Dynamics-based centrality for directed networks

Naoki Masuda[1,2] and Hiroshi Kori[3,2]
[1]Department of Mathematical Informatics, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8656, Japan
[2]PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan
[3]Division of Advanced Sciences, Ochadai Academic Production, Ochanomizu University,
2-1-1 Ohtsuka, Bunkyo-ku, Tokyo 112-8610, Japan

(Received 16 June 2010; revised manuscript received 12 October 2010; published 8 November 2010)

Determining the relative importance of nodes in directed networks is important in, for example, ranking websites, publications, and sports teams, and for understanding signal flows in systems biology. A prevailing centrality measure in this respect is the PageRank. In this work, we focus on another class of centrality derived from the Laplacian of the network. We extend the Laplacian-based centrality, which has mainly been applied to strongly connected networks, to the case of general directed networks such that we can quantitatively compare arbitrary nodes. Toward this end, we adopt the idea used in the PageRank to introduce global connectivity between all the pairs of nodes with a certain strength. Numerical simulations are carried out on some networks. We also offer interpretations of the Laplacian-based centrality for general directed networks in terms of various dynamical and structural properties of networks. Importantly, the Laplacian-based centrality defined as the stationary density of the continuous-time random walk with random jumps is shown to be equivalent to the absorption probability of the random walk with sinks at each node but without random jumps. Similarly, the proposed centrality represents the importance of nodes in dynamics on the original network supplied with sinks but not with random jumps.

DOI: 10.1103/PhysRevE.82.056107 PACS number(s): 89.75.Hc, 89.75.Fb, 64.60.aq, 02.10.Ox

I. INTRODUCTION

A network is a set of nodes and a set of links that connect pairs of nodes (see [1–3] for reviews). In applications including information science, sociology, and biology, it is often necessary to determine important nodes in a network. Various definitions of the importance of nodes, or centrality measures, have been proposed since the first classical studies on social network analysis in the 1950s [2–5].

It is often more suitable to consider links to be directed, where the direction of link represents relationships such as the control of one node over another, unidirectional flow, and citation. Many centrality measures including degree centrality, betweenness centrality, and eigenvector centrality can be adopted to the case of directed networks. Nevertheless, the most popular centrality for directed networks appears to be PageRank, which takes nontrivial values only in directed networks. It was originally developed for ranking websites [6]. In other words, the PageRank of a node is large when the node receives many links from important nodes that do not have too many outgoing links.

In the present study, we focus on another important class of centrality for directed networks, i.e., those derived from the Laplacian of the network. This class of centrality has a long history [7–12] and is mathematically close to the PageRank (see Sec. V). Furthermore, for strongly connected networks, i.e., networks in which there exists a path of directed links between an arbitrary ordered pair of nodes, the Laplacian-based centrality value of a node, which we also call the influence of a node, represents its importance in various dynamics on networks [13–15].

The Laplacian-based centrality measure has mostly been analyzed for strongly connected networks [7–9,13–15]. However, real directed networks may not be strongly connected. This is typically the case when the network is sparse (i.e., number of links is relatively small) or of small size. Although the Laplacian-based centrality in the original form is applicable when all the nodes are reached along directed paths from a certain specified node, such a network is not generic. The Laplacian-based centrality has been generalized to the case of general directed networks [10–12]. In the generalized version, nodes in an uppermost component have positive centrality values, whereas nodes in a downstream component have zero centrality values (see Secs. II and III for definitions of uppermost and downstream components). However, we may want to compare the importance of nodes in downstream components. We may also wish to compare a node 1 in an uppermost component and a node 2 in a downstream component that is not under the control of node 1.

In this paper, we extend the Laplacian-based centrality measure (i.e., influence) to the case of general directed networks. Networks do not have to be strongly connected and can be composed of disconnected components. The extended centrality measure, called as the influence or extended influence without ambiguity, is a one-parameter family of the centrality measure with parameter q such that the previous definition [10–12] is recovered in the limit $q \to 0$. The extended influence is a relative of the PageRank; the influence and the PageRank correspond to continuous-time and discrete-time simple random walks, respectively. The present paper is organized as follows. In Secs. II and III, we review previous works on the influence for strongly connected and general directed networks, respectively. In Sec. IV, we present new interpretations of the centrality measure introduced in Sec. III. In Sec. V, we extend the concept of the influence by borrowing the idea used in the PageRank to introduce some global connectivity to the original network. We also show that the proposed influence can be interpreted as the dynamical properties of nodes on the original network without additional global connectivity. In Secs. VI and VII, we apply the influence to toy examples and relatively large networks, respectively. In Sec. VIII, we summarize and discuss our results, with an emphasis on the comparison of the influence and the PageRank.

II. INFLUENCE FOR NETWORKS WITH SINGLE ZERO LAPLACIAN EIGENVALUE

Consider a directed and weighted network having N nodes. The weight of the link from node i to node j is denoted by w_{ij} and assumed to be nonnegative. $w_{ij}>0$ represents the strength with which node i governs node j. w_{ij} and w_{ji} are generally different from each other.

The Laplacian-based centrality measure, called the influence of node i and denoted as v_i, is defined as the solution of the following set of N linear equations:

$$v_i = \frac{\sum_{j=1}^{N} w_{ij} v_j}{\sum_{j'=1}^{N} w_{j'i}}, \quad (1 \le i \le N). \tag{1}$$

The normalization is given by $\sum_{i=1}^{N} v_i = 1$. We can rewrite Eq. (1) as

$$(v_1 \cdots v_N) L = 0, \tag{2}$$

where $L=(L_{ij})$ is the asymmetric Laplacian defined by

$$L_{ij} = \delta_{ij} \sum_{j' \neq i} w_{j'i} - (1 - \delta_{ij}) w_{ji}. \tag{3}$$

v_i represents the importance of nodes in various dynamics on networks, such as the voter model, a random walk, DeGroot's model of consensus formation, and the response of synchronized networks [13].

If a network is strongly connected, that is, if any node j can be reached from an arbitrary node i along directed links, the Perron-Frobenius theorem guarantees that $(v_1 \cdots v_N)$ is unique and $v_i>0$ $(1 \le i \le N)$. In particular, for undirected networks, which are strongly connected as long as they are connected, we have $v_i = 1/N$. Therefore, the influence is a centrality measure that is relevant only in directed networks.

To discuss the uniqueness of the zero eigenvector $(v_1 \cdots v_N)$ of L, we use the concept of the root node [10]. Consider the set G_{r} of m nodes in a given network $(1 \le m \le N)$. We define G_{r} to be a set of root nodes if an arbitrary node can be reached along directed links from a node included in G_{r} and G_{r} is minimal. In the example shown in Fig. 1, {1, 2} qualifies as G_{r}. {1, 3} is another example of G_{r}. {1, 4, 5} does not qualify because it is not minimal. The minimality indicates that some nodes cannot be reached from G'_{r}, where G'_{r} is the set of nodes with $m-1$ nodes defined by removing an arbitrary node from G_{r}. For strongly connected networks $(m=1)$, G_{r} can be a set of any single node. As this exercise suggests, G_{r} for a given network is generally not unique. However, m is uniquely determined from a network [10]. The directed chain shown

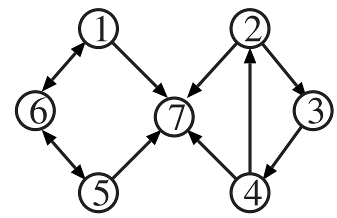

FIG. 1. A network with two root nodes.

in Fig. 2 is a network that is not strongly connected with $m=1$. In Fig. 2, we obtain $v_1=1$ for the unique root node 1 and $v_i=0$ $(2 \le i \le N)$.

The multiplicity of the zero eigenvalue of L, also called the geometric multiplicity of the eigenvalue [16,17], is equal to m [10,18,19]. Therefore, the influence given by Eq. (2) is well defined only for networks with $m=1$, and most previous papers that treat Eq. (2) concentrate on strongly connected networks [7–9,13–15,20]. In this case, $(v_1 \cdots v_N)$ can be readily calculated by the power iteration or the enumeration of the directed spanning tree [13,14].

III. CASE OF MULTIPLE ZERO LAPLACIAN EIGENVALUES

In this section, we treat networks with multiple zero Laplacian eigenvalues. Such a network is not strongly connected. The influence explained in Sec. II was extended to accommodate this case by Agaev–Chebotarev [10,11] and Borm *et al.* [12]. We develop a new centrality measure in Sec. V by generalizing their definitions. In this section, we explain their centrality measure and examine its properties.

Consider a continuous-time simple random walk on the network generated by reversing the direction of all the links of the original network. We select each node i $(1 \le i \le N)$ as the initial location of the random walker with probability $1/N$. For directed networks that are not necessarily strongly connected, Agaev–Chebotarev [10,11] and Borm *et al.* [12] defined a centrality measure, which we call the influence and denote by v_i without ambiguity, as the long-term probability that the walker visits node i. For a strongly connected network, v_i is equal to the stationary density of the random walk and coincides with v_i defined by Eq. (2) [13,20]. For a network with a single root node i_0, node i_0 is the unique absorbing boundary, and any random walker is eventually trapped at node i_0. Therefore, $v_{i_0}=1$ and $v_i=0$ $(i \neq i_0)$, which is again consistent with Eq. (2) [13].

Because the generator of the continuous-time random walk is equal to $-L$, we obtain

$$(v_1 \cdots v_N) = \lim_{t \to \infty} \frac{1}{N} (1 \cdots 1) \exp(-Lt). \tag{4}$$

The spectral decomposition of L yields

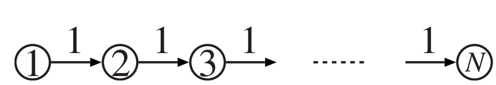

FIG. 2. Directed chain with N nodes.

$$e^{-Lt} = \sum_{m'=1}^{m} \boldsymbol{u}_{m'}^{(0)}\boldsymbol{v}_{m'}^{(0)} + (\boldsymbol{u}_1^{(\lambda_2)}\boldsymbol{v}_1^{(\lambda_2)} + \cdots)e^{-\lambda_2 t} + \cdots, \quad (5)$$

where $\boldsymbol{v}_{m'}^{(0)}$ and $\boldsymbol{u}_{m'}^{(0)}$ $(1 \le m' \le m)$ are the zero left and right eigenvectors of L, respectively. Without loss of generality, we assume that $\boldsymbol{v}_{m'}^{(0)}$ and $\boldsymbol{u}_{m'}^{(0)}$ are normalized and orthogonalized such that

$$\sum_{j=1}^{N} v_{m',j}^{(0)} = 1, \quad (1 \le m' \le m), \quad (6)$$

and

$$\boldsymbol{v}_{m_1}^{(0)}\boldsymbol{u}_{m_2}^{(0)} = \delta_{m_1 m_2}. \quad (7)$$

λ_2 is the spectral gap (i.e., smallest positive eigenvalue) of L. $\boldsymbol{u}_1^{(\lambda_2)}$ and $\boldsymbol{v}_1^{(\lambda_2)}$ are a pair of left and right eigenvalues of L corresponding to λ_2, where $\boldsymbol{v}_1^{(\lambda_2)}\boldsymbol{u}_1^{(\lambda_2)} \neq 0$. Other modes that decay at least as fast as $e^{-\lambda_2 t}$ with t are omitted in Eq. (5). Note that Eq. (5) is also valid when L is not diagonalizable and has a nondiagonal Jordan normal form. The combination of Eqs. (4) and (5) leads to

$$(v_1 \cdots v_N) = \frac{1}{N}(1 \cdots 1)\sum_{m'=1}^{m} \boldsymbol{u}_{m'}^{(0)}\boldsymbol{v}_{m'}^{(0)}. \quad (8)$$

When $m=1$, by substituting $\boldsymbol{u}_1^{(0)}=(1\cdots 1)^{\mathsf{T}}$ (T denotes the transpose) into Eq. (8), we obtain $(v_1 \cdots v_N) = \boldsymbol{v}_1^{(0)}$. Therefore, Eq. (8) extends Eq. (2). We note that Eq. (8) is also applicable to disconnected networks for which $m \ge 2$.

To gain insights into Eq. (8), we consider the decomposition of directed networks into strongly connected components (SCCs). We define the uppermost SCC as an SCC that is not downstream to any other SCC along directed links. The number of uppermost SCCs in a given network is equal to m [10]. The choice of G_{r}, the set of root nodes, is unique up to the arbitrariness of the choice of a node in each uppermost SCC. This is consistent with the fact that the set of any single node is qualified as G_{r} in a strongly connected network.

In general, we can permute the indices of the nodes such that L is the irreducible normal form [8,17] given by

$$L = \begin{pmatrix} L_{11} & 0 & 0 & \cdots & 0 \\ L_{21} & L_{22} & 0 & \cdots & 0 \\ L_{31} & L_{32} & L_{33} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ L_{b1} & L_{b2} & L_{b3} & \cdots & L_{bb} \end{pmatrix}. \quad (9)$$

The diagonal block $L_{b'b'}$ $(1 \le b' \le b)$ corresponds to the b'th SCC. We denote the number of nodes in the b'th SCC by $N_{b'}$. Then, $L_{b'b''}$ is an $N_{b'} \times N_{b''}$ matrix and $\sum_{b'=1}^{b} N_{b'} = N$. The lower triangular nature of Eq. (9) implies that the SCCs are ordered in Eq. (9) such that links may exist from a node in the b'th SCC to the b''th SCC only when $b' \le b''$.

Because m out of b SCCs do not receive links from other SCCs, the uppermost SCCs occupy the first m rows of blocks in Eq. (9), and we obtain

$$L_{m_1 m_2} = 0, \quad (m_1 > m_2, 1 \le m_1 \le m). \quad (10)$$

Equation (9) constrained by Eq. (10) is called the Frobenius normal form [21]. In addition, $L_{m'm'}$ $(1 \le m' \le m)$ is the Laplacian matrix of the m'th SCC, which has a single zero eigenvalue. The eigenequations for this submatrix are represented by

$$L_{m'm'}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 0, \quad (v_{m',1} \cdots v_{m',N_{m'}})L_{m'm'} = 0. \quad (11)$$

It is easy [10] to verify that the m left zero eigenvectors of L are given by

$$\mathbf{v}_{m'}^{(0)} = (\underbrace{0 \cdots 0}_{\substack{N_1 + \ldots + N_{m'-1} \\ \text{zeros}}} \; v_{m',1} v_{m',2} \cdots v_{m',N_{m'}} \; \underbrace{0 \cdots 0}_{\substack{N_{m'+1} + \ldots + N_b \\ \text{zeros}}}),$$

$$\sum_{j=1}^{N_{m'}} v_{m',j} = 1, \quad (1 \le m' \le m). \quad (12)$$

To satisfy the normalization condition $\boldsymbol{v}_{m_1}^{(0)}\boldsymbol{u}_{m_2}^{(0)} = \delta_{m_1,m_2}$ and the first $\sum_{m'=1}^{m} N_{m'}$ rows of $L\boldsymbol{u}_{m'}^{(0)}=0$, we should take

$$\boldsymbol{u}_{m'}^{(0)} = (\underbrace{0 \cdots 0}_{\substack{N_1 + \ldots + N_{m'-1} \\ \text{zeros}}} \; \underbrace{1 \cdots 1}_{N_{m'} \text{ ones}} \; \underbrace{0 \cdots 0}_{\substack{N_{m'+1} + \ldots + N_m \\ \text{zeros}}} \; \bar{\boldsymbol{u}}_{m'}^{m+1\mathsf{T}} \cdots \bar{\boldsymbol{u}}_{m'}^{b\mathsf{T}})^{\mathsf{T}}, \quad (13)$$

where $\bar{\boldsymbol{u}}_{m'}^{b'}$ $(m+1 \le b' \le b)$ is the $N_{b'}$-dimensional column vector determined by

$$\begin{pmatrix} L_{m+1,m'} \\ \vdots \\ L_{b,m'} \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \begin{pmatrix} L_{m+1,m+1} & 0 & \cdots & 0 \\ L_{m+2,m+1} & L_{m+2,m+2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ L_{b,m+1} & L_{b,m+2} & \cdots & L_{bb} \end{pmatrix}\begin{pmatrix} \bar{\boldsymbol{u}}_{m'}^{m+1} \\ \vdots \\ \bar{\boldsymbol{u}}_{m'}^{b} \end{pmatrix}$$
$$= 0. \quad (14)$$

To show that Eq. (14) has a unique nonnegative solution, we decompose the diagonal block $L_{m'm'}$ $(m+1 \le m' \le N)$ as

$$L_{m'm'} = \tilde{L}_{m'm'} + D_{m'}, \quad (15)$$

where $\tilde{L}_{m'm'}$ is the Laplacian of the m'th SCC and $D_{m'}$ is the diagonal matrix whose ith element is equal to the total number of incoming links from SCCs $1, \ldots, m'-1$ to the ith node in the m'th SCC. Equation (15) implies that $L_{m'm'}$ is diagonally dominant. Therefore, by applying the Jacobi or Gauss-Seidel iteration to the first N_{m+1} rows of Eq. (14), we can uniquely calculate $\bar{\boldsymbol{u}}_{m'}^{m+1}$ [16]. Furthermore, $L_{m'm'}$ is an M-matrix [22]. Because all the elements of $L_{m+1,m'}(1\cdots 1)^{\mathsf{T}}$ that appear in the first N_{m+1} rows of Eq. (14) are not positive, all the elements of $\bar{\boldsymbol{u}}_{m'}^{m+1}$ are guaranteed to be nonnegative [22]. By substituting the obtained $\bar{\boldsymbol{u}}_{m'}^{m+1}$ in Eq. (14) and applying the Jacobi or Gauss-Seidel iteration to the next N_{m+2} rows, we can uniquely determine $\bar{\boldsymbol{u}}_{m'}^{m+2}$. By repeating the same procedure, we can successively determine $\bar{\boldsymbol{u}}_{m'}^{(0)}$, whose elements are unique and nonnegative.

The projection of L onto the eigenspaces yields

$$I = \left(\sum_{m'=1}^{m} \boldsymbol{u}_{m'}^{(0)} \boldsymbol{v}_{m'}^{(0)} \right) + \boldsymbol{u}_1^{(\lambda_2)} \boldsymbol{v}_1^{(\lambda_2)} + \cdots , \tag{16}$$

where I is the $N \times N$ unit matrix. Note that Eq. (16) is valid even if L is not diagonalizable. By multiplying the N-dimensional column vector $(1 \cdots 1)^{\top}$, a zero right eigenvector of L, from the right to both sides of Eq. (16) and using Eqs. (6) and (7), we obtain

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{m'=1}^{m} \boldsymbol{u}_{m'}^{(0)}. \tag{17}$$

Equation (17) implies $\sum_{m'=1}^{m} u_{m',j}^{(0)} = 1$ $(1 \le j \le N)$, where $u_{m',j}^{(0)}$ is the jth element of $\boldsymbol{u}_{m'}^{(0)}$ and represents the probability that the random walk starting from node j is trapped by the m'th uppermost SCC. $u_{m',j}^{(0)}$ can be interpreted as the magnitude of the influence that the m'th SCC exerts on node j. Note that $u_{m',j}^{(0)} = 0$ if node j cannot be reached from the m'th uppermost SCC along directed links in the original network.

By substituting Eq. (12) in Eq. (8), we obtain

$$v_i = v_{m',i} \frac{\sum_{j=1}^{N} u_{m',j}^{(0)}}{N} \tag{18}$$

for node i that belongs to the m'th uppermost SCC. For these nodes, $v_i > 0$ is satisfied. For nodes that do not belong to an uppermost SCC, we obtain $v_i = 0$. Equation (17) guarantees that $\sum_{i=1}^{N} v_i = 1$. Equation (18) generalizes the definition for strongly connected networks given by Eq. (1). We interpret the right-hand side of Eq. (18) to be the multiplication of the influence of node i within the m'th SCC (i.e., $v_{m',i}$) and the relative influence of the m'th SCC in the entire network (i.e., $\sum_{j=1}^{N} u_{m',j}^{(0)} / N$).

For pedagogical purposes, the calculations of v_i for two toy networks with $N=4$ and $m=2$ are presented in the Appendix.

IV. INTERPRETATION OF INFLUENCE FOR NETWORKS WITH MULTIPLE ROOT NODES

Borm and colleagues defined the Laplacian centrality measure on the basis of the continuous-time simple random walk on networks. In this section, we further motivate this definition by showing that v_i given by Eq. (18) have other interpretations, as is the case for v_i formulated for strongly connected networks [13].

A. Collective responses in the DeGroot model of consensus formation

The DeGroot model represents dynamical opinion formation in a population of interacting individuals [23]. The dynamics of the continuous-time version of the DeGroot model [24], also known as Abelson's model [18], are defined by

$$\dot{\boldsymbol{x}}(t) = - L \boldsymbol{x}(t), \tag{19}$$

where $\boldsymbol{x}(t) \equiv [x_1(t) \cdots x_N(t)]^{\top} \in \boldsymbol{R}^N$ represents the time-dependent opinion vector. For networks with $m=1$, including strongly connected networks, the consensus, i.e., synchrony, is asymptotically reached. In this case, the final synchronized opinion is given by $\boldsymbol{v}\boldsymbol{x}(0)$. Therefore, v_i is equal to the fraction of the initial opinion at node i reflected in the final opinion of the entire network [13,23–25].

When $m \ge 2$, synchrony is neutrally but not asymptotically stable. Therefore, the consensus of the entire network is not generally reached from a general initial condition. The final opinion vector is given by

$$\lim_{t\to\infty} \boldsymbol{x}(t) = \lim_{t\to\infty} e^{-Lt} \boldsymbol{x}(0) = \left(\sum_{m'=1}^{m} \boldsymbol{u}_{m'}^{(0)} \boldsymbol{v}_{m'}^{(0)} \right) \boldsymbol{x}(0). \tag{20}$$

If we set $x_j(0) = \delta_{ij}$ $(1 \le j \le N)$ to introduce a different opinion of unit strength at node i to the initial all-0 consensus state, the average response of the nodes induced by a different opinion at node i is equal to

$$\lim_{t\to\infty} \frac{1}{N} \sum_{j=1}^{N} x_j(t) = \frac{1}{N} (1 \cdots 1) \sum_{m'=1}^{m} \boldsymbol{u}_{m'}^{(0)} \boldsymbol{v}_{m'}^{(0)} \boldsymbol{e}_i, \tag{21}$$

where $\boldsymbol{e}_i$ is the N-dimensional unit column vector such that the ith element is equal to 1 and the other elements are equal to 0. Because Eq. (21) coincides with Eq. (18), the amount of the initial opinion of node i reflected in the final opinion of the entire network is given by Eq. (18).

B. Stationary density of voter model

The so-called link dynamics is a stochastic interacting particle system on networks in which each node takes one of the two opinions A and B [26]. In each time step, one link is randomly selected from the network with a probability proportional to the weight of the link. Then, the state of the source node of the link replaces that of the target node of the link if their states are different. Note that opinions A and B are equally strong in the dynamics. The dynamics halt when A or B takes over the entire network. The fixation probability of node i is defined as the probability that B takes over the network when the initial configuration is such that node i takes B and the other $N-1$ nodes take A. When $m=1$, v_i is equal to the fixation probability of node i [13,20].

When $m \ge 2$, the fixation of B introduced at node i never occurs. If node i is located in a downstream SCC, B eventually vanishes because A in the uppermost SCCs is permanent and replaces B in the downstream SCCs. If node i is located in an uppermost SCC, this SCC ends up with being entirely occupied by B with a positive probability. However, other uppermost SCCs are permanently occupied by A, such that the consensus is never reached.

In this situation, consider the expected fraction of B in the network in the stationary state when we start from the initial configuration with a single B at node i. The probability that B takes over the m'th uppermost SCC to which node i belongs is equal to $v_{m',i}$. Under the condition that the m'th uppermost SCC is entirely occupied by B and the other $m-1$ uppermost

SCCs are entirely occupied by A, the master equation for the probability p_j^{LD} that node j in a downstream SCC is occupied by B is given by

$$\frac{dp_j^{\mathrm{LD}}}{dt} = (1 - p_j^{\mathrm{LD}}) \sum_{j'=1}^{N} w_{j'j} p_{j'}^{\mathrm{LD}} - p_j^{\mathrm{LD}} \sum_{j'=1}^{N} w_{j'j}(1 - p_{j'}^{\mathrm{LD}}),$$

$$= \sum_{j'=1}^{N} w_{j'j} p_{j'}^{\mathrm{LD}} - p_j^{\mathrm{LD}} \sum_{j'=1}^{N} w_{j'j}, \tag{22}$$

where we set $p_{j'}^{\mathrm{LD}}=1$ when node j' belongs to the m'th uppermost SCC and $p_{j'}^{\mathrm{LD}}=0$ when node j' belongs to one of the other uppermost SCCs.

Equation (22) implies that in the equilibrium, $(p_1^{\mathrm{LD}} \cdots p_N^{\mathrm{LD}})^{\top}$ is a right zero eigenvector of L and identical to $\boldsymbol{u}_{m'}^{(0)}$ given by Eq. (13). Therefore, the stationary fraction of opinion B in the network is given by Eq. (18).

C. Enumeration of spanning trees

When $m=1$, the matrix-tree theorem implies that v_i is proportional to the sum of the weights of all the possible directed spanning trees rooted at node i [9,10,13]. The weight of a spanning tree is defined as the multiplication of all the weights of the $N-1$ links included in the spanning tree.

The Markov chain tree theorem extends this result to the case $m \geq 2$ [27]. According to this theorem, $v_{m',i} u_{m',j}^{(0)}$ for general directed networks is proportional to the sum of the weights of all the arborescences such that node i is a root node of the arborescence and the arborescence passes node j. An arborescence is a subgraph of the original networks with N nodes such that the indegree of each node restricted to the arborescence is at most one, it has no cycles, and it contains the maximal number of links. The nodes whose indegrees are zero within the arborescence are called the root nodes of the arborescence. They form G_{r} such that the concept of the root node for the arborescence and that for the network Laplacian [10] are identical. Therefore, the number of links in an arborescence is equal to $N-m$, and the arborescence is composed of m disconnected directed trees each of which emanates from a root node. Intuitively, $v_{m',i} u_{m',j}^{(0)}$ represents the number of different ways in which node i influences node j.

The influence of node i defined by Eq. (18) is proportional to the summation of all the arborescences with the modified weight. The modified weight of an arborescence is defined by the multiplication of all the weights of the $N-m$ links included in the arborescence and the number of nodes included in the directed tree rooted at node i in the arborescence. If node i is not the root of the arborescence, we set the weight of this arborescence to zero.

V. INFLUENCE OF NODES IN DOWNSTREAM COMPONENTS

A. Definition of the extended influence

With the definition given by Eq. (18), nodes that do not belong to any uppermost SCC have $v_i=0$. In practice, however, we often need to assess the relative importance of different nodes in downstream SCCs and that of nodes in different downstream SCCs. There also arise occasions when we want to compare uninfluential nodes in an uppermost SCC and influential nodes in a downstream SCC.

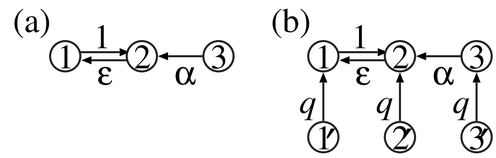

FIG. 3. (a) Example of directed network with $N=3$. (b) Network with $2N=6$ nodes obtained by adding source nodes to the network shown in (a).

An extreme situation in which this is the case is realized by the network shown in Fig. 3(a). Whenever $\epsilon > 0$ and $\alpha > 0$, we obtain $v_1 = v_2 = 0$ and $v_3 = 1$. However, when ϵ and α are small, node 1 may be regarded to be more central than node 3 because node 1 is much more central than node 2 and node 3 only weakly influences noncentral node 2. To cope with such a situation, we extend the influence to a one-parameter family of centrality measure by adopting the concept behind the definition of the PageRank.

The PageRank of node i, denoted by R_i, is defined as the stationary density of the discrete-time simple random walk as follows [6]:

$$R_i = (1-q) \sum_{j=1}^{N} \frac{w_{ji}}{\sum_{\ell=1}^{N} w_{j\ell}} R_j + \delta_{\sum_{\ell=1}^{N} w_{i\ell},0}(1-q) R_i + \frac{q}{N},$$

$$(1 \leq i \leq N), \tag{23}$$

where $\delta_{i,j}=1$ if $i=j$ and $\delta_{i,j}=0$ if $i \neq j$. The so-called teleportation probability q represents the probability that the random walker jumps from any node to an arbitrary node in one step. The same concept underlies the definition of the centrality based on the adjacency matrix [3,4]. According to the second term on the right-hand side in Eq. (23), the random walker stays at the node without outgoing links with probability $1-q$. The introduction of q is necessary for treating networks that are not strongly connected.

The PageRank is originally designed for web graphs. Therefore, receiving links increases R_i, which is opposite to the contention of the influence. To relate the PageRank to the influence, we consider the PageRank in the network generated by reversing all the links of the original network [13]. We denote this quantity for node i by R_i^{rev}, which is determined by

$$R_i^{\mathrm{rev}} = (1-q) \sum_{j=1}^{N} \frac{w_{ij}}{\sum_{\ell=1}^{N} w_{\ell j}} R_j^{\mathrm{rev}} + \delta_{\sum_{\ell=1}^{N} w_{\ell i},0}(1-q) R_i^{\mathrm{rev}} + \frac{q}{N},$$

$$(1 \leq i \leq N). \tag{24}$$

As explained in Sec. III, the influence corresponds to the continuous-time random walk on the link-reversed network. In a strongly connected network, the influence of each node is equal to the stationary density of the continuous-time random walk on the link-reversed network [13,20]. As in the definition of the PageRank, let us introduce random global jumps to the continuous-time random walk on the link-reversed network. We do so by assuming that the walker jumps from any node to an arbitrary node with rate q. Note that q represents a probability in the PageRank, whereas it is a rate in the influence. In the following, we allow q to exceed unity unless otherwise stated. The destination of the random jump is chosen from all the nodes with equal probability $1/N$. We denote by $\hat{v}_i$ the stationary density of the modified random walk at node i. The normalization is given by $\sum_{i=1}^{N}\hat{v}_i=1$. The stationary density is obtained from

$$\frac{d\hat{v}_i}{dt}=\sum_{j=1;j\neq i}^{N}\hat{v}_j w_{ij}-\hat{v}_i\sum_{j=1;j\neq i}^{N}w_{ji}+\frac{q}{N}-q\hat{v}_i=0. \tag{25}$$

We define the extended influence by the solution of Eq. (25). We note that the link-reversed version of Eq. (25), with a different structure of the global jump, was proposed as an alternative of the PageRank to be applied to web graphs [28].

In the vector notation, Eq. (25) is represented by

$$(\hat{v}_1\cdots\hat{v}_N)(L+qI)=\frac{q}{N}(1\cdots 1), \tag{26}$$

or equivalently,

$$(\hat{v}_1\cdots\hat{v}_N)\left(L+qI-\frac{q}{N}J\right)=0, \tag{27}$$

where J is the N by N matrix whose all the elements are equal to unity. If $q>0$, $L+qI$ is strictly diagonally dominant, and Eq. (26) can be solved by the Jacobi or Gauss-Seidel iteration. A large q guarantees exponentially fast convergence of the iteration [16].

We note that

$$L+qI=\sum_{m'=1}^{m}q\boldsymbol{u}_{m'}^{(0)}\boldsymbol{v}_{m'}^{(0)}+(\lambda_2+q)\boldsymbol{u}_1^{(\lambda_2)}\boldsymbol{v}_1^{(\lambda_2)}+\cdots, \tag{28}$$

which leads to

$$(L+qI)^{-1}=\sum_{m'=1}^{m}\frac{1}{q}\boldsymbol{u}_{m'}^{(0)}\boldsymbol{v}_{m'}^{(0)}+\frac{1}{\lambda_2+q}\boldsymbol{u}_1^{(\lambda_2)}\boldsymbol{v}_1^{(\lambda_2)}+\cdots. \tag{29}$$

In the limit $q\to 0$, Eq. (29) implies that $\hat{v}_i\approx v_i$, where v_i is defined by Eq. (8). For the first term on the right-hand side of Eq. (8) to be comparable with the remaining terms, q must be at least approximately Re λ_2. If this is the case, $\hat{v}_i$ can quantitatively differentiate various nodes including those in downstream components.

In the limit $q\to\infty$, Eq. (26) gives $\hat{v}_i=1/N$ $(1\le i\le N)$. When q is a large finite value, Eq. (26) is expanded as

$$\begin{aligned}(\hat{v}_1\cdots\hat{v}_N)&=\frac{1}{N}(1\cdots 1)\sum_{\ell=0}^{\infty}(-1)^{\ell}\left(\frac{L}{q}\right)^{\ell}\\&=\frac{1}{N}(1\cdots 1)+\frac{1}{Nq}(k_1^{\text{out}}-k_1^{\text{in}}\cdots k_N^{\text{out}}-k_N^{\text{in}})+O\left(\frac{1}{q^2}\right),\end{aligned} \tag{30}$$

where k_i^{out} and k_i^{in} are the outdegree and the indegree of node i, respectively. The Taylor expansion is justified when $q>\text{Re}\,\lambda_N$, where λ_N is the eigenvalue of L with the largest modulus. If q is large relative to Re λ_N, the influence is determined by the outdegree and the indegree and is independent of the global structure of networks. Therefore, in practice, q should not be too large as compared to Re λ_N. This is surprising because a large q implies a strong global connectivity. As a rule of thumb, we recommend setting $\text{Re}\,\lambda_2<q<\text{Re}\,\lambda_N$. A suitable range of the teleportation probability q for the PageRank can be also obtained by applying the criterion $\text{Re}\,\lambda_2<q<\text{Re}\,\lambda_N$ to the PageRank matrix implied in Eq. (23).

B. Interpreting the extended influence without regard to global jumps

We have extended the influence by introducing global jumps to the continuous-time random walk on the link-reversed network. However, the meaning of the teleportation term in terms of the dynamical and structural properties of the nodes in the network and its rationale are somewhat vague. We show that the extended influence defined by Eq. (26) allows another interpretation: absorption probability of the random walk on the link-reversed network with a sink attached to each node but without global jumps. A similar interpretation was made for the PageRank in Ref. [29].

We assume N additional source nodes indexed by $1',\ldots,N'$ and directed links with weight $q>0$ from node i' to node i $(1\le i\le N)$ in the original network. The extension of the network shown in Fig. 3(a) is depicted in Fig. 3(b). The extended network has $2N$ nodes. Nodes $1',\ldots,N'$ are the unique root nodes of the extended network. Node i' forms the ith uppermost SCC in the extended network. The multiplicity of the zero Laplacian eigenvalue of the extended network is equal to $m=N$.

We then reverse all the links and consider the probability that the random walker starting from an arbitrary node with equal probability $1/2N$ is absorbed at node i'. This probability is given by $v_{i'}$. Because it is obvious and uninformative that the random walker starting from the auxiliary node i' is necessarily absorbed to node i', we would like to exclude this factor. Therefore, we examine the quantity given by

$$2\left(v_{i'}-\frac{1}{2N}\right)=2v_{i'}-\frac{1}{N}. \tag{31}$$

The subtraction of $1/(2N)$ in Eq. (31) accounts for the exclusion of the random walker starting from and absorbed to node i'. The multiplicative factor 2 accounts for the fact that we effectively start the random walk from nodes $1,\ldots,N$ with equal probability $1/N$.

Equation (18) implies that the calculation of $v_{i'}$ involves $v_{i,1}$, that is, the first element of the left zero eigenvector of L corresponding to the ith uppermost SCC. Because the uppermost SCC consists of single node i', $v_{i,1}$ is equal to unity. The calculation of $v_{i'}$ also involves $\boldsymbol{u}_i^{(0)} \equiv (\boldsymbol{e}_i\, \bar{\boldsymbol{u}}_i)^{\mathsf{T}}$, that is, the zero eigenvector of the $2N$-dimensional Laplacian. $\bar{\boldsymbol{u}}_i^{\mathsf{T}}$ is an N-dimensional column vector. By substituting these expressions and Eq. (18) in Eq. (31), the quantity given by Eq. (31) is equal to

$$2\left(v_{i,1}\frac{\sum_{j=1}^{2N} u_{i,j}^{(0)}}{2N} - \frac{1}{2N}\right) = \frac{\sum_{j=1}^{N} \bar{u}_{i,j}}{N}. \tag{32}$$

We calculate $\bar{\boldsymbol{u}}_i^{\mathsf{T}}$ from

$$\begin{pmatrix} O & O \\ -qI & L+qI \end{pmatrix}\begin{pmatrix} \boldsymbol{e}_i \\ \bar{\boldsymbol{u}}_i \end{pmatrix} = 0, \tag{33}$$

where O is the $N \times N$ zero matrix and L is the Laplacian of the original network. Equation (33) is equivalent to

$$(L+qI)\bar{\boldsymbol{u}}_i = q\boldsymbol{e}_i. \tag{34}$$

By combining Eqs. (32) and (34), we obtain Eq. (26).

With this interpretation, we gain an intuitive understanding of the fact that the extended influence $\hat{v}_i$ is a local quantity when q is large. In this situation, the tendency that a random walk exits from each node is strong, and a random walk would not travel a long distance before being absorbed. Therefore, it is natural that $\hat{v}_i$ at large q is efficiently approximated by local quantities of nodes such as the outdegree and the indegree, as discussed using Eq. (30).

VI. TOY EXAMPLES

In this and the next section, we apply the extended influence to various networks.

A. Network with $N=3$

Consider the network shown in Fig. 3(a). We are concerned with the situation in which $0 < \epsilon \ll 1$ such that node 1 is apparently much more central than node 2. If node 3 is absent, $v_1/v_2 = 1/\epsilon$; node 1 is actually much more influential than node 2 [14,25]. However, regardless of the value of $\alpha > 0$, node 3 takes all the share of the influence if we use v_i.

The extended influence $\hat{v}_i$ $(1 \le i \le 3)$ is equal to v_i for the network shown in Fig. 3(b). We obtain

$$\hat{v}_1 = \frac{1}{3\Delta}[q^2 + (2+\alpha)q], \tag{35}$$

$$\hat{v}_2 = \frac{1}{3\Delta}(q^2 + 2\epsilon q), \tag{36}$$

$$\hat{v}_3 = \frac{1}{3\Delta}[q^2 + (1+\epsilon+2\alpha)q + 3\epsilon\alpha], \tag{37}$$

where

$$\Delta = q^2 + (1+\epsilon+\alpha)q + \epsilon\alpha. \tag{38}$$

Therefore, $\hat{v}_1 > \hat{v}_3$ when $(1-\epsilon-\alpha)q > 3\epsilon\alpha$. When ϵ or α is small and $\epsilon+\alpha < 1$, we have an intuitive result that node 1 is more influential than node 3.

B. Directed chain

Consider a directed chain having N nodes defined by $w_{i,i+1}=1$ $(1 \le i \le N-1)$ and $w_{i,j}=0$ $(j \ne i+1)$. The network is schematically shown in Fig. 2. We obtain $v_1=1$ and $v_i=0$ $(2 \le i \le N)$ [14]. However, nodes with small i are located relatively upstream in the chain and intuitively appear influential as compared to nodes with large i. We can calculate the influence either by solving Eq. (34) or by analyzing random walks with N traps on the network obtained by reversing all the links shown in Fig. 2. When the random walker on the link-reversed network starts from node j $(2 \le j \le N)$, the probability that the walker exits from node i to the absorbing node i' is equal to

$$u_{ij}^{(0)} = \begin{cases} \dfrac{1}{(1+q)^{j-1}}, & (i=1), \\ \dfrac{q}{(1+q)^{j-i+1}}, & (2 \le i \le j), \\ 0, & (j+1 \le i \le N). \end{cases} \tag{39}$$

Therefore, we obtain

$$\hat{v}_i = \frac{1}{N}\sum_{j=1}^{N} u_{ij}^{(0)} = \begin{cases} \dfrac{1}{N}\left(\dfrac{1+q}{q} - \dfrac{1}{q(1+q)^{N-1}}\right), & (i=1), \\ \dfrac{1}{N}\left(1 - \dfrac{1}{(1+q)^{N-i+1}}\right), & (2 \le i \le N). \end{cases} \tag{40}$$

We note that $\lim_{q\to 0}\hat{v}_1 = 1$, $\lim_{q\to 0}\hat{v}_i = 0$ $(2 \le i \le N)$, and $\hat{v}_i$ monotonically decreases with i for any $q>0$.

VII. NUMERICAL RESULTS

In this section, we examine the influence in three directed networks: a random graph, a neural network, and an online social network.

A. Descriptions of networks

We generate a directed random network with $N=100$ and expected degree $\langle k \rangle = 3.5$ by connecting each ordered pair of nodes independently with probability $\langle k \rangle/(N-1)$. Because $\langle k \rangle$ is relatively small, the generated network is not strongly connected, whereas it is weakly connected, i.e., not divided into disconnected components. The generated network has three root nodes, each of which forms an SCC. The largest SCC contains 94 nodes and is downstream to the three root nodes. The extremal Laplacian eigenvalues are $\lambda_2=0.046$ and $\lambda_N=8.255$.

We generate a *C. elegans* neural network with $N=279$ on the basis of published data [30]. In this network, there exist two types of links: undirected gap junctions and directed

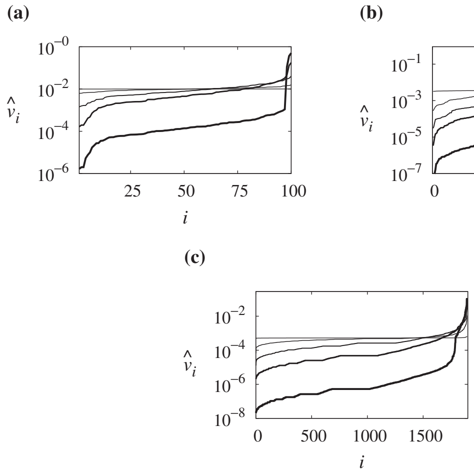

FIG. 4. Values of $\hat{v}_i$ for (a) random graph, (b) neural network, and (c) online social network. We set q=0.001, 0.1, 1, 10, 1000 (from steep thick lines to flat thin lines). For each q, we have sorted $\hat{v}_i$ in the ascending order for demonstration.

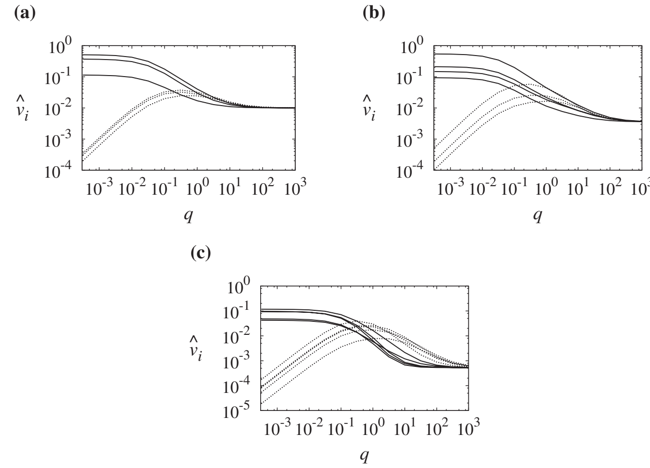

FIG. 5. Dependence of $\hat{v}_i$ of some nodes on q for (a) random graph, (b) neural network, and (c) online social network. The nodes with the largest influence values for q=0.001 and for q=10 correspond to the solid and dashed lines, respectively.

chemical synapses. A pair of neurons can be connected by multiple synapses. We regard this network as a weighted directed network, where the weight of the link from neuron i to neuron j is defined as the summation of the number of gap junctions between i and j and the number of chemical synapses from i to j. The network has 2993 links. The largest SCC has 274 nodes [13]. Four of the five remaining nodes are located upstream to the largest SCC and form individual SCCs. The other node is located downstream to the largest SCC. The extremal Laplacian eigenvalues are $\lambda_2=0.050$ and $\lambda_N=354.105$.

The third network that we use is an online social network among students at University of California, Irvine [31]. This network has N=1899 nodes and 20 296 directed and weighted links. We focus on the largest weakly connected component of this network that contains 1893 nodes and 13 835 links. There exist 103 root nodes, each of which forms an SCC. The largest SCC has 1023 nodes and is downstream to these root nodes. The extremal Laplacian eigenvalues are $\lambda_2=0.146$ and $\lambda_N=92.996$.

B. Analysis of influence in the three networks

The rank plots of the influence for various values of q for the random graph, neural network, and online social network are shown in Figs. 4(a)–4(c), respectively. In the figure, the values of $\hat{v}_i$ are shown in the ascending order for each q for clarity.

When q=0.001 (thickest lines), $\hat{v}_i$ is similar to v_i for the three networks. Therefore, the root nodes have exclusively large $\hat{v}_i$, whereas the other nodes have $\hat{v}_i \approx 0$. Accordingly, we find a sudden jump in the rank plot for each network. Such a small value of q does not allow us to quantitatively compare the centrality of nodes in downstream components. This is also anticipated from the fact that the three networks yield $q=0.001<\mathrm{Re}\lambda_2$. In the other extreme, $\hat{v}_i \approx 1/N$ is roughly satisfied when q=1000 (thinnest lines). This is consistent with the fact that the three networks yield $q=1000>\mathrm{Re}\,\lambda_N$. In this range of q, the influence is not an adequate centrality measure. For intermediate values of q, $\hat{v}_i$ is reasonably dispersed, and nodes that are not the roots are also endowed with positive $\hat{v}_i$. We consider that the influence with intermediate values of q enables us to compare the importance of nodes that are in downstream SCCs and quantify the relative importance of nodes in uppermost SCCs and nodes in downstream SCCs.

The influence with intermediate values of q is distinct from the interpolation of the influence when $q\rightarrow 0$ (i.e., $\hat{v}_i \approx v_i$) and that when $q=\infty$ (i.e., $\hat{v}_i=1/N$). The order of the nodes in terms of the value of $\hat{v}_i$ drastically changes as q varies. To demonstrate this, we examine the dependence of $\hat{v}_i$ on q for some selected nodes.

For the random graph, we select the three root nodes, for which $\hat{v}_i$ is the largest at q=0.001 and the three nodes whose $\hat{v}_i$ is the largest at q=10. The dependence of $\hat{v}_i$ on q for the six nodes is shown in Fig. 5(a). The three root nodes [solid lines in Fig. 5(a)] and the three nodes with the largest $\hat{v}_i$ at q=10 (dashed lines) do not overlap each other. In particular, the root node with the third largest $\hat{v}_i$ for q=0.001 does not have large $\hat{v}_i$ when q is approximately larger than 1. Although the indegree of this root node is equal to zero, the destinations of the links from this root node are presumably nodes with small influence values in the largest SCC. This phenomenon is essentially the same as that shown in Fig. 3.

The neural network has four root nodes. The dependence of $\hat{v}_i$ on q for the root nodes and the three nodes whose $\hat{v}_i$ is among the four largest values at q=10 are shown in Fig. 5(b). In the neural network, one of the four roots is among the nodes with the four largest values of $\hat{v}_i$ at q=10. For the online social network, the relationships between $\hat{v}_i$ and q for the five root nodes with the largest $\hat{v}_i$ at q=0.001 and the five nodes with the largest $\hat{v}_i$ at q=1000 are shown in Fig. 5(c). The results for the neural network and the online social network are qualitatively the same as those for the random graph. In particular, some root nodes (solid lines) do not have particularly large $\hat{v}_i$ when q is approximately larger than unity.

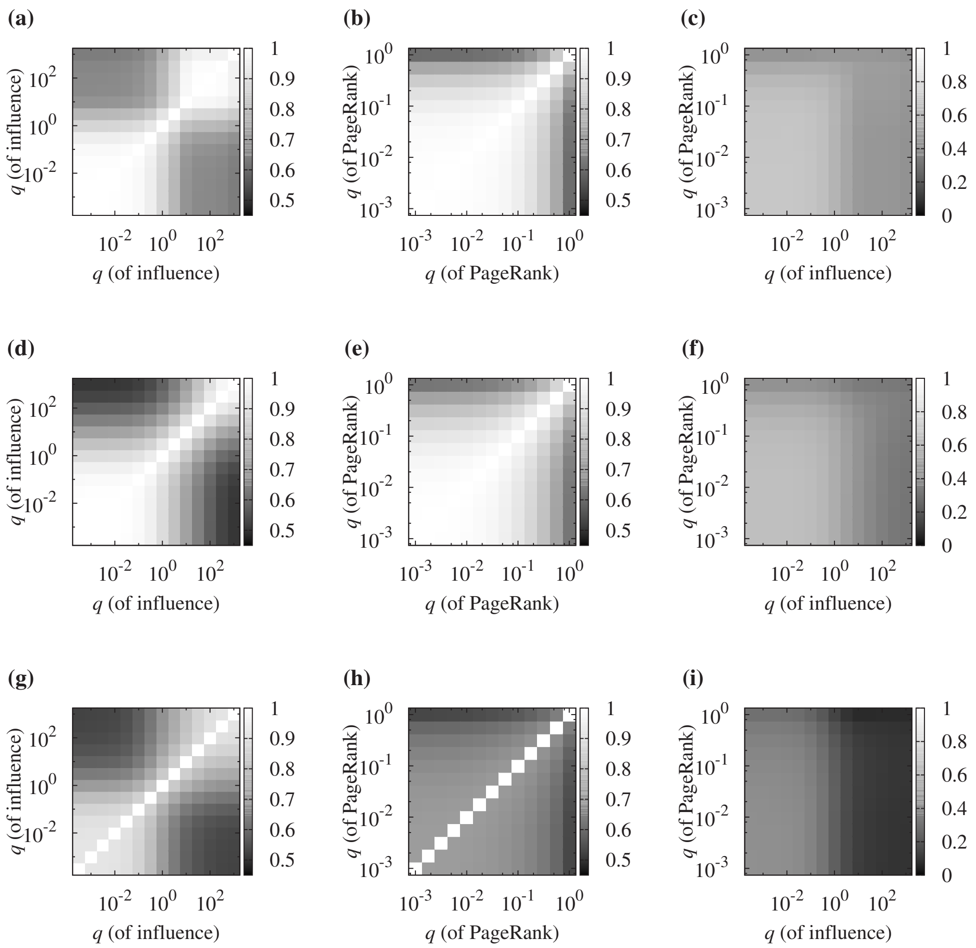

FIG. 6. Kendall rank correlation coefficient for the influence and the PageRank. (a)–(c) Random graph, (d)–(f) neural network, and (g)–(i) online social network. (a, d, g) Correlation between the influence for different values of q. (b, e, h) Correlation between the PageRank for different values of q. (c, f, i) Correlation between the influence and the PageRank for various values of q.

Finally, we quantify the dependence of the influence on q by calculating the Kendall rank correlation coefficient. It is defined as $2\{P/[N(N-1)/2]-1\}$, where P is the number of pairs i, j $(1\leq i<j\leq N)$ such that the sign of $\hat{v}_i-\hat{v}_j$ for $q=q_1$ is the same as that for $q=q_2$. The correlation coefficient falls between -1 and 1. The correlation coefficient for the random graph for various values of q is shown in Fig. 6(a). As anticipated, the correlation decreases with $|q_1-q_2|$. Figure 6(a) also indicates that the ranking on the basis of the influence is fairly insensitive to q in two ranges of q, i.e., for q smaller than ≈ 1 and for q larger than ≈ 10. The ranking is sensitive to q between these two ranges of q. For comparison, the correlation coefficient for the PageRank for various values of q is shown in Fig. 6(b). Similar to the case of the influence, the correlation decreases with $|q_1-q_2|$. The correlation between the influence and the PageRank [Fig. 6(c)] is generally small regardless of the two values of q. On this basis, we claim that the influence and the PageRank are distinct centrality measures. This result generalizes that when directed networks are strongly connected and $q=0$ [13].

The rank correlation coefficient for the neural network and the social network calculated in the same manner is shown in Figs. 6(d)–6(f) and Figs. 6(g)–6(i), respectively. The results are qualitatively the same as those for the random graph [Figs. 6(a)–6(c)].

VIII. CONCLUSIONS

We have proposed a centrality measure (influence) for general directed networks. It is a generalization of a Laplacian-based centrality measure that is often used for strongly connected networks [7–9,13–15]. It also generalizes the formulation of the same centrality measure developed for networks that are not necessarily strongly connected [10–12]. Unlike the previous measure [10–12], the proposed measure is suitable for comparing the importance of nodes that are in downstream SCCs and comparing nodes in different SCCs. It has a free parameter q. For networks that are not strongly connected, we suggest using $\mathrm{Re}\,\lambda_2<q<\mathrm{Re}\,\lambda_N$ (Sec. V A). A small value of q implies that the centrality values concentrate on nodes in uppermost components. A large value of q makes the influence close to a degree centrality, i.e., outdegree minus indegree. The choice of q is up to users' preferences. We acknowledge that various mathematical properties of the matrix associated with the influence (i.e., $L+qI$) have been analyzed in [10,11]. In [11], the use of this matrix for the centrality measure is briefly mentioned.

Arguably, the most frequently used centrality measure for directed networks appears to be the PageRank [6]. Beyond the World Wide Web, for which the PageRank was originally designed, the PageRank has been applied to rank, for example, academic papers and journals (e.g., [32]). The PageRank is interpreted as the stationary density of the discrete-time simple random walk with global jumps on the network.

We have defined the influence as a continuous-time counterpart of the PageRank. Furthermore, we have provided the interpretation of the influence as the absorption probability of the continuous-time random walk to the sink attached to each node but not with global random jumps. As a corollary, the PageRank can be interpreted as the absorption probability of the random walk without teleportation to a sink. In addition, a suitable range of the teleportation probability in the PageRank can be estimated by adapting the criterion $\mathrm{Re}\,\lambda_2 < q < \mathrm{Re}\,\lambda_N$ to the discrete-time random walk.

For the case of strongly connected networks, we refer to our previous work [13,20] for a discussion of continuous-time versus discrete-time random walk. We have shown that q controls the relative importance of nodes in upstream SCCs and nodes in downstream SCCs. The same role is shared by the teleportation probability in the PageRank. Then, why do we feel the need to introduce a new centrality?

First, the extended influence inherits the definition of the influence for strongly connected networks and one-root networks (i.e., influence when $q=0$), and therefore, it represents the importance of nodes in various dynamics and in the enumeration of spanning trees (Sec. IV). Actually, for each dynamics considered in Sec. IV, we can consider a discrete-time version and relate the importance of nodes in the dynamics to the PageRank. We have explained this correspondence for the random walk (Sec. V). In addition, the DeGroot model of opinion formation was originally proposed in discrete time [23]. We should choose one among the two centrality measures depending on whether the continuous-time or discrete-time dynamics are assumed to occur on the network in question.

In the discrete-time interpretation, the indegree is essentially normalized to be unity. Therefore, if the weight of the link represents a value that should not be normalized, such as the rate of interaction, nominal connection strength, amount of signal or monetary fluxes, and the number of wins and losses between a pair of sports teams, the continuous-time interpretation, that is, the influence, appears to be more appropriate. On the other hand, the PageRank is more appropriate in the case of scientometry; if a paper cites many papers, the value of each citation should be considered to be small, and being cited from this paper should not be of great importance. This distinction may underlie the current situation that the PageRank and the Laplacian-based centrality have been used in somewhat different research communities and for different types of data. In this light, we have extended the Laplacian-based centrality so that it is applicable to general directed networks, as is the PageRank.

Second, the PageRank has a subtle arbitrariness in determining the behavior of the random walk that has reached a dangling node. Depending on the implementation, the walker at a dangling node hops to a randomly chosen node even with probability $1-q$ $(0<q\le 1)$ [6] or stays at the same node with probability $1-q$ [33]. The theoretical justification for either assumption is not clear. In the influence, we have the sole control parameter q, and the influence unambiguously corresponds to the discrete-time case in which the walker stays at the dangling node with probability $1-q$.

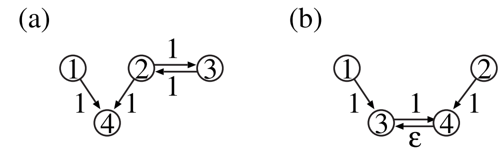

FIG. 7. Two directed networks with two root nodes.

ACKNOWLEDGMENTS

We thank Yoji Kawamura for the helpful discussions and for careful reading of the paper. N.M. acknowledges the support provided by the Grants-in-Aid for Scientific Research (Grant Nos. 20760258 and 20540382, and Innovative Areas "Systems Molecular Ethology") from MEXT, Japan.

APPENDIX: INFLUENCE FOR TWO TOY NETWORKS WITH MULTIPLE ROOT NODES

For the network with four nodes and two root nodes shown in Fig. 7(a), we obtain $m=2$, $b=3$, $N_1=1$, $N_2=2$, $N_3=1$,

$$L = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 2 \end{pmatrix}, \tag{A1}$$

$$\boldsymbol{v}_1^{(0)} = (1\ \ 0\ \ 0\ \ 0), \quad \boldsymbol{u}_1^{(0)} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \frac{1}{2} \end{pmatrix}, \tag{A2}$$

$$\boldsymbol{v}_2^{(0)} = \left(0\ \ \frac{1}{2}\ \ \frac{1}{2}\ \ 0\right), \quad \boldsymbol{u}_2^{(0)} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ \frac{1}{2} \end{pmatrix}. \tag{A3}$$

Therefore, the influence is given by

$$v_1 = \frac{3}{8}, \quad v_2 = v_3 = \frac{5}{16}, \quad v_4 = 0. \tag{A4}$$

Nodes 2 and 3 have the same influence because they are as strong as each other within their SCC. Although the two upstream SCCs are upstream to node 4 in the same manner, v_1 is smaller than v_2+v_3 because v_1 controls two nodes and the SCC of nodes 2 and 3 controls three nodes.

For the network with four nodes and two root nodes shown in Fig. 7(b), we obtain $m=2$, $b=3$, $N_1=1$, $N_2=1$, $N_3=2$,

$$L=\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1+\epsilon & -\epsilon \\ 0 & -1 & -1 & 2 \end{pmatrix}, \tag{A5}$$

$$\boldsymbol{v}_1^{(0)}=(1\ \ 0\ \ 0\ \ 0), \quad \boldsymbol{u}_1^{(0)}=\begin{pmatrix} 1 \\ 0 \\ \dfrac{2}{2+\epsilon} \\ \dfrac{1}{2+\epsilon} \end{pmatrix}, \tag{A6}$$

$$\boldsymbol{v}_2^{(0)}=(0\ \ 1\ \ 0\ \ 0), \quad \boldsymbol{u}_2^{(0)}=\begin{pmatrix} 0 \\ 1 \\ \dfrac{\epsilon}{2+\epsilon} \\ \dfrac{1+\epsilon}{2+\epsilon} \end{pmatrix}. \tag{A7}$$

The influence is given by

$$v_1=\frac{5+\epsilon}{4(2+\epsilon)}, \quad v_2=\frac{3(1+\epsilon)}{4(2+\epsilon)}, \quad v_3=v_4=0. \tag{A8}$$

$v_1 > v_2$ because node 1 is connected to the more influential node of the downstream SCC (i.e., v_3) unlike node 2, which links to the less influential node of the downstream SCC (i.e., v_4). Note that when $\epsilon \approx 0$, the effect of node 1 on node 4 is similar to that of node 2 on node 4, despite the fact that node 1 does not directly link to node 4, whereas node 2 does. The reverse is not the case; when $\epsilon \approx 0$, node 1 can affect node 3, but node 2 can hardly do so.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[3] M. E. J. Newman, *Networks—An introduction* (Oxford University Press, Oxford, 2010).
[4] L. Katz, Psychometrika **18**, 39 (1953).
[5] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, New York, 1994); *Network Analysis*, edited by U. Brandes and T. Erlebach (Springer, Berlin, 2005).
[6] S. Brin and L. Page, *Proceedings of the Seventh International World Wide Web Conference* (Elsevier, Amsterdam, 1998), pp. 107–117; A. N. Langville and C. D. Meyer, *Google's PageRank and Beyond* (Princeton University Press, Princeton, 2006).
[7] H. E. Daniels, Biometrika **56**, 295 (1969); K. A. Berman, SIAM J. Algebraic Discrete Methods **1**, 359 (1980).
[8] J. W. Moon and N. J. Pullman, SIAM Rev. **12**, 384 (1970).
[9] N. Biggs, Bull. London Math. Soc. **29**, 641 (1997).
[10] R. P. Agaev and P. Y. Chebotarev, Autom. Remote Control (Engl. Transl.) **61**, 1424 (2000).
[11] P. Chebotarev and R. Agaev, e-print arXiv:math/0508171.
[12] N. E. Borm, R. Van Den Brink, and M. Slikker, Ann. Operat. Res. **109**, 61 (2002).
[13] N. Masuda, Y. Kawamura, and H. Kori, New J. Phys. **11**, 113002 (2009).
[14] N. Masuda, Y. Kawamura, and H. Kori, Phys. Rev. E **80**, 046114 (2009).
[15] K. Klemm, M. Á. Serrano, V. M. Eguíluz, and M. San Miguel, e-print arXiv:1002.4042.
[16] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. (Johns Hopkins University Press, Baltimore, 1996).
[17] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, England, 1985).
[18] R. P. Abramson, in *Contributions to Mathematical Psychology*, edited by N. Frederiksen and H. Gulliksen (Holt, Rinehart, and Winston, New York, 1964), pp. 141–160.
[19] G. B. Ermentrout, SIAM J. Appl. Math. **52**, 1665 (1992).
[20] N. Masuda and H. Ohtsuki, New J. Phys. **11**, 033012 (2009).
[21] R. B. Bapat and T. E. S. Raghavan, *Nonnegative Matrices and Applications* (Cambridge University Press, Cambridge, England, 1997).
[22] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences* (Academic Press, San Diego, CA, 1979).
[23] M. H. DeGroot, J. Am. Stat. Assoc. **69**, 118 (1974); N. E. Friedkin, Am. J. Sociol. **96**, 1478 (1991).
[24] R. Olfati-Saber, J. Fax, and R. Murray, Proc. IEEE **95**, 215 (2007).
[25] H. Kori, Y. Kawamura, H. Nakao, K. Arai, and Y. Kuramoto, Phys. Rev. E **80**, 036207 (2009).
[26] T. Antal, S. Redner, and V. Sood, Phys. Rev. Lett. **96**, 188104 (2006); V. Sood, T. Antal, and S. Redner, Phys. Rev. E **77**, 041121 (2008).
[27] F. T. Leighton and R. L. Rivest, IEEE Trans. Inf. Theory **32**, 733 (1986).
[28] Y. Liu, B. Gao, T. Y. Liu, Y. Zhang, Z. Ma, S. He, and H. Li, *SIGIR08: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (ACM, New York, 2008), pp. 451–458; Y. Liu, T. Y. Liu, B. Gao, Z. Ma, and H. Li, Inf. Retrieval **13**, 22 (2010).
[29] K. Avrachenkov, N. Litvak, D. Nemirovsky, and N. Osipova, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **45**, 890 (2007).
[30] B. L. Chen, D. H. Hall, and D. B. Chklovskii, Proc. Natl. Acad. Sci. U.S.A. **103**, 4723 (2006).
[31] P. Panzarasa, T. Opsahl, and K. M. Carley, J. Am. Soc. Inf. Sci. Technol. **60**, 911 (2009).
[32] I. Palacios-Huerta and O. Volij, Econometrica **72**, 963 (2004); A. Fersht, Proc. Natl. Acad. Sci. U.S.A. **106**, 6883 (2009).
[33] S. Fortunato, A. Flammini, F. Menczer, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **103**, 12684 (2006).