# Impact of hierarchical modular structure on ranking of individual nodes in directed networks

To cite this article: Naoki Masuda *et al* 2009 *New J. Phys.* **11** 113002

View the article online for updates and enhancements.

## You may also like

# Impact of hierarchical modular structure on ranking of individual nodes in directed networks

**Naoki Masuda**[1,2,5]**, Yoji Kawamura**[3] **and Hiroshi Kori**[2,4]

[1] Graduate School of Information Science and Technology, 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] Institute for Research on Earth Evolution, Japan Agency for Marine-Earth Science and Technology, 3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan
[4] Division of Advanced Sciences, Ochadai Academic Production, Ochanomizu University, 2-1-1, Ohtsuka, Bunkyo-ku, Tokyo 112-8610, Japan
E-mail: masuda@mist.i.u-tokyo.ac.jp

**Abstract.** Many systems, ranging from biological and engineering systems to social systems, can be modeled as directed networks, with links representing directed interaction between two nodes. To assess the importance of a node in a directed network, various centrality measures based on different criteria have been proposed. However, calculating the centrality of a node is often difficult because of the overwhelming size of the network or because the information held about the network is incomplete. Thus, developing an approximation method for estimating centrality measures is needed. In this study, we focus on modular networks; many real-world networks are composed of modules, where connection is dense within a module and sparse across different modules. We show that ranking-type centrality measures, including the PageRank, can be efficiently estimated once the modular structure of a network is extracted. We develop an analytical method to evaluate the centrality of nodes by combining the local property (i.e. indegree and outdegree of nodes) and the global property (i.e. centrality of modules). The proposed method is corroborated by real data. Our results provide a linkage between the ranking-type centrality values of modules and those of individual nodes. They also reveal the hierarchical structure of

[5] Author to whom any correspondence should be addressed.

networks in the sense of subordination (not nestedness) laid out by connectivity among modules of different relative importance. The present study raises a novel motive for identifying modules in networks.

**Contents**

## 1. Introduction

A variety of systems of interacting elements can be represented as networks. A network is a collection of nodes and links; a link connects a pair of nodes. Generally speaking, some nodes play central functions, such as binding different parts of the network together and controlling dynamics in the network. To identify important nodes in a network, various centrality measures based on different criteria have been proposed [1]–[3].

Links of many real networks such as the World Wide Web (WWW), food webs, neural networks, protein interaction networks and many social networks are directed or asymmetrically weighted. In contrast to the case of undirected networks, a link in directed networks indicates an asymmetrical relationship between two nodes, for example the control of the source node of a link over the target node. The direction of a link indicates the relative importance of the two nodes. Central nodes in a network in this sense would be, for example, executive personnel in an organizational network or top predators in a food web. Generally, more (less) central nodes are located at an upper level (a lower level) in the hierarchy of the network, where hierarchy refers to the distinction between the upper and lower levels in terms of the centrality value as relevant in, for example, biological [4, 5] and social [6] systems. This type of centrality measure is necessarily specialized for directed networks and includes the popularity or prestige measures for social networks [1], ranking systems for webpages such as the PageRank [7, 8] and HITS [9, 10], adaptations of the PageRank to citation networks of academic papers [11, 12] and

journals [11], [13]–[15], and ranking systems of sports teams [16]. We call them ranking-type centrality measures.

Under practical restrictions such as overwhelming network size or incomplete information about the network, it is often difficult to obtain exact ranking-type centrality values of nodes. In such situations, the simplest approximators are perhaps those based on the degree of nodes (i.e. the number of links owned by a node). For example, the indegree of a node can be an accurate approximator of the PageRank of websites [17] and ranks of academic journals [14, 15]. However, such local approximations often fail [18]–[20], implying a significant effect of the global structure of networks.

A ubiquitous global structure of networks that adversely affects local approximations is the modular structure. Both in undirected [21]–[24] and directed [1], [24]–[27] networks, nodes are often classified into modules (also called communities), such that the nodes are densely connected within a module and sparsely connected across different modules. In modular networks, some modules may be central in a coarse-grained network, where each module is regarded as a supernode [28]. However, relationships between the centrality of individual nodes and that of modules are not well understood. Using these relationships, we will be able to assess the centrality of individual nodes only on the basis of coarse-grained information about the organization of modules or under limited computational resources.

In this study, we analyze the ranking-type centrality measures for directed modular networks. We are concerned with the modular structure of the network in the sense of the partitioning of it into parts, and not the overlapping community structure [22, 23, 25]. We determine the centrality of modules, which reflects the hierarchical structure of the networks in the sense of subordination [4]–[6], not nestedness [30]–[33]. Then, we show that module membership is a chief determinant of the centrality of individual nodes. A node tends to be central when it belongs to a high-rank module and it is locally central by, for example, having a large degree. To clarify these points, we analyse centrality in modular networks. On the basis of the matrix tree theorem, the centrality value of a node is derived from the number of spanning trees rooted at the node. We use this relationship to develop an approximation scheme for the ranking-type centrality values of nodes in modular networks. The approximated value turns out to be a combination of local and global effects, i.e. the degree of nodes and the centrality of modules. For analytical tractability, we formulate our theory using the ranking-type centrality measure called the *influence*, but the results are also applicable to the PageRank. We corroborate the effectiveness of the proposed scheme using the *Caenorhabditis elegans* (*C. elegans*) neural network, an email social network, and the WWW.

## 2. Ranking-type centrality measures

We consider a directed and weighted network of $N$ nodes denoted by $G = \{V, E\}$. A set of nodes is denoted by $V = \{1, \ldots, N\}$, and $E$ is a set of directed links, i.e. node $i$ sends a directed link to node $j$ with weight $w_{ij}$ if and only if $(i, j) \in E$. The weight represents the amplitude of the direct influence of node $i$ on node $j$. We set $w_{ij} = 0$ when $(i, j) \notin E$.

Depending on applications, different centrality measures can be used to rank the nodes in a network. We analyze the effect of the modular structure on ranking of nodes using a centrality measure called *influence* because it facilitates theoretical analysis. The existence of a one-to-one mapping from the influence to the PageRank [7, 8, 17, 18] and to variations of the PageRank used for ranking academic journals and articles [11], [13]–[15], which we will explain in this

section, enables us to adapt our results to each case of such ranking-type centrality measures. To show that our results are not specific to the proposed measure, we study the influence and the PageRank simultaneously.

We define the influence of node $i$, denoted by $v_i$, by the solution of the following set of $N$ linear equations:

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

where $k_i^{\text{in}} \equiv \sum_{j'=1}^{N} w_{j'i}$ is the indegree of node $i$, and $\sum_{i=1}^{N} v_i = 1$ provides the normalization. $v_i$ is large if (i) node $i$ directly affects many nodes (i.e. many terms probably with a large $w_{ij}$ on the rhs of equation (1)), (ii) the nodes that receive directed links from node $i$ are influential (i.e. large $v_j$ on the rhs), and (iii) node $i$ has a small indegree.

Equation (1) is the definition for strongly connected networks; $G$ is defined to be strongly connected if there is a path, i.e. a sequence of directed links, from any node $i$ to any node $j$. If $G$ is not strongly connected, there is no path from a certain node $i$ to a certain node $j$. Then, node $i$ cannot influence node $j$ even indirectly, and the problem of determining the influence of nodes is reduced to considering each strongly connected component individually. Therefore, we assume that $G$ is strongly connected.

The influence $v_i$ represents the importance of nodes in different types of dynamics on networks (see appendix A for details). Firstly, $v_i$ is equal to the fixation probability of a new opinion introduced at node $i$ in a voter-type interacting particle system [20]. Secondly, if all links are reversed so that a random walker visits influential nodes with high probabilities, then $v_i$ is the stationary density of the continuous-time simple random walk. Thirdly, $v_i$ is the so-called reproductive value used in population ecology [29, 34]. Fourthly, $v_i$ is the contribution of an opinion at node $i$ to the opinion of the entire population in the consensus in the continuous-time version of the DeGroot model [35]–[37]. Fifthly, $v_i$ is equal to the amplitude of the collective response in the synchronized dynamics when an input is given to node $i$ [38].

The influence can be mapped to the PageRank. The PageRank, denoted by $R_i$ for node $i$, is defined self-consistently by

$$R_i = \frac{q}{N} + (1-q) \sum_{j=1}^{N} \frac{w_{ji}}{k_j^{\text{out}}} R_j + \delta_{k_i^{\text{out}},0}(1-q) R_i, \quad (1 \leqslant i \leqslant N), \tag{2}$$

where $k_i^{\text{out}} \equiv \sum_{j'=1}^{N} w_{ij'}$ is the outdegree of node $i$, $\delta_{i,j} = 1$ if $i = j$, and $\delta_{i,j} = 0$ if $i \neq j$. The second term on the rhs of equation (2) is present only when $k_j^{\text{out}} > 0$. Note that the direction of the link in the PageRank has a meaning opposite to that in the influence; $R_i$ of a webpage is incremented by an incoming link (hyperlink), whereas $v_i$ is incremented by an outgoing link. The introduction of $q > 0$ homogenizes $R_i$ and is necessary for the PageRank to be defined for directed networks that are not strongly connected, such as real web graphs. The normalization is given by $\sum_{i=1}^{N} R_i = 1$. $R_i$ is regarded as the stationary density of the discrete-time simple random walk on the network [7, 8, 17], where $q$ is the probability of a jump to a randomly selected node.

An essential difference between the two measures lies in normalization. In the influence, the total credit that node $j$ gives its neighbors is equal to $\sum_{i=1}^{N} w_{ij} v_j = k_j^{\text{in}} v_j$, while that in the PageRank is equal to $\sum_{i=1}^{N} (w_{ji}/k_j^{\text{out}}) R_j = R_j$. In the PageRank, the multiplicative factor of the total credit that node $j$ gives other nodes is set to $\sum_{i=1}^{N} w_{ji}/k_j^{\text{out}} = 1$ to prevent nodes

with many outgoing links from biasing ranks of nodes. In the ranking of webpages, creation of a webpage $i$ with many hyperlinks does not indicate that node $i$ gives a large amount of credit to recipients of a link. Each neighbor of node $i$ receives the credit $R_i/k_i^{\text{out}}$ from node $i$. We should refer to the PageRank when nodes can select the number of recipients of credit (e.g. the WWW and citation-based ranking of academic papers and journals). We should use the influence when the importance of all links is proportional to their weights (e.g. opinion formation and synchronization mentioned above).

The PageRank is equal to the influence in a network modified from the original network $G$ (see appendix B for derivation). In particular, the PageRank in $G$ for $q = 0$ is given by

$$R_i = k_i^{\text{out}} v_i(G^{\text{rev}}), \tag{3}$$

where $v_i(G^{\text{rev}})$ is the influence of node $i$ for the network $G^{\text{rev}}$, which is obtained by reversing all links of $G$. We use this relation to extend our results for the influence to the PageRank.

The influence has a nontrivial sense only in directed networks because $w_{ij} = w_{ji}$ in equation (1) leads to $v_i = 1/N$ [20, 39, 40]. Furthermore, any network with $k_i^{\text{in}} = k_i^{\text{out}}$ ($1 \leqslant i \leqslant N$) results in $v_i = 1/N$. Therefore, from equation (3), $R_i = k_i^{\text{in}}/(\langle k \rangle N)$ for such a network, where $\langle k \rangle \equiv \sum_{i=1}^{N} k_i^{\text{in}}/N = \sum_{i=1}^{N} k_i^{\text{out}}/N$ is the mean degree. In this case, $v_i$ and $R_i$ are not affected by the global structure of the network.

In directed or asymmetrically weighted networks, $v_i$ and $R_i$ are heterogeneous in general. The mean-field approximation (MA) is the simplest ansatz based on the local property of a node. By using $\sum_{j=1}^{N} w_{ij} v_j \approx \sum_{j=1}^{N} w_{ij} \bar{v} = k_i^{\text{out}} \bar{v}$, where $\bar{v} \equiv \sum_{i=1}^{N} v_i/N = 1/N$, we obtain $v_i \propto k_i^{\text{out}}/k_i^{\text{in}}$. Combination of this and equation (3) yields the MA for the PageRank: $v_i \approx k_i^{\text{in}}/(\langle k \rangle N)$.

We can calculate $v_i$ by enumerating spanning trees. To show this, note that equation (1) implies that $v_i$ is the left eigenvector with eigenvalue zero of the Laplacian matrix defined by $L_{ii} = \sum_{j'=1}^{N} w_{j'i}$ and $L_{ij} = -w_{ji}$ ($i \neq j$), i.e.

$$\sum_{i=1}^{N} v_i L_{ij} = 0, \quad (1 \leqslant j \leqslant N). \tag{4}$$

The $(i, j)$ cofactor of $L$ is defined by

$$\text{Co}(i, j) \equiv (-1)^{i+j} \det L(i, j), \tag{5}$$

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 $\sum_{j=1}^{N} L_{ij} = 0$, ($1 \leqslant i \leqslant N$), $\text{Co}(i, j)$ does not depend on $j$. Using equation (5) and the fact that $L$ is degenerate, we obtain

$$\sum_{i=1}^{N} \text{Co}(i, i) L_{ij} = \sum_{i=1}^{N} \text{Co}(i, j) L_{ij}$$
$$= \det L = 0, \quad (1 \leqslant j \leqslant N). \tag{6}$$

Therefore, $(\text{Co}(1, 1), \ldots, \text{Co}(N, N))$ is the left eigenvector of $L$ with eigenvalue zero, which yields

$$v_i \propto \text{Co}\,(i, i) = \det L(i, i). \tag{7}$$

From the matrix tree theorem [41, 42], $\det L(i, i)$ is equal to the sum of the weight of all possible directed spanning trees rooted at node $i$. The weight of a spanning tree is equal to the product of the weight of the $N - 1$ links forming the spanning tree.
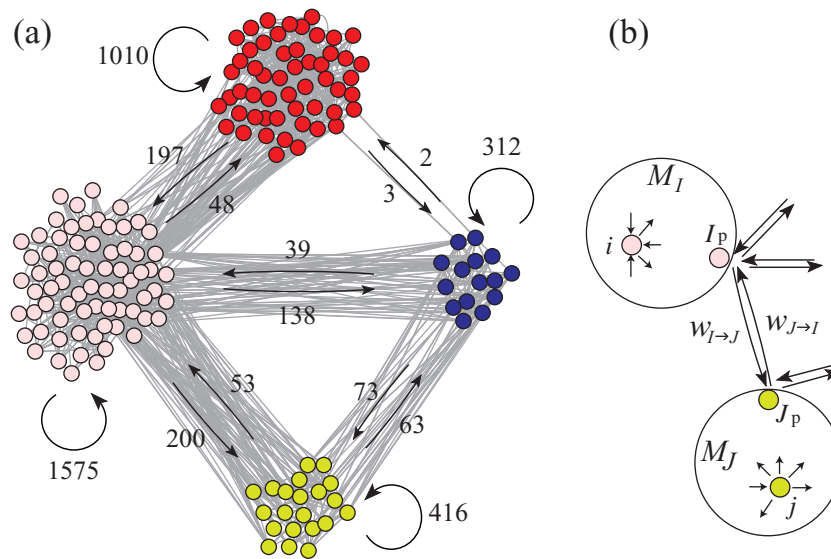
**Figure 1.** (a) A part of the *C. elegans* neural network composed of the four largest modules. The link weight is equal to the sum of the number of chemical synapses and that of the gap junctions. The original network has 274 nodes, 2959 links and 13 modules, while the depicted subnetwork has 159 nodes and 1363 links. The values indicate the summed link weights from one module to another. (b) Approximation of intermodular connectivity by links between portal nodes.

## 3. Centrality in modular directed networks

Most directed networks in the real world are more structured than those captured by the MA. A ubiquitous global structure of networks is a modular structure. Modular networks consist of several densely connected subgraphs called modules (also called communities), and modules are connected to each other by relatively few links. As an example, a subnetwork of the *C. elegans* neural network [43, 44] containing four modules is shown in figure 1(a). Modular structure is common in both undirected [21]–[24] and directed [24]–[27] networks.

Modular structure of directed networks often leads to hierarchical structure. By hierarchy, we refer to modules being located at different levels in terms of the value of the ranking-type centrality. It is relatively easy to traverse from a node in an upper level to one in a lower level along directed links, but not vice versa. The hierarchical structure leads to the deviation of $v_i$ from the value obtained from the MA.

As an example, consider the directed $P$-partite network shown in figure 2. Layer $P'$ ($1 \leqslant P' \leqslant P$) contains $N/P$ nodes, where $N$ is divided by $P$. The nodes in the same layer are connected bidirectionally with weight $w$. Each node in layer $P'$ ($1 \leqslant P' \leqslant P - 1$) sends directed links to all nodes in layer $P' + 1$ with weight unity, and each node in layer $P'$ ($2 \leqslant P' \leqslant P$) sends directed links to all the $N/P$ nodes in layer $P' - 1$ with weight $\epsilon$. The following results do not change if two adjacent layers are only connected by an asymmetrically weighted bridge, as shown in figure 1(b). Because of the symmetry, all nodes in layer $P'$ have the same influence
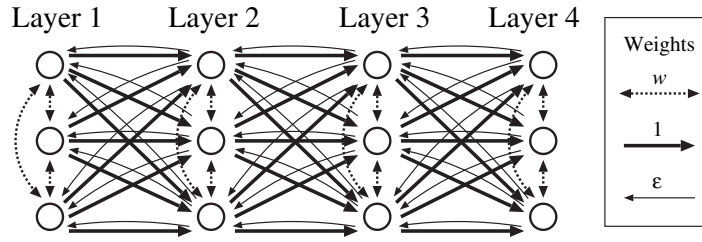
**Figure 2.** Hierarchical multipartite network with $N = 12$ and $P = 4$.

$v_{P'}$. From equation (1), we obtain

$$v_{P'} = \frac{\epsilon^{P'-1}(1-\epsilon)P}{(1-\epsilon^P)N}. \tag{8}$$

When $\epsilon < 1$, a node in a layer with small $P'$ is more influential than a node in a layer with large $P'$. The MA yields

$$\frac{k_i^{\text{out}}}{k_i^{\text{in}}} = \begin{cases} \epsilon^{-1}, & (\text{node } i \in \text{layer } 1), \\ \epsilon, & (\text{node } i \in \text{layer } P), \\ 1, & (\text{otherwise}). \end{cases} \tag{9}$$

The actual $v_{P'}$ decreases exponentially throughout the hierarchy, whereas $k_i^{\text{out}}/k_i^{\text{in}}$ does not. We observe a similar discrepancy in the case of the PageRank.

We develop an improved approximation for the influence in modular networks by combining the MA and the correction factor obtained from the global modular structure of networks. Consider a network of $m$ modules $M_I$ ($1 \leqslant I \leqslant m$). For mathematical tractability, we assume that each module communicates with the other modules via a single portal node $I_{\text{p}} \in M_I$, as illustrated in figure 1(b); the network shown in figure 1(b) is an approximation of that shown in figure 1(a). We denote the weight of the link $(I_{\text{p}}, J_{\text{p}})$ by $w_{I \to J}$ ($I \neq J$).

We obtain $v_i$ in this modular network by enumerating spanning trees rooted at node $i \in M_I$. We denote such a spanning tree by $T$. The intersection of $T$ and $M_I$ is a spanning tree restricted to $M_I$ and rooted at node $i$. This restricted spanning tree reaches all nodes in $M_I$. $T$ enters $M_J$ ($J \neq I$) via a directed path from node $I_{\text{p}}$ to node $J_{\text{p}}$. This path is provided by a spanning tree in the network of $m$ modules, where each module is represented by a single node. The other nodes in $M_J$ are spanned by the intersection of $T$ and $M_J$, which forms a spanning tree restricted to $M_J$ and rooted at node $J_{\text{p}}$. Therefore, $T$ is a concatenation of (i) an intramodular spanning tree in $M_I$ and rooted at node $i$, (ii) $m-1$ intramodular spanning trees in $M_J$ and rooted at node $J_{\text{p}}$ ($J \neq I$), and (iii) a spanning tree in the network of $m$ modules rooted at node $I_{\text{p}}$. Let $\mathcal{N}_\ell(M_I)$ ($\ell$ for local) denote the number of spanning trees in $M_I$ with an arbitrary root, and $\mathcal{N}_{\text{g}}$ (g for global) denote the number of spanning trees in a network of $m$ modules with an arbitrary root. Then, the number of spanning trees in $G$ rooted at node $i$ is equal to

$$\mathcal{N}_\ell(M_I)v_i^\ell \left[ \prod_{J=1, J\neq I}^{m} \left( \mathcal{N}_\ell(M_J) v_{J_{\text{p}}}^\ell \right) \right] \mathcal{N}_{\text{g}} v_{M_I}^{\text{g}}, \tag{10}$$

where $v_i^\ell$ is the influence of node $i \in M_I$ within $M_I$ and $v_{M_I}^{\text{g}}$ is the influence of $M_I$ in the network of $m$ modules. The first, second and third factors in equation (10) correspond to the numbers of

spanning trees of types (i)–(iii), respectively. Therefore, we obtain

$$v_i \propto v_i^\ell \left( \prod_{J=1, J \neq I}^{m} v_{J_\mathrm{p}}^\ell \right) v_{M_I}^\mathrm{g}. \tag{11}$$

For nodes $i, i' \in M_I$, equation (11) yields $v_i/v_{i'} = v_i^\ell/v_{i'}^\ell$; the relative influence of nodes in the same module is equal to their relative influence within the module. For nodes in different modules, i.e. node $i$ in module $M_I$ and node $j$ in module $M_J$ ($I \neq J$), equation (11) leads to

$$\frac{v_i}{v_j} = \frac{v_i^\ell v_{M_I}^\mathrm{g} v_{J_\mathrm{p}}^\ell}{v_j^\ell v_{M_J}^\mathrm{g} v_{I_\mathrm{p}}^\ell}. \tag{12}$$

If each module is homogeneous, we approximate $v_i^\ell \approx v_{I_\mathrm{p}}^\ell$, $v_j^\ell \approx v_{J_\mathrm{p}}^\ell$ and obtain $v_i/v_j \approx v_{M_I}^\mathrm{g}/v_{M_J}^\mathrm{g}$; the global structure of the network laid out by links across modules determines the influence of each node. If each module is heterogeneous in degree, we use the MA, i.e. $v_i^\ell \approx (k_i^\mathrm{out}/k_i^\mathrm{in})/\sum_{i';\mathrm{node}\, i' \in M_I}(k_{i'}^\mathrm{out}/k_{i'}^\mathrm{in})$ and $v_j^\ell \approx (k_j^\mathrm{out}/k_j^\mathrm{in})/\sum_{j';\mathrm{node}\, j' \in M_J}(k_{j'}^\mathrm{out}/k_{j'}^\mathrm{in})$. By assuming that $I_\mathrm{p}$ ($J_\mathrm{p}$) is a typical node in $M_I$ ($M_J$), we set $v_{I_\mathrm{p}}^\ell \approx 1/\sum_{i';\mathrm{node}\, i' \in M_I}(k_{i'}^\mathrm{out}/k_{i'}^\mathrm{in})$ and $v_{J_\mathrm{p}}^\ell \approx 1/\sum_{j';\mathrm{node}\, j' \in M_J}(k_{j'}^\mathrm{out}/k_{j'}^\mathrm{in})$. Then, equation (12) is transformed into

$$\frac{v_i}{v_j} \approx \frac{\left(k_i^\mathrm{out}/k_i^\mathrm{in}\right) v_{M_I}^\mathrm{g}}{\left(k_j^\mathrm{out}/k_j^\mathrm{in}\right) v_{M_J}^\mathrm{g}}. \tag{13}$$

Therefore, we define an approximation scheme, called the MA–Mod, for node $i$ in module $M_I$ as

$$v_i \propto \frac{k_i^\mathrm{out}}{k_i^\mathrm{in}} v_{M_I}^\mathrm{g}. \tag{14}$$

Equation (14) can be used for general modular networks in which different modules can be connected by more than one link.

Two crucial assumptions underlie equation (14). Firstly, a module is assumed to be an uncorrelated and possibly heterogeneous random network so that the MA is effective within the module. Note that the degree of nodes can be heterogeneously distributed. Secondly, most links are assumed to be intramodular so that the local MA is simply given by $v_i^\ell \propto k_i^\mathrm{out}/k_i^\mathrm{in}$.

To obtain $v_{M_I}^\mathrm{g}$ for general networks, we define $w_{I \to J} = \sum_{i \in M_I, j \in M_J} w_{ij}$ and approximate $v_i \approx v_{M_I}^\mathrm{g}/[\sum_{I'=1}^{m} N_{I'} v_{M_{I'}}^\mathrm{g}]$ (node $i \in M_I$), where $N_{I'}$ is the number of nodes in $M_{I'}$. Then, $\sum_{i=1}^{N} v_i = 1$ is satisfied. Equation (1) is transformed into

$$v_{M_I}^\mathrm{g} = \frac{\sum_{J \neq I} w_{I \to J} v_{M_J}^\mathrm{g}}{k_I^\mathrm{in}}, \quad (1 \leqslant I \leqslant m), \tag{15}$$

where

$$k_I^\mathrm{in} \equiv \sum_{J' \neq I} w_{J' \to I} = \sum_{i \in M_I, j \notin M_I} w_{ji}. \tag{16}$$

Equation (15) has the same form as equation (1). By solving the set of $m$ linear equations, we obtain $v_{M_I}^\mathrm{g}$.

Equation (14) adapted for $v_i(G^{\mathrm{rev}})$ leads to $v_i(G^{\mathrm{rev}}) \approx (k_i^{\mathrm{in}}/k_i^{\mathrm{out}})v_{M_I}^g(G^{\mathrm{rev}})$. By combining this with equation (3), we obtain the MA–Mod scheme for the PageRank with $q = 0$:

$$R_i \propto \frac{k_i^{\mathrm{in}}}{k_I^{\mathrm{out}}} R_{M_I}^g, \tag{17}$$

where

$$k_I^{\mathrm{out}} \equiv \sum_{J' \neq I} w_{I \to J'} = \sum_{i \in M_I, j \notin M_I} w_{ij}. \tag{18}$$

## 4. Application to real data

We examine the effectiveness of the MA–Mod scheme using three datasets from different fields.

### 4.1. Neural network

In the network of nematode *C. elegans*, a pair of neurons may be connected by chemical synapses, which are directed links, or by gap junctions, which are undirected links. We calculate the influence of neurons on the basis of a connectivity dataset [43, 44]. The link weight $w_{ij}$ is assumed to be the sum of the number of chemical synapses from neuron $i$ to neuron $j$ and that of the gap junctions between neurons $i$ and $j$. The following results are qualitatively the same if we ignore the gap junction or the link weight (see appendix C for the results). The largest strongly connected component, which we simply call the neural network, contains 274 nodes and 2959 links.

It is difficult to determine whether the influence or the PageRank is more appropriate from current biological evidence. If postsynaptic neurons linearly integrate different synaptic inputs, the influence may be an appropriate measure. In contrast, post-synaptic neurons may effectively select one synaptic input by a nonlinear mechanism. If each input is selected with the same probability in a long run and the activity level does not differ much across neurons, the PageRank may be appropriate. We examine both scenarios using power iteration (see appendix D for the methodology).

Among 274 neurons, 54, 79 and 87 neurons are classified as sensory neurons, interneurons and motor neurons, respectively [44]. By definition, sensory neurons directly receive external input such as touch and chemical substances, motor neurons send direct commands to move the body, and interneurons mediate information processing in various ways. The other neurons are polymodal neurons or neurons whose functions are unknown. Neurons with a large $v_i$ are mostly sensory neurons. For example, among the ten neurons with the largest $v_i$, eight are sensory neurons (ALMR, ASJL, ASJR, AVM, IL2VL, PHAL, PHAR and PVM) and two are interneurons (AIML and AIMR). Generally speaking, these neurons have a large $v_i$ not simply because their $k_i^{\mathrm{out}}/k_i^{\mathrm{in}}$ is large. The average of $v_i/[(k_i^{\mathrm{out}}/k_i^{\mathrm{in}})/\sum_{j=1}^N (k_j^{\mathrm{out}}/k_j^{\mathrm{in}})]$ over the ten neurons is equal to 3.456 (see table C.2 in appendix C for the values for individual neurons). These neurons are located at upper levels of the neural network in the global sense. The conclusion remains qualitatively the same if we use $R_i(G^{\mathrm{rev}})$. Recall that the PageRank is calculated for $G^{\mathrm{rev}}$ because the meaning of the direction of the link in the influence is opposite to that in the PageRank.

The average value of $v_i$ ($R_i(G^{\mathrm{rev}})$) for sensory neurons interneurons, and motor neurons is equal to 0.009235 (0.006621), 0.003614 (0.005415) and 0.001032 (0.001323), respectively.
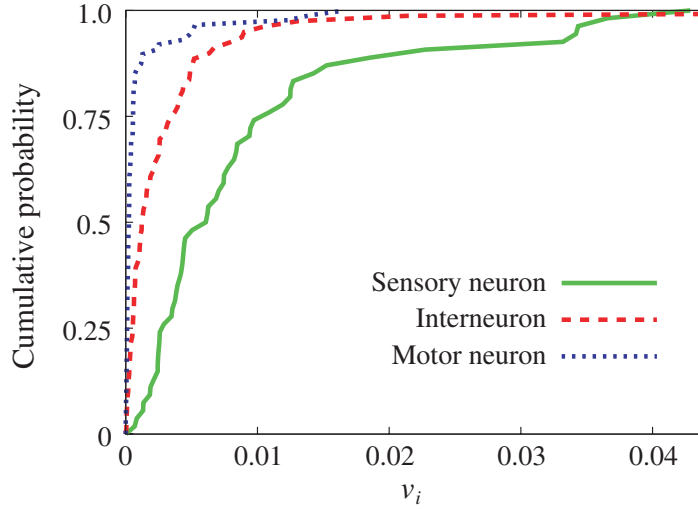
**Figure 3.** Cumulative distribution of $v_i$ for 54 sensory neurons, 79 interneurons and 87 motor neurons in *C. elegans* neural network.

The cumulative distributions of $v_i$ for different classes of neurons are shown in figure 3. Even though many synapses from motor neurons to interneurons and sensory neurons exist, as well as synapses from interneurons to sensory neurons, these numerical results indicate that the neural network is principally hierarchical. Generally speaking, sensory neurons, which directly receive external stimuli, are located at upper levels of the hierarchy, motor neurons are located at lower levels, and interneurons are located in between. Sensory neurons serve as a source of signals flowing to interneurons and motor neurons down the hierarchy.

The relation between $v_i$ and the MA is shown in figure 4(a) by the squares. They appear to be strongly correlated. However, the Pearson correlation coefficient (PCC; see appendix E for definition) between $v_i$ and the MA is not large ($= 0.5389$), as shown in table 1, because $v_i$ tends to be larger than the MA for nodes with a large $v_i$. Note that the data are plotted in the log–log scale in figure 4.

The neural network has a modular structure [45]. To use the MA–Mod scheme (equation (14)), we apply a community detection algorithm [27] to the neural network. We have selected this algorithm [27] because a directed link in the present context indicates the flow rather than the connectedness on which a recent algorithm [26] is based. As a result, we obtain $m = 13$ modules, calculate $v_{M_1}^g$, ..., $v_{M_m}^g$ from the network of the $m$ modules, and use equation (14). $v_i$ is plotted against the MA–Mod in figure 4(a), indicated by circles. The data fitting has improved compared to the case of the MA, in particular for small values of $v_i$. The PCC between $v_i$ and the MA–Mod is greater than that between $v_i$ and the MA (table 1). In this example, this holds true for the raw data and the logarithmic values of the raw data. As a benchmark, we assess the performance of the global estimator $v_i \approx v_{M_I}^g / [\sum_{I'=1}^{m} N_{I'} v_{M_{I'}}^g]$ (node $i \in M_I$), which we call the Mod. The Mod ignores the variability of $v_i$ within the module and is exact for networks with completely homogeneous modules, such as the network shown in figure 2. The performance of the Mod is poor in the neural network, as indicated by the triangles in figure 4(a) and the PCC listed in table 1.

The values of the PCC between the actual and approximated $R_i(G^{\text{rev}})$ are also listed in table 1. The results for the PageRank are qualitatively the same as those for the influence. With both measures, the module membership is a crucial determinant of centrality in individual nodes.
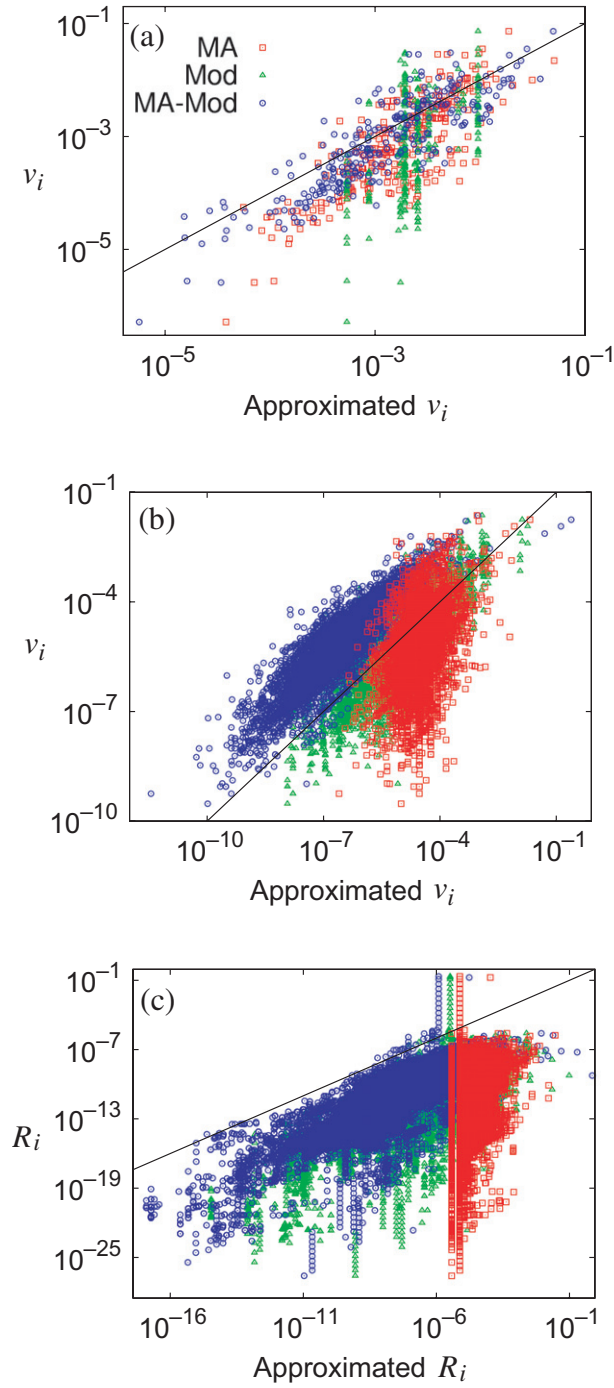
**Figure 4.** (a) $v_i$ for neural network, (b) $v_i$ for email social network and (c) $R_i$ for WWW with $q = 0$. The quantities placed on the horizontal axis are the MA (i.e. the normalized $k_i^{out}/k_i^{in}$ for $v_i$ and the normalized $k_i^{in}$ for $R_i$) (red squares), Mod (green triangles) and MA–Mod (blue circles).

Note that, on the basis of the Mod for the influence given by

$$v_i \approx \frac{v_{M_I}^{g}}{\sum_{I'=1}^{m} N_{I'} v_{M_{I'}}^{g}}, \tag{19}$$

**Table 1.** The PCC between the centrality measures and different estimators.

| Network | *C. elegans* | | Email | | WWW | | |
|---|---|---|---|---|---|---|---|
| $N$ | 274 | | 9079 | | 53 968 | | |
| $m$ | 13 | | 637 | | 2977 | | |
| Centrality | $v_i$ | $R_i(G^{\text{rev}})$ | $v_i$ | $R_i(G^{\text{rev}})$ | $v_i(G^{\text{rev}})$ | $R_i$ | $R_i$ |
| $q$ | N/A | 0 | N/A | 0 | N/A | 0 | 0.15 |
| MA | 0.5389 | 0.3593 | 0.5066 | 0.3997 | 0.0073 | 0.0007 | 0.2162 |
| Mod | 0.2927 | 0.4346 | 0.5010 | 0.2452 | 0.0003 | −0.0003 | 0.4104 |
| MA–Mod | 0.7295 | 0.5005 | 0.5066 | 0.2671 | 0.0000 | 0.0000 | 0.3166 |
| MA (log) | 0.8024 | 0.7073 | 0.3636 | 0.5353 | 0.3109 | 0.1289 | 0.4627 |
| Mod (log) | 0.5195 | 0.5503 | 0.8075 | 0.7147 | 0.7800 | 0.7147 | 0.4098 |
| MA–Mod (log) | 0.8736 | 0.8252 | 0.8798 | 0.9022 | 0.7964 | 0.7812 | 0.6256 |

the Mod for the PageRank is given by

$$\frac{R_i(G^{\text{rev}})}{k_i^{\text{in}}} \approx \frac{R_{M_I}^{\text{g}}(G^{\text{rev}})}{k_I^{\text{in}} \sum_{I'=1}^{m} N_{I'} v_{M_{I'}}^{\text{g}}}, \tag{20}$$

i.e.

$$R_i(G^{\text{rev}}) \propto \frac{k_i^{\text{in}} R_{M_I}^{\text{g}}(G^{\text{rev}})}{k_I^{\text{in}}} \approx \frac{R_{M_I}^{\text{g}}(G^{\text{rev}})}{N_I}. \tag{21}$$

We approximate $k_i^{\text{in}}$ by $k_I^{\text{in}}/N_I$ because the information about local degree is unavailable for the Mod.

### 4.2. Email social network

Our second example is the largest strongly connected component of an email social network [46]. A directed link exists between a sender and a recipient of an email. The network has modular structure [25]. In the weighted network that we consider here, the link weight is defined by the number of emails. The following results do not qualitatively change even if we neglect the link weight (see appendix F). The largest strongly connected component has 9079 nodes and 23 808 links and is partitioned into 637 modules.

Whether the influence or the PageRank is appropriate for ranking nodes depends on assumptions about human behavior. If recipients spend the same amount of time on each incoming email (i.e. the link of weight unity), $v_i$ is relevant. In contrast, recipients may have a fixed amount of time for dealing with all incoming emails. Then, a recipient may equally distribute the total time available to each email depending on the number of incoming emails. Under this assumption, the PageRank is relevant. We analyze both $v_i$ and $R_i(G^{\text{rev}})$.

In figure 4(b), the values of $v_i$ are plotted against those obtained by different estimators. On the log–log scale, the MA–Mod performs considerably better than the MA. Remarkably, even the Mod, in which nodes in the same module share an estimated centrality value, performs better than the MA. This is a strong indication that the structure of the coarse-grained network of modules is a more important determinant of $v_i$ than the local structure (i.e. degree) in this example. The values of the PCC summarized in table 1 support our claim. The PCC for the

MA–Mod and the Mod is considerably larger than that for the MA on the logarithmic scale, which implies that the MA–Mod is especially effective for nodes with small $v_i$. The values of the PCC between $R_i(G^{\text{rev}})$ and the different estimators are listed in table 1. These results are qualitatively the same as those for $v_i$.

### 4.3. WWW

Our last example comprises the largest strongly connected components of a WWW dataset [47]. The original network contains 325 729 nodes and 1 469 680 links, and the largest strongly connected component contains 53 968 nodes and 296 229 links. The MA fits the PageRank (with $q > 0$) of some WWW data when nodes of the same degree are grouped together [17] but not other data [18]. Because of the modular structure of the WWW [25], the MA–Mod is expected to perform better than the MA.

In figure 4(c), $R_i$ for $q = 0$ is plotted against the MA, Mod and MA–Mod. For nodes with small PageRanks, the MA–Mod, and even the Mod, are considerably better correlated with $R_i$ than the MA is (note the use of the log–log scale in figure 4(c); also see table 1). These nodes are located at lower levels of hierarchy. The results are qualitatively the same if we use the influence (table 1). Note that we reverse the links and calculate $v_i(G^{\text{rev}})$ because a directed link in the WWW indicates an impact of the target node on the source node.

The MA–Mod for the PageRank can be extended to the case $q > 0$. From equation (2), the MA for the PageRank is given by

$$R_i \approx \frac{q}{N} + (1-q)\frac{k_i^{\text{in}}}{\langle k \rangle}\bar{R} = \frac{q + (1-q)(k_i^{\text{in}}/\langle k \rangle)}{N}, \tag{22}$$

which implies that $k_i^{\text{in}}$ in the MA for $q = 0$ is replaced by $q\langle k \rangle + (1-q)k_i^{\text{in}}$ for general $q$. We define the MA–Mod for $q > 0$ by

$$R_i \propto \frac{\left[q\langle k \rangle + (1-q)k_i^{\text{in}}\right]R_{M_I}^{\text{g}}}{q\sum_J k_J^{\text{out}}/m + (1-q)k_I^{\text{out}}}. \tag{23}$$

Note that this ansatz is heuristic, whereas equation (17) used for $q = 0$ has an analytical basis. The PCCs between the PageRank with $q = 0.15$ and the three estimates are listed in table 1. The MA–Mod performs better than the MA. The advantage of the MA–Mod over the MA is smaller for $q = 0.15$ than for $q = 0$, because a larger $q$ implies a heavier neglect of the network structure.

The definition of the PageRank given by equation (2) is not continuous with respect to the outdegree; the term $\delta_{k_i^{\text{out}},0}(1-q)R_i$ is present for $k_i^{\text{out}} = 0$ (i.e. dangling node) and absent for $k_i^{\text{out}} > 0$. Therefore, dangling nodes can have large PageRanks. To improve the MA–Mod for $q > 0$, we should separately treat dangling nodes and other nodes in the same module. We do not explore this point because this situation seems to be specific to the working definition of the PageRank.

In practice, nodes with a small $R_i$ could be irrelevant to the performance of a search engine, which outputs a list of websites with the largest PageRanks. However, nodes with small PageRanks constitute the majority of a network when the PageRank follows a power-law distribution. This is the case for the real WWW data, which are scale-free networks [17, 18]. Our method is considerably better than the MA, especially for nodes with small PageRanks.

In general, the WWW is nested, with each level defined by webpages, directories, hosts and domains. At the host level, for example, most links are directed toward nodes within the same

host [48]. Therefore, a host can be regarded as a module in the network. By calculating the importance of the host, called the BlockRank, the PageRank can be efficiently computed [48]. In spirit, our $R_{M_I}^{g}$ is similar to the BlockRank, although our $R_{M_I}^{g}$ is used for identifying the hierarchical levels of networks and systematically approximating $R_i$.

It should be noted that, in general, our approximation scheme runs much faster than the direct calculation of $v_i$ or $R_i$ for large networks. This is because the community detection algorithm [27] is fast and the power iteration used for calculating $v_i$ and $R_i$ converges more quickly for a smaller network in most (but not all) cases. In the WWW, which is a large network, our approximation scheme for the PageRank with $q = 0$ ran more than 100 times faster than the direct calculation on our computer.

## 5. Discussion and conclusions

We have shown that the hierarchical structure of directed modular networks considerably affects ranking-type centrality measures of individual nodes. Using the information about connectivity among modules, we have significantly improved the estimation of centrality values. Our theoretical development is based on the measure that we have proposed (i.e. influence), but the conclusions hold true for both the influence and the PageRank. Our method can be implemented for variants of the PageRank including the eigenfactor [14, 15] and the so-called invariant method [11, 13] used for ranking academic journals.

The hierarchy discussed in this study is different from the nestedness of networks. Many networks are hierarchical in the sense that they are nested and have multiple scales [30]–[33]. A modular network is hierarchical in this sense, at least to a limited extent; two hierarchical levels are defined by the scale of the entire network and that of a single module. In contrast, we are concerned with hierarchical relationships among modules defined by the directionality of networks. This concept of hierarchy has been studied with relation to, for example, food webs [4], transcription networks [5], and social dynamics [6], but its application to networks is relatively unrewarding in spite of its intuitive appeal. The influence and the PageRank quantify the hierarchical position of individual nodes and of modules.

In real networks, nodes and links are subject to change. Change affects nodes near the perturbed nodes, but may not significantly affect modules. In social networks, large groups change slowly over time compared to small groups [23]. In addition, where a knowledge of a network is incomplete, a modest understanding at the level of the modular structure may be adequate. Nodes in a module may also have a common function. These are main reasons for investigating the modular structure of networks. We have shown that the modular structure is also important in the context of directed networks, hierarchy and ranking. Defining a module is complex in the case of directed networks when compared to undirected networks, and module detection in directed networks is currently under investigation (see [24] for a review). We hope that our results aid the development of the concept of modules and related algorithms in directed networks.

## Acknowledgments

## Appendix A. The influence is obtained from various dynamical models on networks

### A.1. Fixation probability of evolutionary dynamics

$v_i$ represents the probability that an 'opinion' introduced at node $i$ spreads to the entire network. We consider stochastic competitive dynamics between two equally strong types of opinion $A$ and $B$; each node takes either $A$ or $B$ at a given time. In the so-called link dynamics (LD) [39, 40], which is a network version of the standard voter model, one link $(i, j) \in E$ is selected for reproduction with an equal probability in each time step. Then, the type at node $i$ replaces that at node $j$. This process is repeated until $A$ or $B$ takes over the entire network.

$v_i$ coincides with the fixation probability denoted by $F_i^{\text{LD}}$, which is the probability that a new type $A$ introduced at node $i$ in the network of the resident-type $B$ nodes takes over the entire network [20]. To calculate $F_i^{\text{LD}}$, fix a network and consider the initial configuration in which $A$ is located at node $i$ and $B$ is located at the other $N - 1$ nodes. In the first time step, one of the following events occurs. With the probability $w_{ij} / \sum_{i',j'} w_{i'j'}$, the link $(i, j) \in E$ is selected for reproduction. Then, type $A$ is located at nodes $i$ and $j$. Let $F_{\{i,j\}}^{\text{LD}}$ denote the fixation probability of type $A$ for this new configuration. With the probability $w_{ji} / \sum_{i',j'} w_{i'j'}$, the link $(j, i) \in E$ is selected, type $A$ becomes extinct, and the dynamics terminates. With the remaining probability $\sum_{i' \neq i, j' \neq i} w_{i'j'} / \sum_{i',j'} w_{i'j'}$, the configuration of types $A$ and $B$ on the network does not change. Therefore, we obtain

$$F_i^{\text{LD}} = \sum_j \frac{w_{ij}}{\sum_{i',j'} w_{i'j'}} F_{\{i,j\}}^{\text{LD}} + \frac{\sum_j w_{ji}}{\sum_{i',j'} w_{i'j'}} \times 0 + \frac{\sum_{i' \neq i, j' \neq i} w_{i'j'}}{\sum_{i',j'} w_{i'j'}} F_i^{\text{LD}}. \tag{A.1}$$

Because $F_{\{i,j\}}^{\text{LD}} = F_i^{\text{LD}} + F_j^{\text{LD}}$ [20], equation (A.1) leads to equation (1) with $v_i$ replaced by $F_i^{\text{LD}}$.

### A.2. Continuous-time simple random walk

Consider a simple random walk on the network in continuous time. In a brief time interval $\Delta t$, a walker at node $i$ is attracted to its neighbor $j$, where $(j, i) \in E$, with the probability $\Delta t$. Note that the direction of the link is the opposite of the conventional one because the directed link in the present study indicates the influence of the source node of the link on the target node of the link. The master equation for the density of the random walker at node $i$, denoted by $F_i^{\text{RW}}$ ($1 \leqslant i \leqslant N$), is represented by

$$\frac{\mathrm{d}F_i^{\text{RW}}}{\mathrm{d}t} = \sum_{j=1}^{N} w_{ij} F_j^{\text{RW}} - k_i^{\text{in}} F_i^{\text{RW}}. \tag{A.2}$$

Because the network $G$ is strongly connected, $F_i^{\text{RW}}$ converges to the unique stationary density. By setting the lhs of equation (A.2) to 0, we obtain $F_i^{\text{RW}} = v_i$.

The simple random walk is closely associated with the fixation problem. The so-called dual process of the LD is the coalescing random walk. In the coalescing random walk, each of the $N$ walkers basically performs the continuous-time simple random walk on the network with the direction of all links reversed. Therefore, the random walker can traverse from node $i$ to node $j$ when $(j, i) \in E$. If two random walkers meet on a node, they coalesce into one walker. There is only one walker after a sufficiently long time, and the duality between the two stochastic processes guarantees $F_i^{\text{RW}} = F_i^{\text{LD}}$ [20].

## A.3. Reproductive value

In population ecology, the number of offspring an individual contributes to is quantified as the reproductive value of the individual. The reproductive value of node $i$ is defined by equation (1) [29, 34]. In practice, a node represents a class of individuals defined by, for example, sex, age, or habitat.

## A.4. DeGroot model in social dynamics

The DeGroot model [35]–[37] is a discrete-time model that represents the propagation of information or opinions in social systems. The state of the individual at node $i$ is represented by a real value $p_i(t)$; $p_i(t)$ parameterizes the information that the individual at node $i$ has at time $t$. The weight $w_{ij}$ is the probability that the individual at node $j$ copies the opinion at node $i$ in the next time step. The normalization is given by $\sum_{i=1}^{N} w_{ij} = 1$. The states of the $N$ nodes evolve according to

$$p_i(t) = \sum_{j=1}^{N} w_{ji} p_j(t-1). \tag{A.3}$$

If the network is strongly connected and aperiodic, a consensus is reached asymptotically, i.e. $p_1(\infty) = \ldots = p_N(\infty)$ [36, 37].

The extent to which the initial information at node $i$ influences the limiting common information in the continuous-time version of the DeGroot model is equal to $v_i$. To show this, we start with the discrete-time dynamics given by equation (A.3). Suppose that $F_i^{\mathrm{DG}}$ ($1 \leqslant i \leqslant N$) satisfies $p_1(\infty) = \ldots = p_N(\infty) = \sum_{i=1}^{N} F_i^{\mathrm{DG}} p_i(0)$ for arbitrary $p_1(0), \ldots, p_N(0)$. Because the configurations $\{p_1(0), \ldots, p_N(0)\}$ and $\{p_1(1), \ldots, p_N(1)\}$ starting with $\{p_1(0), \ldots, p_N(0)\}$ end up with the identical $p_1(\infty) = \ldots = p_N(\infty)$, we obtain

$$\sum_{i=1}^{N} F_i^{\mathrm{DG}} p_i(0) = \sum_{i=1}^{N} F_i^{\mathrm{DG}} \left( \sum_{j=1}^{N} w_{ji} p_j(0) \right). \tag{A.4}$$

Since $p_1(0), \ldots, p_N(0)$ are arbitrary, we obtain

$$F_i^{\mathrm{DG}} = \sum_{j=1}^{N} w_{ij} F_j^{\mathrm{DG}}, \tag{A.5}$$

Equation (A.5) is of the same form as equation (1). However, the condition $\sum_{i=1}^{N} w_{ij} = 1$ is imposed in equation (A.5) because the dynamics are defined in the discrete time. The continuous-time counterpart of the DeGroot model is defined in [37] as follows:

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t} = \sum_{j=1}^{N} w_{ji} \left( p_j(t) - p_i(t) \right). \tag{A.6}$$

If $p_1(\infty) = \ldots = p_N(\infty) = \sum_{i=1}^{N} F_i^{\mathrm{DG}} p_i(0)$, we obtain $\sum_{i=1}^{N} F_i^{\mathrm{DG}} \mathrm{d}p_i(t)/\mathrm{d}t = 0$, which leads to

$$\sum_{i=1}^{N} \left( \sum_{j=1}^{N} F_j^{\mathrm{DG}} w_{ij} - F_i^{\mathrm{DG}} \sum_{j=1}^{N} w_{ji} \right) p_i(0) = 0 \tag{A.7}$$

for arbitrary $p_1(0), \ldots, p_N(0)$. Therefore, $F_i^{\mathrm{DG}} = v_i$.

*A.5. Collective responses in coupled oscillator dynamics*

In accordance with [38], we consider $N$ coupled phase oscillators obeying

$$\dot{\phi}_i = \omega_i + \sum_{j=1}^{N} \Gamma_{ij} \left( \phi_i - \phi_j \right) + \sigma p_i(t), \quad i = 1, \ldots, N, \tag{A.8}$$

where $\phi_i \in [0, 2\pi)$ is the phase of the oscillator $i$, $\omega_i$ is the intrinsic frequency of the oscillator $i$, $\Gamma_{ij}$ is the effect of node $j$ on node $i$, and $p_i(t)$ is the input at time $t$ applied to node $i$. We assume that (i) in the absence of the input (i.e. $\sigma = 0$), the system is fully phase-locked, i.e. $\phi_i = \phi_i^0 + \Omega t$ for all $i$ with some constants $\phi_i^0$ and $\Omega$ and that (ii) the input is small, i.e. $\sigma \ll 1$, so that the system is always close to the phase-locked state. Using the synchronization condition, i.e. $\omega_i + \sum_{j=1}^{N} \Gamma_{ij}(\phi_i^0 - \phi_j^0) = \Omega$ $(1 \leqslant i \leqslant N)$, which is implied by assumption (i), we linearize equation (A.8) as

$$\dot{\psi}_i = \sum_{j=1}^{N} L_{ij} \psi_j + \sigma p_i(t), \tag{A.9}$$

where $\psi_i \equiv \phi_i - \phi_i^0 - \Omega t$ is a small perturbation in the phase, and $L$ is the Jacobian matrix given by $L_{ij} = [\sum_{j' \neq i} \Gamma'_{ij'}(\phi_i^0 - \phi_{j'}^0)]\delta_{ij} - \Gamma'_{ij}(\phi_i^0 - \phi_j^0)(1 - \delta_{ij})$. Note that the effective weight of the link from node $j$ to node $i$ is given by $w_{ji} = -\Gamma'_{ij}(\phi_i^0 - \phi_j^0)$. Because assumption (i) implies the stability of the phase-locked state, the real parts of all the eigenvalues of $L$ are negative, except a zero eigenvalue. We define the collective phase by $\Theta \equiv \sum_{i=1}^{N} v_i \phi_i$. Combination of $\sum_{i=1}^{N} v_i L_{ij} = 0$ $(1 \leqslant j \leqslant N)$, which is derived from equation (1), and equation (A.9) yields

$$\dot{\Theta} = \sum_{i=1}^{N} v_i \dot{\phi}_i = \Omega + \sigma \sum_{i=1}^{N} v_i p_i(t). \tag{A.10}$$

Assumption (ii) implies that $\dot{\phi}_1 \approx \ldots \approx \dot{\phi}_N \approx \dot{\Theta}$. Therefore, equation (A.10) describes the dynamical behavior of each oscillator and that of the entire network. The response of the collective behavior to the input applied to node $i$ is weighted by $v_i$.

## Appendix B. Relationship between the influence and the PageRank

To determine the relationship between the influence and the PageRank, we rewrite equation (2) as

$$R_i = \sum_{j=1}^{N} \left[ \frac{q}{N} + (1-q)\frac{w_{ji}}{k_j^{\text{out}}} + (1-q)\delta_{i,j}\delta_{k_i^{\text{out}},0} \right] R_j. \tag{B.1}$$

From the original network $G$, we define a complete and asymmetrically weighted network $G'$ using the matrix of link weights $w'_{ij} = q/N + (1-q)w_{ji}/k_j^{\text{out}} + (1-q)\delta_{i,j}\delta_{k_i^{\text{out}},0}$. Because $\sum_{j=1}^{N} w'_{ji} = 1 (1 \leqslant i \leqslant N)$, $R_i$ in $G$ is equal to $v_i$ in $G'$, which we denote by $v_i(G')$ for clarity. Because self-loops do not affect the calculation of the influence, we can replace $w'_{ij}$ by $q/N + (1-q)w_{ji}/k_j^{\text{out}}$.

In particular, $R_i$ for $q = 0$ is equal to $v_i(G')$, where $G'$ is defined by $w'_{ij} = w_{ji}/k_j^{\text{out}}$. In this case, the PageRank and the influence are connected by the simple relationship given by equation (3).

**Table C.1.** PCC between centrality measures and different estimators for *C. elegans* neural networks.

| Gap junction | Yes | | No | | No | |
|---|---|---|---|---|---|---|
| Link | Unweighted | | Weighted | | Unweighted | |
| $N$ | 274 | | 237 | | 237 | |
| $m$ | 7 | | 20 | | 15 | |
| Centrality | $v_i$ | $R_i(G^{\mathrm{rev}})$ | $v_i$ | $R_i(G^{\mathrm{rev}})$ | $v_i$ | $R_i(G^{\mathrm{rev}})$ |
| MA | 0.7420 | 0.6331 | 0.2145 | 0.0875 | 0.5153 | 0.4240 |
| Mod | 0.3727 | 0.2542 | 0.1577 | 0.1741 | 0.1583 | 0.2224 |
| MA–Mod | 0.8235 | 0.7401 | 0.3328 | 0.1189 | 0.4949 | 0.4659 |
| MA (log) | 0.8478 | 0.7942 | 0.6899 | 0.6152 | 0.7976 | 0.7726 |
| Mod (log) | 0.5190 | 0.1937 | 0.7018 | 0.7338 | 0.6743 | 0.6519 |
| MA–Mod (log) | 0.8995 | 0.8752 | 0.8475 | 0.8137 | 0.8785 | 0.8586 |

## Appendix C. Detailed analysis of the *Caenorhabditis elegans* neural network

The relative contributions of a chemical synapse and that of a gap junction to signal transduction in the *C. elegans* neural circuitry are unknown. In the main text, we have assumed that the neural network is a weighted network in which a chemical synapse has the same link weight as a gap junction. Here, we examine three other variants of the *C. elegans* neural network. In these three neural networks, we neglect the link weight and/or gap junctions. The omission of the link weight reflects the possibility that the intensity of the communication between two neurons may saturate as the number of synapses increases. The omission of gap junctions reflects the possibility that gap junctions may not contribute to signal processing as significantly as chemical synapses. Note that the largest strongly connected component shrinks to a network of 237 nodes with 1936 synapses by the omission of gap junctions.

For the three neural networks, the values of the PCC between the centralities of the nodes and the three approximators are listed in table C.1. We have examined both $v_i$ and $R_i(G^{\mathrm{rev}})$ with $q = 0$. In general, the MA–Mod predicts $v_i$ and $R_i(G^{\mathrm{rev}})$ better than the MA in the three networks. The results listed in table C.1 are consistent with those presented in the main text.

For the four neural networks, including the one in the main text, the ten most influential neurons are listed in table C.2. This list of ten neurons is largely consistent across different definitions of neural network. For the majority of these neurons, $v_i$ is larger than the value predicted from the MA.

## Appendix D. Power iteration

If we use a standard numerical method such as Gaussian elimination, the computation time required for calculating $v_i$ and $R_i$ from equations (1) and (2) is $O(N^3)$. For sparse networks carrying out power iteration (also called Jacobi iteration) may be much faster. The convergence of this iteration is guaranteed, as explained below for the influence. The proof for the PageRank is almost the same.

**Table C.2.** The most influential neurons in *C. elegans* neural networks. $v_i/(MA)$ indicates $v_i$ divided by the value obtained from the MA.

| With gap junction | | | | | | No gap junction | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Weighted | | | Unweighted | | | Weighted | | | Unweighted | | |
| Neuron | $v_i$ | $\frac{v_i}{(MA)}$ | Neuron | $v_i$ | $\frac{v_i}{(MA)}$ | Neuron | $v_i$ | $\frac{v_i}{(MA)}$ | Neuron | $v_i$ | $\frac{v_i}{(MA)}$ |
| AIMR | 0.08876 | 4.160 | PHAL | 0.04279 | 3.226 | AIMR | 0.06841 | 3.924 | VC04 | 0.05899 | 4.594 |
| ASJL | 0.04287 | 3.588 | PHAR | 0.04117 | 3.449 | ASJL | 0.04835 | 3.467 | VC05 | 0.04439 | 3.841 |
| ALMR | 0.03657 | 3.296 | AIMR | 0.04062 | 2.356 | ALMR | 0.03965 | 2.843 | AIMR | 0.03718 | 2.227 |
| PHAR | 0.03435 | 7.740 | ASIL | 0.02748 | 2.072 | VC04 | 0.03334 | 2.988 | AIML | 0.02722 | 1.325 |
| PHAL | 0.03419 | 6.259 | ASIR | 0.02695 | 2.540 | PVM | 0.03246 | 2.116 | AWAL | 0.02715 | 1.510 |
| ASJR | 0.03319 | 4.094 | AIML | 0.02152 | 1.432 | AVM | 0.02847 | 1.047 | AVG | 0.02426 | 2.951 |
| IL2VL | 0.02647 | 0.456 | IL2VL | 0.02061 | 0.706 | AIML | 0.02304 | 2.447 | AVM | 0.01715 | 1.028 |
| AVM | 0.02273 | 1.816 | ALMR | 0.01982 | 2.135 | AVG | 0.02257 | 8.826 | ASKR | 0.01701 | 3.975 |
| AIML | 0.02133 | 2.231 | VC05 | 0.01719 | 2.160 | ASJR | 0.02217 | 2.649 | ALMR | 0.01692 | 1.647 |
| PVM | 0.01860 | 1.816 | VC04 | 0.01505 | 2.838 | ADLL | 0.01777 | 0.593 | IL2VL | 0.01546 | 0.669 |

We rewrite equation (1) as

$$v_i = \sum_{j=1}^{N} \frac{w_{ij}}{\sum_{j'=1}^{N} w_{j'i}} v_j. \tag{D.1}$$

Equation (D.1) indicates that $v_i$ is the $i$th element of the right eigenvector of the matrix $M \equiv (M_{ij}) = (w_{ij}/\sum_{j'=1}^{N} w_{j'i})$ for the eigenvalue equal to unity. Multiplying $M$ by the diagonal matrix $(\delta_{ij}/\sum_{i'=1}^{N} w_{i'j})$ on the right and its inverse on the left does not alter the spectrum of $M$. This operation yields a new matrix whose $(i, j)$ element is given by $(w_{ij}/\sum_{i'=1}^{N} w_{i'j})$. The spectral radius of the new matrix is at most unity because its maximum row sum matrix norm [49, p 295] is equal to unity. Consequently, the spectral radius of $M$ is equal to unity.

Consider the power iteration scheme in which the $(t + 1)$th estimate of $v_i$ $(1 \leqslant i \leqslant N)$ is given by the rhs of equation (D.1) in which the $t$th estimate of $v_i$ $(1 \leqslant i \leqslant N)$ is substituted. If the network is strongly connected and aperiodic, the non-negative matrix $M$ is primitive, i.e. the eigenvalue of the largest modulus, which is equal to unity in the present case, is unique [49, p 516]. Then, the convergence of power iteration to the correct $(v_1, \ldots, v_N)$ is guaranteed [49, p 523]. The Perron–Frobenius theorem [49] guarantees that the Perron vector $(v_1, \ldots, v_N)$ is uniquely determined and that $v_i > 0$ $(1 \leqslant i \leqslant N)$. The power iteration converges quickly if the modulus of the second eigenvalue of $M$ is considerably smaller than that of the largest eigenvalue, i.e. unity.

## Appendix E. PCC

The PCC between $v_i$ and an estimator $v_i^{\text{est}}$, such as MA, Mod and MA–Mod, is defined by

$$\frac{1/N \sum_{i=1}^{N} \left( v_i v_i^{\text{est}} - 1/N^2 \right)}{\sqrt{1/N \sum_{i=1}^{N} (v_i - 1/N)^2} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( v_i^{\text{est}} - 1/N \right)^2}}. \tag{E.1}$$

Note that $\sum_{i=1}^{N} v_i/N = \sum_{i=1}^{N} v_i^{\text{est}}/N = 1/N$.

**IOP** Institute of Physics Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

**Table F.1.** PCC between centrality measures and different estimators for unweighed email social network.

| $N$ | 9079 | |
|---|---|---|
| $m$ | 599 | |
| Centrality | $v_i$ | $R_i(G^{\text{rev}})$ |
| $q$ | N/A | 0 |
| MA | 0.6628 | 0.5536 |
| Mod | 0.6537 | 0.3290 |
| MA–Mod | 0.6692 | 0.4774 |
| MA (log) | 0.2552 | 0.4203 |
| Mod (log) | 0.8719 | 0.7754 |
| MA–Mod (log) | 0.8898 | 0.9042 |

## Appendix F. Results for unweighed email social network

The values of the PCC between the two centrality measures and different estimators for the unweighed email social network are listed in table F.1. The results are qualitatively the same as those for the weighted network shown in the main text.

## References

[1] Wasserman S and Faust K 1994 *Social Network Analysis* (New York: Cambridge University Press)
[2] Newman M E J 2003 *SIAM Rev.* **45** 167
[3] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D-U 2006 *Phys. Rep.* **424** 175
[4] Garlaschelli D, Caldarelli G and Pietronero L 2003 *Nature* **423** 165
[5] Lagomarsino M C, Jona P, Bassetti B and Isambert H 2007 *Proc. Natl Acad. Sci. USA* **104** 5516
[6] Castellano C, Fortunato S and Loreto V 2009 *Rev. Mod. Phys.* **81** 591
[7] Brin S and Page L 1998 *Proc. 7th Int. World Wide Web Conf.* (*Brisbane, Australia, 14–8 April*) pp 107–17
[8] Berkhin P 2005 *Internet Math.* **2** 73
[9] Kleinberg J M 1999 *J. ACM* **46** 604
[10] Kleinberg J and Lawrence S 2001 *Science* **294** 1849
[11] Palacios-Huerta I and Volij O 2004 *Econometrica* **72** 963
[12] Chen P, Xie H, Maslov S and Redner S 2007 *J. Inform.* **1** 8
[13] Pinski G and Narin F 1976 *Inf. Process. Manage.* **12** 297
[14] Davis P M 2008 *J. Am. Soc. Inf. Sci. Technol.* **59** 2186
[15] Fersht A 2009 *Proc. Natl Acad. Sci. USA* **106** 6883
[16] Park J and Newman M E J 2005 *J. Stat. Mech.* P10014
[17] Fortunato S, Flammini A, Menczer F and Vespignani A 2006 *Proc. Natl Acad. Sci. USA* **103** 12684
[18] Donato D, Laura L, Leonardi S and Millozzi S 2004 *Eur. Phys. J.* B **38** 239
[19] Restrepo J G, Ott E and Hunt B R 2006 *Phys. Rev. Lett.* **97** 094102
[20] Masuda N and Ohtsuki H 2009 *New J. Phys.* **11** 033012
[21] Newman M E J 2004 *Eur. Phys. J.* B **38** 321
[22] Palla G, Derényi I, Farkas I and Vicsek T 2005 *Nature* **435** 814
[23] Palla G, Barabási L-A and Vicsek T 2007 *Nature* **446** 664
[24] Fortunato S 2009 *Phys. Rep.* in press

**IOP** Institute of Physics $\Phi$ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

[25] Palla G, Farkas I J, Pollner P, Derényi I and Vicsek T 2007 *New J. Phys.* **9** 186
[26] Leicht E A and Newman M E J 2008 *Phys. Rev. Lett.* **100** 118703
[27] Rosvall M and Bergstrom C T 2008 *Proc. Natl Acad. Sci. USA* **105** 1118
[28] Everett M G and Borgatti S P 1999 *J. Math. Sociol.* **23** 181
[29] Taylor P D 1990 *Am. Nat.* **135** 95
[30] Ravasz E, Somera A L, Mongru D A, Oltvai Z N and Barabási A-L 2002 *Science* **297** 1551
[31] Guimerà R and Amaral L A N 2005 *Nature* **433** 895
[32] Sales-Pardo M, Guimerà R, Moreira A A and Amaral L A N 2007 *Proc. Natl Acad. Sci. USA* **104** 15224
[33] Clauset A, Moore C and Newman M E J 2008 *Nature* **453** 98
[34] Taylor P D 1996 *J. Math. Biol.* **34** 654
[35] DeGroot M H 1974 *J. Am. Stat. Assoc.* **69** 118
[36] Jackson M O 2008 *Social and Economic Networks* (Princeton, NJ: Princeton University Press)
[37] Olfati-Saber R, Fax J A and Murray R M 2007 *Proc. IEEE* **95** 215
[38] Kori H, Kawamura Y, Nakao H, Arai K and Kuramoto Y 2009 *Phys. Rev.* E **80** 036207
[39] Antal T, Redner S and Sood V 2006 *Phys. Rev. Lett.* **96** 188104
[40] Sood V, Antal T and Redner S 2008 *Phys. Rev.* E **77** 041121
[41] Biggs N 1997 *Bull. L. Math. Soc.* **29** 641
[42] Agaev R P and Chebotarev P Y 2000 *Autom. Rem. Cont.* **61** 1424
[43] Chen B L, Hall D H and Chklovskii D B 2006 *Proc. Natl Acad. Sci. USA* **103** 4723
[44] http://www.wormatlas.org
[45] Müller-Linow M, Hilgetag C C and Hütt M-T 2008 *PLoS Comput. Biol.* **4** e1000190
[46] Ebel H, Mielsch L-I and Bornholdt S 2002 *Phys. Rev.* E **66** 035103
[47] Albert R, Jeong H and Barabási A-L 1999 *Nature* **401** 130
[48] Kamvar S D, Haveliwala T H, Manning C D and Golub G H 2003 *Proc. 12th Int. World Wide Web Conf.* May 2003
[49] Horn R A and Johnson C R 1985 *Matrix Analysis* (Cambridge: Cambridge University Press)