \ldots = k_{N-1}^{\mathrm{out}}/k_{N-1}^{\mathrm{in}} = 1$, and $k_N^{\mathrm{out}}/k_N^{\mathrm{in}} = \epsilon$, is inconsistent with Eq. (11), except when $\epsilon \to 0$.

3.3 Directed multipartite network

Consider the directed L-partite network, as schematically shown in Fig. 3(a). Layer ℓ $(1 \leq \ell \leq L)$ contains N_ℓ nodes. Each node in layer ℓ sends directed links to all the $N_{\ell+1}$ nodes in layer $\ell + 1$, where layer $L + 1$ is identified as layer 1. Because of symmetry, all nodes in layer ℓ have the same value of influence, denoted by v_ℓ. From Eq. (1), we obtain

$$
N_{\ell-1}v_\ell = N_{\ell+1}v_{\ell+1}, \quad (1 \leq \ell \leq L), \tag{12}
$$

where $N_0 \equiv N_L$. By combining Eq. (12) with the normalization condition $\sum_{\ell=1}^{L} N_\ell v_\ell = 1$, we obtain

$$v_\ell = \frac{1}{N_{\ell-1} N_\ell \sum_{\ell'=1}^{L} N_{\ell'}^{-1}}. \tag{13}$$

3.3.1 Super-star

The super-star, which was introduced in [16] to study the fixation probability in networks, is a variant of the directed multipartite network. The super-star shown in Fig. 3(b) is generated as a superposition of a certain number of identical directed multipartite networks with $N_1 = N_3 = N_4 = \ldots = N_L = 1$ and $N_2 = z \ (\geq 1)$. Each multipartite network is called a leave. The leaves are superposed such that they share a single node in layer 1. The indegree and outdegree of this node are equal to the number of leaves.

It can be easily shown that v_i is independent of the number of leaves. Therefore, we consider the case of a single leave. Then, Eq. (13) yields

$$v_1 = v_4 = v_5 = \ldots = v_L \quad = \quad \frac{z}{z(L-1)+1}, \tag{14}$$

$$v_2 = v_3 \quad = \quad \frac{1}{z(L-1)+1}. \tag{15}$$

Surprisingly, the node in layer 1 does not have a particularly large influence value. Given $z \geq 2$, the nodes in the expanded layer (i.e., layer 2) and the node that receives convergent links from this layer (i.e., layer 3) have small influence values. These relationships are not predicted by the MA. The MA yields $k_2^{\text{out}}/k_2^{\text{in}} = k_4^{\text{out}}/k_4^{\text{in}} = k_5^{\text{out}}/k_5^{\text{in}} = \ldots = k_N^{\text{out}}/k_N^{\text{in}} = 1$, $k_1^{\text{out}}/k_1^{\text{in}} = z$, and $k_3^{\text{out}}/k_3^{\text{in}} = 1/z$; the actual v_1 and v_2 values are essentially smaller than the values predicted by the MA.

3.3.2 Funnel

The funnel, shown in Fig. 3(c), was introduced in [16] along with the super-star; it is also a directed multipartite network. The funnel has $N_\ell = z^{L-\ell}$ nodes in layer ℓ $(1 \leq \ell \leq L)$. Using Eq. (13), we obtain

$$v_\ell = \begin{cases} \dfrac{z-1}{z^L-1}, & (\ell = 1), \\[2mm] \dfrac{z-1}{z^L-1} z^{2\ell-L-2}, & (2 \leq \ell \leq L). \end{cases} \tag{16}$$

The nodes in layer L are most influential, and the node in layer 2 is least influential. The nodes in layer 1 are intermediately influential. For a large z, they are as influential as a node in layer $\approx L/2$. These relationships are not predicted by the MA, which yields the followin results: $k_1^{\text{out}}/k_1^{\text{in}} = k_L^{\text{out}}/k_L^{\text{in}} = z^{L-2}$ and $k_2^{\text{out}}/k_2^{\text{in}} = \ldots = k_{L-1}^{\text{out}}/k_L^{\text{in}} = z^{-2}$.

3.4   Directionally biased random network

The networks considered in the previous sections have inherent global directionality due to the presence of asymmetrically weighted or unidirectional links from node to node or from layer to layer. The directionality of networks is a main cause for the deviation in the v_i values from those predicted by the MA. To examine the effect of directionality in further detail, we study the directionally biased random network [17]. To generate a network from this model, we prepare a strongly connected directed random graph with mean indegree and mean outdegree z and specify a root node, which is placed in layer 1. The root node is the source of directed links to about z nodes, which are placed in layer 2. We align all the nodes according to their distance from the root. Except in the layers near the last layer, the number of nodes in layer ℓ grows roughly as $z^{\ell-1}$. We set the weights of the forward links, i.e., links from layer ℓ to layer $\ell+1$, as unity. We set the weights of the backward links, i.e., links from layer ℓ to layer ℓ', where $\ell > \ell'$, as ϵ. The weights of the parallel links, i.e., those connecting two nodes in the same layer, are arbitrary; they do not affect the value of v_i in the following derivation. When $\epsilon = 1$, the network is an unweighted directed random graph, if the weight of the parallel link is equal to unity at $\epsilon = 1$. When $\epsilon = 0$, the network is purely feedforward and no longer strongly connected. The feedforwardness is parameterized by ϵ.

The directionally biased random network is approximated by using a modified tree as follows [17]. We assume that each node has z outgoing links and that each node except the root node has only one "parent" node, namely, the node in the previous layer from where it receives a feedforward link. Further assume that there are L layers and that layer ℓ ($1 \leq \ell \leq L$) has $z^{\ell-1}$ nodes. The number of nodes is equal to $(z^L - 1)/(z - 1)$. At this point, the constructed network is a tree. Then, we add backward links with weight ϵ to this tree. When z is large, most backward links in the original network originate from layer L, because layer L has a majority of nodes. Therefore, we assume that, in the approximated network, the backward links with weight ϵ originate only from the the nodes in layer L. This approximation is accurate when z is sufficiently large. The other links in the approximated network have the weight of unity. For a sufficiently large z, all nodes in the same layer have almost the same connectivity pattern. In terms of incoming links, a node receives approximately one forward link from the previous layer and z backward links from layer L. The approximated network is schematically shown in Fig. 4. On an average, each node in layer L is the source of an directed link to each node with an effective weight ϵ'. Because $k_i^{\text{out}} = \epsilon z$ for a node in layer L is approximated by $\epsilon'(z^L-1)/(z-1) \approx \epsilon' z^{L-1}$, we obtain $\epsilon' \approx \epsilon z^{-L+2}$.

The influence of a node in layer ℓ, denoted by v_ℓ, satisfies the following

relationships:

$$z^{L-1}\epsilon' v_1 = zv_2, \tag{17}$$

$$(z^{L-1}\epsilon' + 1)v_\ell = zv_{\ell+1}, \quad (2 \le \ell \le L-1). \tag{18}$$

On substituting $\epsilon' \approx \epsilon z^{-L+2}$ in Eq. (18) and considering the normalization given by

$$\sum_{\ell=1}^{L} z^{\ell-1} v_\ell = 1, \tag{19}$$

we obtain

$$v_\ell \approx \begin{cases} (\epsilon z + 1)^{-L+1}, & (\ell = 1), \\ (\epsilon z + 1)^{-L+\ell-1} z^{-\ell+2}\epsilon, & (2 \le \ell \le L). \end{cases} \tag{20}$$

For a small ϵ, the network is close to feedforward, and v_1 is relatively large; v_ℓ varies as $v_\ell \propto (\epsilon + z^{-1})^\ell$. When z is sufficiently large, we obtain $v_\ell \propto \epsilon^\ell$, which coincides with the results obtained for the network shown in Fig. 2(b) (Sec. 3.1).

To test our theory, we generate a directionally biased random network with $N = 5000$ and $z = 10$. For $\epsilon = 0.5$, the values of v_i of all the nodes are plotted against the values obtained from the MA in Fig. 5(a). Although the values obtained from the MA are strongly correlated with v_i, there is some variation in v_i for a fixed $k_i^{\text{out}}/k_i^{\text{in}}$. The average and the standard deviation of v_i in each layer are plotted by the circles and the corresponding error bars, respectively, in Fig. 5(b). The influence of a node decreases exponentially with ℓ, as predicted by Eq. (20) (Eq. (20) for $\ell \ge 2$ is represented by the line in Fig. 5(b)). The average and the standard deviation of v_i obtained by the MA are plotted by the squares and the corresponding error bars, respectively. The values obtained from the MA are scaled by a multiplicative factor C, where C is selected such that $v_i = Ck_i^{\text{out}}/k_i^{\text{in}}$ for the root node (i.e., $\ell = 1$). Figure 5(b) shows that $k_i^{\text{out}}/k_i^{\text{in}}$ is generally small for node i in a downstream layer. However, the decrease in v_i with ℓ is much more than that in $k_i^{\text{out}}/k_i^{\text{in}}$. The hierarchical nature of the network is revealed by the v_i values and not satisfactorily by the local degree. The results for $\epsilon = 0.1$ shown in Figs. 5(c) and (d) provide further evidence for our claim.

3.5 Small-world networks

In this section, we analyze the influence in the directed unweighted small-world network model, which is a variant of the Watts-Strogatz model [14]. To generate a network, we start with an undirected cycle of N nodes, in which each node is connected to its immediate neighbor on both sides. At this stage, $k_i^{\text{in}} = k_i^{\text{out}} = 2$ is satisfied for all i. Then, we add a directed

shortcut to the network, as schematically shown in Fig. 6(a). The source and the target of the shortcut are denoted by nodes s and t, respectively. The distance between node s and node t along the cycle is assumed to be $\min(N_1, N_2)$, where $N_2 \equiv N - N_1$.

We enumerate the number of directed spanning trees rooted at node r, which is $\overline{N}$ nodes away from node s along the cycle, where $0 \leq \overline{N} \leq \max(N_1, N_2)$. There are N spanning trees that do not use the shortcut, as derived in Sec. 3.2. Any spanning tree that uses the shortcut includes the directed path $r \to \ldots \to s$, which contains $\overline{N} + 1$ nodes. The choice of the other links is arbitrary with the restriction that a spanning tree must be formed. The $N_1 - \overline{N} - 1$ nodes between node r and node t in Fig. 6(a) are reached from node r or node t by a directed path along the cycle. There are $N_1 - \overline{N}$ choices regarding the formation of this part of the spanning tree. The $N_2 - 1$ nodes between node s and node t are reached from node s or node t by a directed path along the cycle. There are N_2 choices regarding the formation of this part of the spanning tree. In sum, there are $N + (N_1 - \overline{N})N_2$ spanning trees rooted at v. Therefore, the influence of v is large (small) for the v that is close to the source (target) of the shortcut. In the region on the cycle where no source or target of a shortcut is located, the influence of a node changes linearly with the distance between the source of the shortcut and the node, because $N + (N_1 - \overline{N})N_2 \propto \overline{N}$. We call such a region, including the two border points, the segment.

Next, we consider small-world networks with two directed shortcuts. There are three qualitatively different possible arrangements of the shortcuts, as shown in Figs. 6(b), 6(c), and 6(d). The lengths of the four segments are denoted by N_1, N_2, N_3, and N_4, such that $N_1 + N_2 + N_3 + N_4 = N$.

In the network shown in Fig. 6(b), a node is located at either of the three essentially different positions denoted by a, b, and c. We first enumerate spanning trees rooted at node a. The distance from node a to the source of a shortcut, i.e., node s, is denoted by $\overline{N}$ $(0 \leq \overline{N} \leq N_1)$. There are N spanning trees that do not use the shortcuts. There are $(N_1 - \overline{N})(N_2 + N_3 + N_4)$ spanning trees that use the shortcut $s \to t$ but not $s' \to t'$. There are $(N_1 + N_2 - \overline{N})N_3$ spanning trees that use the shortcut $s' \to t'$ but not $s \to t$. There are $(N_1 - \overline{N})N_2 N_3$ spanning trees that use both shortcuts, which can be explained as follows. The directed path $a \to s \to s'$ is included in such a spanning tree. The $N_1 - \overline{N} - 1$ nodes between node a and node t are reached along the cycle from node a or node t. The $N_2 - 1$ nodes between node t and node t' are reached along the cycle from node t or node t'. The $N_3 - 1$ nodes between node t' and node s' are reached along the cycle from node t' or node s'. In sum, the number of spanning trees rooted at node a is equal to

$$N + (N_1 - \overline{N})(N_2 + N_3 + N_4) + (N_1 + N_2 - \overline{N})N_3 + (N_1 - \overline{N})N_2 N_3, \quad (21)$$

which is proportional to the influence of node a. If N is sufficiently large and

9

the two shortcuts are randomly placed, the last term in Eq. (21) is of the highest order because $N_i = O(N)$ $(1 \leq i \leq 4)$. As in the case of the network with one shortcut, the influence changes linearly within one segment.

Similarly, the number of spanning trees rooted at node b is equal to

$$N + N_1(N_2 + N_3 + N_4 - \overline{N}) + (N_1 + N_2 + \overline{N})N_3 + N_1N_2N_3, \qquad (22)$$

where $\overline{N}$ $(0 \leq \overline{N} \leq N_4)$ is the distance from node b to node s along the cycle. The number of spanning trees rooted at node c is equal to

$$N + (N_2 - \overline{N})N_3 + N_1\overline{N}, \qquad (23)$$

where $\overline{N}$ $(0 \leq \overline{N} \leq N_2)$ is the distance from node c to node t along the cycle. Owing to the absence of a third-order term in Eq. (23), the influence values of the nodes located on the segment between the two targets of the shortcuts are very small.

The quantities given in Eqs. (21), (22), and (23) are linear in $\overline{N}$. Therefore, the influence changes linearly within each segment. This is true for the other two types of arrangements of shortcuts shown in Figs. 6(c) and 6(d).

This linear relationship also holds true in the case of more than two shortcuts. To show this, we consider a general directed small-world network and focus on a segment on the cycle, which is schematically shown in Fig. 6(e). Without loss of generality, we assume that a node v in the segment is located $\overline{N}$ and $N_1 - \overline{N}$ nodes away from the two border points of the segment. We distinguish three types of spanning trees rooted at node r. First, some spanning trees include both $r \rightarrow \ldots \rightarrow 1$ and $r \rightarrow \ldots \rightarrow 2$. We denote the number of such spanning trees by S_1. Second, some spanning trees include $r \rightarrow \ldots \rightarrow 1$ and not $r \rightarrow \ldots \rightarrow 2$. For these spanning trees, node 2 is reached via a path $r \rightarrow 1 \rightarrow \ldots \rightarrow 2$. To enumerate such spanning trees, denote by S_2 the number of directed trees that span the network excluding the nodes in the segment between node r and node 2. Third, the other spanning trees include $r \rightarrow \ldots \rightarrow 2$ and not $r \rightarrow \ldots \rightarrow 1$. Denote by S_3 the number of directed trees that span the network excluding the nodes in the segment between node r and node 1. The number of spanning trees rooted at node r is equal to

$$S_1 + S_2(N_1 - \overline{N}) + S_3\overline{N}.$$

Therefore, the influence of node r changes linearly with $\overline{N}$ $(0 \leq \overline{N} \leq N_1)$.

The thick line in Fig. 7(a) indicates the numerically obtained values of v_i for a small-world network with three shortcuts. We set $N = 5000$. The nodes are aligned according to their position in the cycle. In accordance with the theoretical prediction, v_i changes linearly with i within each segment. v_i is very small for $1250 \leq i \leq 1666$, because these nodes are between two targets of shortcuts.

In theory, it is assumed that each node is initially connected to only its nearest neighbors on the cycle (i.e., $k_i^{\text{in}} = k_i^{\text{out}} = 2$). However, values of v_i are almost the same if the underlying cycle has $k_i^{\text{in}} = k_i^{\text{out}} = 4$ (i.e., each node is connected to two neighbors on each side) or $k_i^{\text{in}} = k_i^{\text{out}} = 6$. The results for $k_i^{\text{in}} = k_i^{\text{out}} = 4$ and those for $k_i^{\text{in}} = k_i^{\text{out}} = 6$ are indicated by the medium and thin lines, respectively, in Fig. 7(a). The three lines are observed to almost overlap with each other.

To examine the effect of shortcuts in more general small-world networks, we generate a small-world network by rewiring many links [14]. We place $N = 5000$ nodes on a cycle and connect a node to its five immediate neighbors on each side, such that $k_i^{\text{in}} = k_i^{\text{out}} = 10$. Then, out of 50000 directed links, we rewire 500 randomly selected ones to create directed shortcuts. The sources and targets of shortcuts are chosen randomly from the network with the restriction that self loops and multiple links must be avoided. Because of rewiring, the mean degree $\langle k \rangle = 10$.

As shown in Fig. 7(b), the MA strongly disagrees with the observed v_i [9]. The values of v_i are plotted against the circular positions of the nodes in Fig. 7(c). v_i changes gradually along the cycle, which is consistent with our analytical results. The peaks and troughs in Fig. 7(c) correspond to the sources and targets of the shortcuts, respectively. For a node near a source (target), v_i is large (small), whereas the MA estimate $\propto k_i^{\text{out}}/k_i^{\text{in}}$ is not as affected by the position of the shortcuts as v_i. The relationship between $v_i/(k_i^{\text{out}}/k_i^{\text{in}})$ and v_i is shown in Fig. 7(d). The MA is exact along the horizontal line, i.e., $v_i = (k_i^{\text{out}}/k_i^{\text{in}})/(\sum_{j=1}^{N} k_j^{\text{out}}/k_j^{\text{in}})$. Nodes with large (small) v_i tend to be located near sources (targets) of shortcuts. For such nodes, v_i is usually larger (smaller) than the value obtained by the MA.

4 Entrainment of a network by a pacemaker

As an application of the influence as a centrality measure, we examine a system of coupled phase oscillators having a pacemaker [17, 18]. Consider a dynamical system of phase oscillators given by

$$\dot{\phi}_i = \omega_i + \frac{\kappa}{\langle k \rangle} \sum_{j=1}^{N} w_{ji} \sin\left(\phi_j - \phi_i\right), \qquad (1 \le i \le N), \qquad (24)$$

where the mean degree $\langle k \rangle = \sum_{i',j'=1}^{N} w_{i'j'}/N$ provides the normalization for the coupling strength κ. The phase and the intrinsic frequency of the i-th oscillator are denoted by $\phi_i \in [0, 2\pi)$ and ω_i, respectively. We assume a pacemaker, i.e., an oscillator that is not influenced by the other oscillators, in the network. Equation (24) emulates a pacemaker system, where the pacemaker is placed at node i_0, if we force $w_{ji_0} = 0$ ($1 \le j \le N$). We examine the possibility of the pacemaker to entrain the other oscillators

into its own intrinsic rhythm. For this, we assume that $\omega_i = \omega$ $(i \neq i_0)$ is identical for the $N - 1$ oscillators and that ω_{i_0} takes a different value. By redefining $\phi_i - \omega t$ as the new ϕ_i and rescaling time, we set $\omega_{i_0} = 1$ and $\omega_i = 0$ $(i \neq i_0)$ without loss of generality.

Depending on the network and the position of the pacemaker, there exists a critical threshold κ_{cr} such that the entrainment is realized for $\kappa > \kappa_{cr}$ [18, 17]. When entrained, the actual frequency of all oscillators becomes exactly the same as that of the pacemaker, i.e., $\omega_{i_0} = 1$. Thus, the condition for the entrainment is given by

$$\dot{\phi}_i = \frac{\kappa}{\langle k \rangle} \sum_{j=1}^{N} w_{ji} \sin\left(\phi_j - \phi_i\right) = 1, \qquad (1 \leq i \leq N, i \neq i_0). \qquad (25)$$

In general, entrainment [17, 19] and synchrony [20] are easy realized for feedforward networks. Because of the intuitive meaning of the influence, κ_{cr} may be small if v_{i_0} is large. We analytically show this for the directionally biased random network with a sufficiently large z. In the directionally biased random network, a node in layer ℓ receives a forward link with weight unity from layer $\ell - 1$. Although a forward link is absent for a node in layer 1, this factor is negligible because layer 1 contains only one node. Most backward links with weight ϵ to a node in layer ℓ, where $\ell \leq L - 1$, originate from layer L, as discussed in Sec. 3.4. The number of parallel links is smaller than that of backward links. We assume that the weight of the parallel link, which was assumed to be arbitrary in Sec. 3.4, is equal to ϵ, such that all the incoming links to a node in layer L, except one forward link, also have weight ϵ. Under this condition, we approximate $\langle k \rangle \approx 1 + \epsilon z$.

Denote by $\kappa_{cr}^{(\ell)}$ the typical critical coupling strength when the pacemaker is located at a node in layer ℓ in a directionally biased random network. First, we consider the case in which node i_0 coincides with the root node in the directionally biased random network, i.e., $\ell = 1$. This case was analyzed in our previous studies [17, 18]. In the entrained state, the phase difference between the oscillators in the same layer is small, and the phases of the oscillators in layers with small ℓ are more advanced. Therefore, we assume that the phases of all the oscillators in the same layer are identical. We set the difference between the phase of the oscillator in layer ℓ and that in layer $\ell + 1$ to $\Delta\phi_\ell$. The entrainment occurs if and only if $\Delta\phi_1$, ..., $\Delta\phi_{L-1}$ stay constant in a long run. We obtain [17, 18]

$$\Delta\phi_\ell = \frac{(1 + \epsilon z)^{L-\ell}}{\kappa}, \qquad (1 \leq \ell \leq L - 1). \qquad (26)$$

By applying the threshold condition $\Delta\phi_1 = 1$, we obtain

$$\kappa_{cr}^{(1)} = (1 + \epsilon z)^{L-1}. \qquad (27)$$

Next, we consider the case in which node i_0 is located in layer L. We expect $\kappa_{\mathrm{cr}}^{(L)}$ to be large because the pacemaker is located downstream of the network. To analyze this case, we redraw the network as a directionally biased random network, such that node i_0 is located at the root. Then, statistically, the network is the same as the original directionally biased random network in terms of the positions of the nodes and the links. However, the effective link weight in the redrawn network is equal to ϵ, because most links in the original network are backward links with weight ϵ. By assuming that all links in the redrawn network have weight ϵ, the result for unweighted directed random graph [17, 18] translates into

$$\Delta\phi_{\bar{\ell}} = \frac{(1+\epsilon z)(1+z)^{L-\bar{\ell}-1}}{\kappa\epsilon}, \qquad (1 \le \bar{\ell} \le L-1), \qquad (28)$$

where $\bar{\ell}$ is the layer number in the redrawn network. Analogous to the derivation of Eq. (27) from Eq. (26), from Eq. (28), we derive

$$\kappa_{\mathrm{cr}}^{(L)} = \frac{(1+\epsilon z)(1+z)^{L-2}}{\epsilon}. \qquad (29)$$

When node i_0 is located in the $(L-M+1)$-th layer $(2 \le M \le L-1)$ in the original network, we redraw the network in a similar manner, such that node i_0 is located at the root. The redrawn network is schematically shown in Fig. 9. For this network, we obtain

$$\Delta\phi_{\bar{\ell}} = \begin{cases} \frac{(1+z)^{L-M}(1+\epsilon z)^{M-\bar{\ell}}}{\kappa}, & (1 \le \bar{\ell} \le M-1), \\ \frac{(1+\epsilon z)(1+z)^{L-\bar{\ell}-1}}{\kappa\epsilon}, & (M \le \bar{\ell} \le L-1), \end{cases} \qquad (30)$$

which yields
$$\kappa_{\mathrm{cr}}^{(L-M+1)} = (1+\epsilon z)^{M-1}(1+z)^{L-M}. \qquad (31)$$

From Eqs. (27), (29), and (31), we obtain

$$\kappa_{\mathrm{cr}}^{(\ell)} = \begin{cases} (1+\epsilon z)^{L-\ell}(1+z)^{\ell-1}, & (1 \le \ell \le L-1), \\ \frac{(1+\epsilon z)(1+z)^{L-2}}{\epsilon}, & (\ell = L). \end{cases} \qquad (32)$$

Under the condition $\epsilon z \gg 1$, Eq. (32) gives

$$\kappa_{\mathrm{cr}}^{(\ell)} \approx \epsilon^{L-\ell} z^{L-1}, \qquad (1 \le \ell \le L). \qquad (33)$$

Moreover, Eq. (20) yields

$$v_\ell \approx \epsilon^{-L+\ell} z^{-L+1}, \qquad (1 \le \ell \le L). \qquad (34)$$

Therefore, we obtain

$$\kappa_{\mathrm{cr}}^{(\ell)} \approx \frac{1}{v_\ell}, \qquad (1 \le \ell \le L). \qquad (35)$$

Equation (35) shows that the pacemaker located at an influential node can easily realize the entrainment. We validate this prediction by direct numerical simulations of the pacemaker system on a directionally biased random network with $N = 200$, $z = 10$, and $\epsilon = 0.1$. To judge whether the entrainment has been realized for a value of κ, we measure the ratio of $\sum_{i=1, i \neq i_0}^{N} [\phi_i (t = T) - \phi_i (t = 0.8T)] / (N - 1)$ to $\phi_{i_0}(t = T) - \phi_{i_0}(t = 0.8T)$, where T is the duration of a run. The first 80% of a run is discarded as transient. The ratio represents the average phase shift of the oscillators, other than the pacemaker, relative to that of the pacemaker. If this value is more than 0.99, we consider the entrainment to be achieved. Because the transient is shorter for larger κ, we set $T = 5 \times 10^5 / \kappa$.

The values of κ_{cr} when the pacemaker is located at different nodes are plotted against v_i in Fig. 8(a). The line in the figure represents $\kappa_{\mathrm{cr}} = v_i^{-1}$, i.e., Eq. (35). The numerically obtained κ_{cr} roughly matches the theoretical one although the condition $\epsilon z \gg 1$ is violated. The same values of κ_{cr} are plotted against the MA estimate $v_i \approx (k_i^{\mathrm{out}}/k_i^{\mathrm{in}}) / \sum_{j=1}^{N} (k_j^{\mathrm{out}}/k_j^{\mathrm{in}})$ in Fig. 8(b). We find a larger spread of data in this plot as compared to that in Fig. 8(a); v_i predicts κ_{cr} more accurately than the MA.

Next, we set the weight of the parallel link to unity, as done in [17]. The values of κ_{cr} for this version of the directionally biased random network are shown in Figs. 8(c) and 8(d). The results are qualitatively the same as those shown in Figs. 8(a) and 8(b). The dependence of κ_{cr} on v_i is weak in the new network (Fig. 8(c)) as compared to the previous network (Fig. 8(a)) mainly for the following reason. Because of the difference in the weights of parallel links, $\langle k \rangle$ in the new network is larger than that in the previous network. Then, $\kappa_{\mathrm{cr}}^{(L)}$ is smaller for the new network, since it is inversely proportional to the effective link weight. We expect that κ_{cr} for nodes in intermediate layers can also be explained using the same approach.

The result $\kappa_{\mathrm{cr}} \approx v_i^{-1}$ is derived for the directionally biased random network. Although it is no guaranteed that this relationship holds true in other types of networks, we test the applicability of the relation $\kappa_{\mathrm{cr}} \approx v_i^{-1}$ in a scale-free network and a neural network.

We generate a directed scale-free network using the configuration model [1]. The degree distributions are independently given for k_i^{in} and k_i^{out} by $p(k^{\mathrm{in}}) \propto k^{-\gamma_{\mathrm{in}}}$ and $p(k^{\mathrm{out}}) \propto k^{-\gamma_{\mathrm{out}}}$, respectively. We set $\gamma_{\mathrm{in}} = \gamma_{\mathrm{out}} = 2.5$ and $N = 200$. The minimum degree is set to 3. The duration of a run and the length of the traisient are equal to those in the case of the directionally biased random networks. The values of κ_{cr} are plotted against v_i and the MA in Figs. 8(e) and 8(f), respectively. The relation $\kappa_{\mathrm{cr}} \approx v_i^{-1}$ fits the data reasonably well, even though the scale-free network is not a directionally biased random network. In contrast, $k_i^{\mathrm{out}}/k_i^{\mathrm{in}}$ poorly predicts κ_{cr}, as shown in Fig. 8(f).

We next examine the *C. elegans* neural network [21, 22] based on chem-

ical synapses, which serve as directed links. The network has the largest strongly connected component with $N = 237$ nodes. The number of synapses from neuron i to neuron j defines w_{ij}. We set $T = 2.5 \times 10^7/\kappa$ to appropriately exclude the transient. The values of κ_{cr} are plotted against v_i and the MA in Figs. 8(g) and 8(h), respectively. The values of κ_{cr} exceeding 10^7 are not plotted because direct numerical simulations need too much time. Similar to the case of scale-free networks, v_i predicts κ_{cr} better than $k_i^{\mathrm{out}}/k_i^{\mathrm{in}}$ does.

5 Conclusions

In this study, we have analyzed the centrality measure called the influence in various networks. The influence extracts the magnitude with which a node controls or impacts the entire network along directed links. We have analytically shown that the source of the shortcut in a directed version of the Watts-Strogatz small-world network and the root node in hierarchical networks have large influence values. This is not accurately predicted if we approximate the influence of a node by its degree. Although by definition, the influence is based on local connectivity, the global structure of networks does affect the influence values. We also analyzed the effect of the location of a pacemaker on the capability of entrainment in a system of coupled phase oscillators. The pacemaker located at a node with a large influence value entrains the other oscillators relatively easily.

In the analysis of some model networks, including the Watts-Strogatz small-world network, we used the method based on the enumeration of directed spanning trees. This method can be applied to the estimation of the PageRank of nodes, because the PageRank can be mapped to the influence if we reverse links and rescale the link weight [7, 9]. Application of our results to other centrality measures for directed networks is warranted for future study.

acknowledgments

References

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002); M. E. J. Newman, SIAM Rev. **45**, 167 (2003); G. Caldarelli, Scale-free networks, Oxford University Press, Oxford (2007).

[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[3] L. C. Freeman, Soc. Netw. **1**, 215 (1979).

[4] S. Wasserman and K. Faust, Social Network Analysis. Cambridge University Press, NY, USA. 1994

[5] J. D. Noh and H. Rieger, Phys. Rev. Lett. **92**, 118701 (2004); M. E. J. Newman, Soc. Netw. **27**, 39 (2005); E. Estrada and J. A. Rodríguez-Velázquez, Phys. Rev. E **71**, 056103 (2005); V. Latora and M. Marchiori, New J. Phys **9**, 188 (2007).

[6] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. Lett. **97**, 094102 (2006).

[7] N. Masuda, Y. Kawamura, and H. Kori, arXiv:0909.0700v1 (2009).

[8] H. Kori, Y. Kawamura, H. Nakao, K. Arai, and Y. Kuramoto, Phys. Rev. E, in press (2009).

[9] N. Masuda and H. Ohtsuki, New J. Phys. **11**, 033012 (2009)

[10] P. D. Taylor, Amer. Natur. **135**, 95 (1990); J. Math. Biol. **34**, 654 (1996).

[11] M. H. DeGroot, J. Am. Stat. Asso. **69**, 118 (1974); N. E. Friedkin, Am. J. Sociol. **96**, 1478 (1991).

[12] http://pajek.imfm.si/doku.php

[13] S. Brin and L. Page, Proc. 7th Int. World Wide Web Conf. (Brisbane, Australia, 14–18 April) p 107–117 (1998); P. Berkhin, Internet Math. **2**, 73 (2005).

[14] D. J. Watts, S. H. Strogatz, Nature (London) **393**, 440 (1998).

[15] N. Biggs, Bull. London Math. Soc. **29**, 641 (1997); R. P. Agaev and P. Yu. Chebotarev, Autom. Rem. Cont. **61**, 1424 (2000).

[16] E. Lieberman, C. Hauert, and M. A. Nowak, Nature (London) **433**, 312 (2005).

[17] H. Kori and A. S. Mikhailov, Phys. Rev. E **74**, 066115 (2006).

[18] H. Kori and A. S. Mikhailov, Phys. Rev. Lett. **93**, 254101 (2004).

[19] N. Masuda and H. Kori, J. Comput. Neurosci. **22**, 327 (2007).

[20] D.-U. Hwang, M. Chavez, A. Amann, and S. Boccaletti, Phys. Rev. Lett. **94**, 138701 (2005); T. Nishikawa and A. E. Motter, Phys. Rev. E **73**, 065106(R) (2006); Physica D **224**, 77 (2006); X. Wang, Y.-C. Lai, and C. H. Lai, Phys. Rev. E **75**, 056205 (2007).

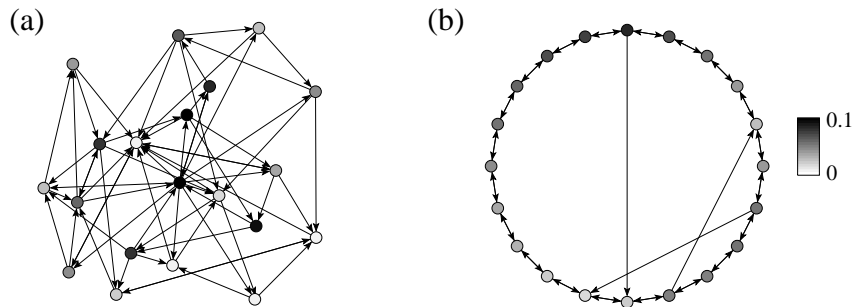[21] B. L. Chen, D. H. Hall, and D. B. Chklovskii, Proc. Natl. Acad. Sci. USA **103**, 4723 (2006).

[22] http://www.wormatlas.org

Figure 1: Influence of nodes in networks having $N = 20$. A dark color of the node represents a large value of v_i. (a) Directed network generated by the configuration model. The indegree and outdegree follow independent power-law distributions with the scaling exponent 2.5 and the minimum degree 2. (b) Directed Watts-Strogatz network with three shortcuts. See Sec. 3.5 and Sec. 4 for details of the network models. The networks are visualized by Pajek [12].

18
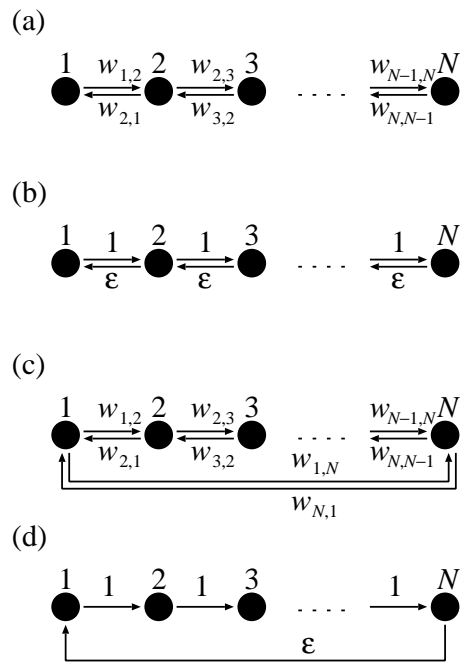
Figure 2: Schematic of (a) weighted chain, (b) special case of weighted chain, (c) weighted cycle, and (d) special case of weighted cycle.
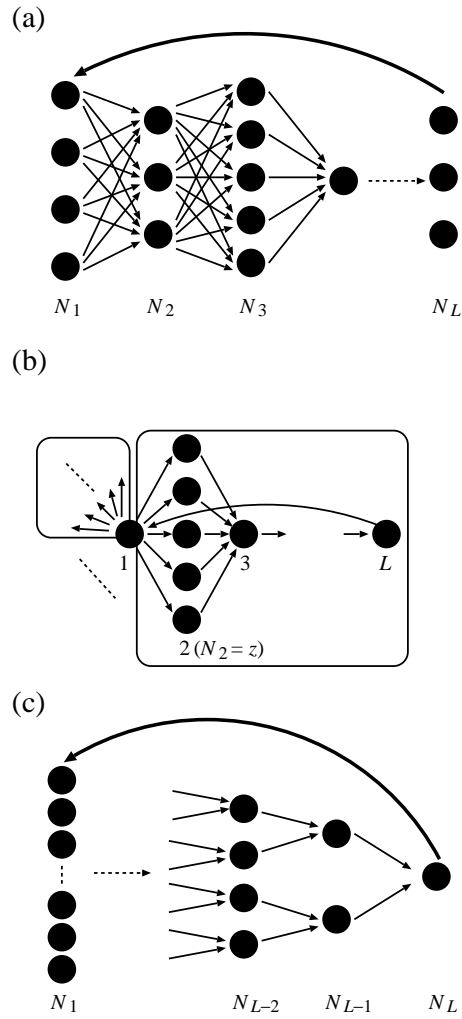
Figure 3: Schematic of (a) multipartite network, (b) super-star, and (c) funnel.
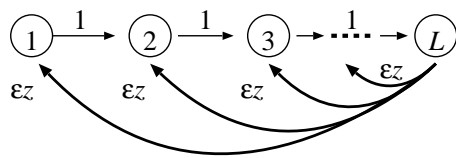
Figure 4: Schematic of directionally biased random network under tree approximation.
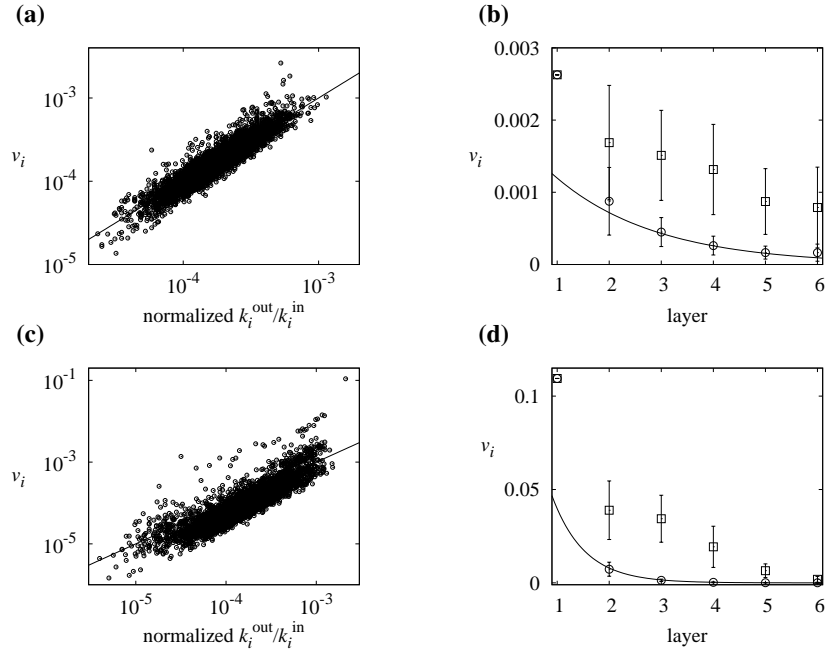
Figure 5: v_i for a directionally biased random network with $N = 5000$. We set (a, b) $\epsilon = 0.5$ and (c, d) $\epsilon = 0.1$. In (a) and (c), v_i is plotted against the MA results. The lines represent the MA: $v_i = (k_i^{\text{out}}/k_i^{\text{in}})/\sum_{j=1}^{N}(k_j^{\text{out}}/k_j^{\text{in}})$. In (b) and (d), v_i (circles) and rescaled $k_i^{\text{out}}/k_i^{\text{in}}$ (squares) averaged over all the nodes in each layer are plotted against layer number. The error bars indicate the standard deviation obtained from v_i of all the nodes in the same layer.
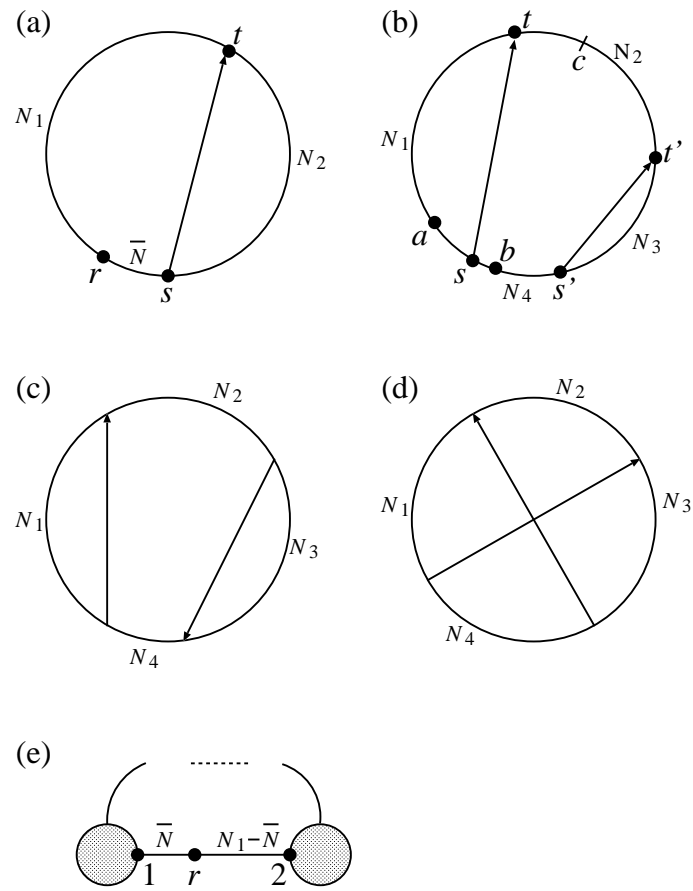
Figure 6: Schematics of directed small-world network having (a) one short-cut, (b, c, d) two shortcuts, and (e) general number of shortcuts.
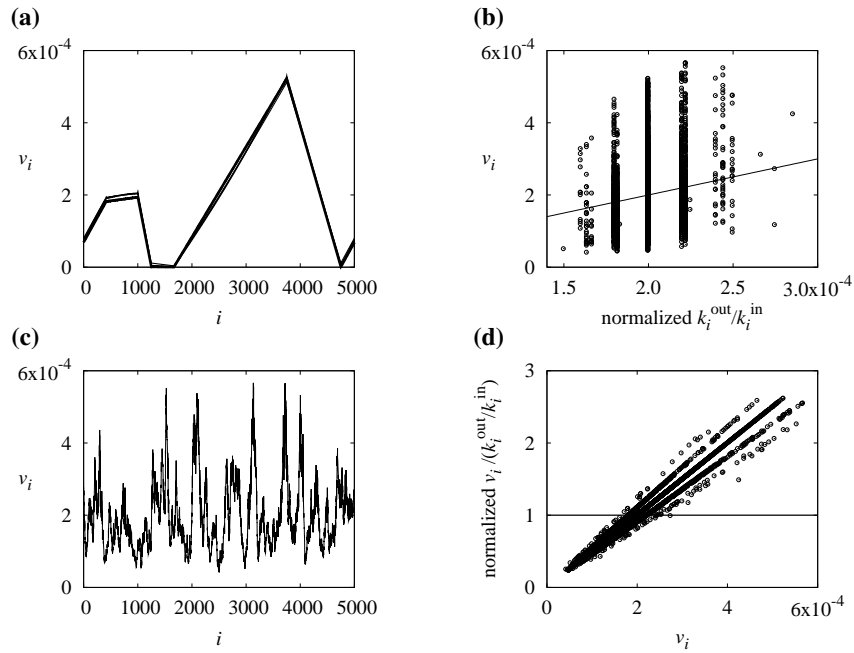
Figure 7: v_i in directed small-world networks with $N = 5000$. (a) Results for small-world network with three added shortcuts. The thick, medium, and thin lines correspond to the networks in which the mean degree of the underlying cycle is equal to 2, 4, and 6, respectively. (b, c, d) Results for directed small-world network with $\langle k \rangle = 10$ and 500 rewired shortcuts. In (a) and (c), v_i is plotted against the position of the node in the underlying cycle. In (b), v_i is plotted against $(k_i^{\text{out}}/k_i^{\text{in}})/\sum_{j=1}^{N}(k_j^{\text{out}}/k_j^{\text{in}})$; the line represents the MA result. In (d), the relationship between $v_i/(k_i^{\text{out}}/k_i^{\text{in}})$ and v_i is shown.
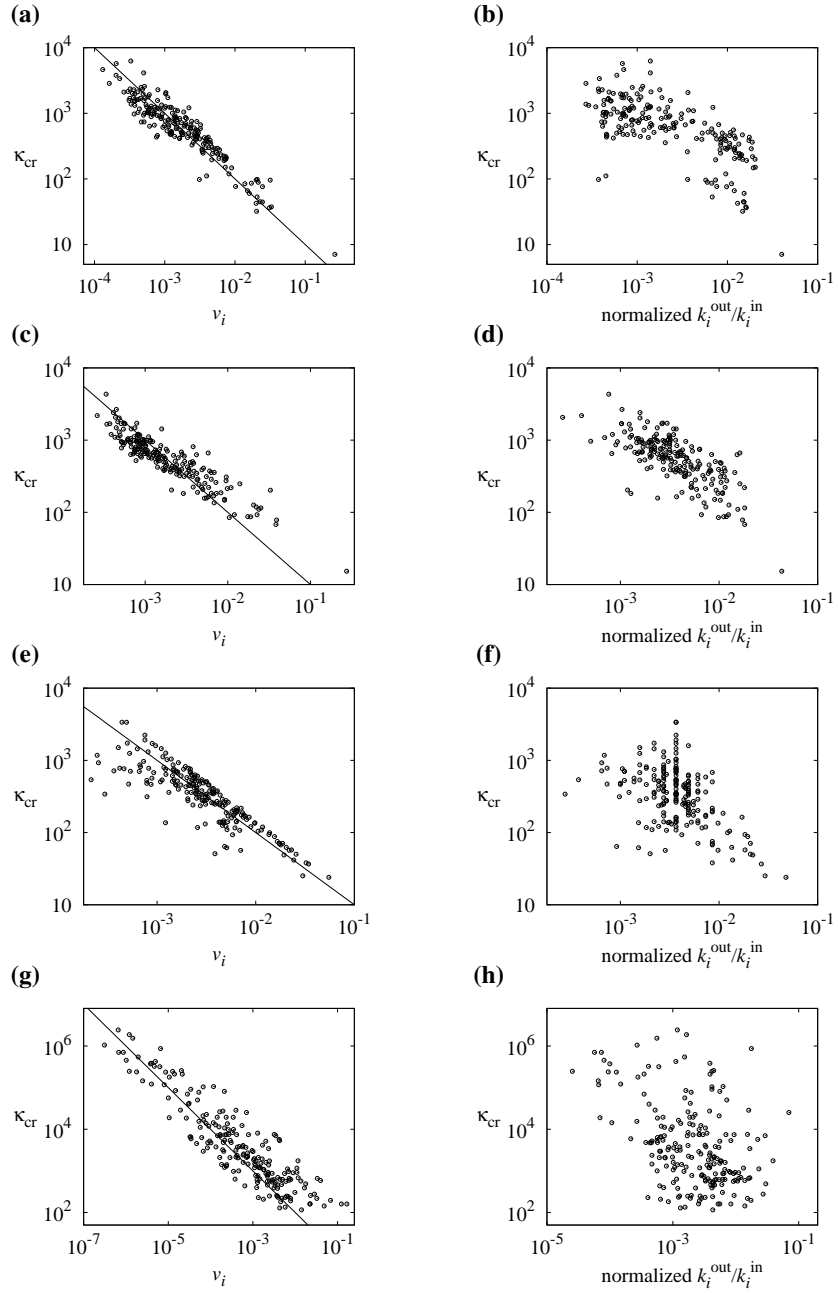
Figure 8: Relationships between κ_{cr} and v_i. (a, b) Directionally biased random network with $N = 200$ and weight of parallel links equal to ϵ. (c, d) Directionally biased random network with $N = 200$ and weight of parallel links equal to unity. (e, f) Directed scale-free network with $N = 200$. (g, h) *C. elegans* neural network with $N = 237$. The data are plotted against v_i in (a, c, e, g) and against $(k_i^{\mathrm{out}}/k_i^{\mathrm{in}})/\sum_{j=1}^{N}(k_j^{\mathrm{out}}/k_j^{\mathrm{in}})$ in (b, d, f, h). The lines in (a, c, e, g) represent $\kappa_{\mathrm{cr}} \approx v_i^{-1}$.
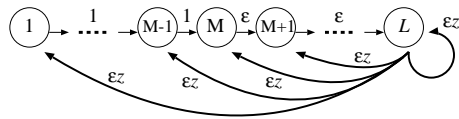
25

Figure 9: Schematic of redrawn directionally biased random network when the pacemaker is located in layer $L - M + 1$ in the original directionally biased random network.