Chapter 2
Resistance Distance, Information Centrality, Node Vulnerability and Vibrations in Complex Networks

Ernesto Estrada and Naomichi Hatano

Abstract We discuss three seemingly unrelated quantities that have been introduced in different fields of science for complex networks. The three quantities are the resistance distance, the information centrality and the node displacement. We first prove various relations among them. Then we focus on the node displacement, showing its usefulness as an index of node vulnerability. We argue that the node displacement has a better resolution as a measure of node vulnerability than the degree and the information centrality.

2.1 Introduction

The study of complex networks is a truly multidisciplinary subject which covers many areas of nature, technology, and society [1, 27, 30]. These networks are graph-theoretic representations of complex systems in which the nodes of a graph represent the entities of the system and the links represent the relationship between them [1, 27, 30]. The use of the graphs for studying complex systems is not new. For instance, the study of social networks is a discipline with a long tradition of using graphs [19] and has provided many theoretical tools that are now used in the analysis of networks in many disciplines. In the physical sciences, graph analysis of relatively small systems has also been in use for long time. Some well known examples include the entire area of chemical graph theory [31] and the use of the graphs in

E. Estrada (✉)
Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK
e-mail: ernesto.estrada@strath.ac.uk

E. Estrada
Department of Physics, University of Strathclyde, Glasgow, UK

N. Hatano
Institute of Industrial Science, University of Tokyo, Tokyo 153-8505, Japan
e-mail: hatano@iis.u-tokyo.ac.jp

statistical mechanics [20]. Then, it is not rare that concepts arising in one discipline are rediscovered and used in another with success. For instance, the concept of node centrality [18, 32], which arises in the study of social networks, is now widely used in the analysis of biological, ecological, and infrastructural networks [5, 7, 9, 11, 13, 21, 22]. Another example is given by the Wiener index, which was introduced in 1947 [34] and defined as the sum of the distances of all shortest paths in the graph representing hydrocarbon molecules. This index has proved to be useful in describing the boiling points and other physico-chemical properties of organic molecules [10]. The mean Wiener index is nowadays known as the average shortest-path distance and it has been instrumental in the definition of the concept of 'small-world' networks [33]. Here we are interested in analysing three concepts arising from different scientific disciplines, in the new context of complex networks. The first of these concepts is the resistance distance introduced in mathematical chemistry by Klein and Randić in 1993 [25] on the basis of electrical network theory. The resistance distance is defined as the effective resistance between two nodes in a graph when a battery is connected across them and the links are considered as unit resistors. The second concept is the information centrality developed by Stephenson and Zelen in 1989 [29], which tries to capture the information that can be transmitted between any two points in a connected network. The third, seemingly unrelated concept is the one of physical vibrations in a network [14, 15]. We consider the displacement of every node in a network due to vibrations/oscillations as a measure of the perturbations that are caused by external factors such as social agitation, economic crisis and physiological conditions. The main objective of this work is to show that these three seemingly unrelated concepts are mathematically connected. Then, we can consider the physically appealing concept of the node vibration as a fundamental concept for complex networks, which is useful in defining: (i) a topological metric, e.g. the resistance distance; (ii) a node centrality, e.g. the information centrality; and (iii) a measure of node vulnerability.

2.2 Resistance Distance in Networks

Let us associate a connected network with an electrical network in such a way that we replace each link of the network with a resistor of electrical resistance equal to one ohm. Then we can calculate the resistance Ω_{ij} between any pair of nodes i and j in the network by the Kirchhoff and Ohm laws. Such resistance is known to be a distance function [25] and called the resistance distance. It was introduced in a seminal paper by Klein and Randić a few years ago [25] and has been intensively studied in mathematical chemistry [14, 15, 25, 29, 35]. The Moore–Penrose generalised inverse (or the pseudo-inverse) $\mathbf{L}^+$ of the graph Laplacian $\mathbf{L}$, which has been proved to exist for any connected graph, gives the following formula [15, 25, 29] for computing the resistance distance:

$$\Omega_{ij} = \left(\mathbf{L}^+\right)_{ii} + \left(\mathbf{L}^+\right)_{jj} - \left(\mathbf{L}^+\right)_{ij} - \left(\mathbf{L}^+\right)_{ji} \tag{2.1}$$

Fig. 2.1 Simple graph used
for illustration of the concepts
of the shortest path and the
resistance distance
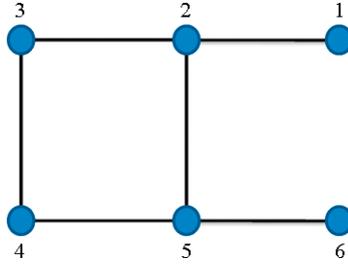
for $i \neq j$, where $\mathbf{L} = \mathbf{D} - \mathbf{A}$ with $\mathbf{D}$ the diagonal matrix of degrees k_i and $\mathbf{A}$ the adjacency matrix of the network.

Let $\mathbf{L}(i)$ be the matrix resulting from removing the ith row and column of the Laplacian and let $\mathbf{L}(i, j)$ the matrix resulting from removing both the ith and jth rows and columns of $\mathbf{L}$. Then, it has been proved that the resistance distance can be also calculated as

$$\Omega_{ij} = \frac{\det \mathbf{L}(i, j)}{\det \mathbf{L}(i)}. \tag{2.2}$$

The resistance matrix $\boldsymbol{\Omega}$ is the matrix whose non-diagonal elements are the resistance distance Ω_{ij} with the diagonal elements $\Omega_{ii} = 0$ [25]. In Fig. 2.1, we illustrate a simple graph having six nodes whose resistance distance matrix is given below. For the sake of comparison, we also give the topological-distance matrix in which the element d_{ij} is given by the number of links in the shortest path between the nodes i and j.

$$\boldsymbol{\Omega} = \begin{bmatrix} 0.00 & 1.00 & 1.75 & 2.00 & 1.75 & 2.75 \\ & 0.00 & 0.75 & 1.00 & 0.75 & 1.75 \\ & & 0.00 & 0.75 & 1.00 & 2.00 \\ & & & 0.00 & 0.75 & 1.75 \\ & & & & 0.00 & 1.00 \\ & & & & & 0.00 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 0 & 1 & 2 & 3 & 2 & 3 \\ & 0 & 1 & 2 & 1 & 2 \\ & & 0 & 1 & 2 & 3 \\ & & & 0 & 1 & 2 \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix}.$$

It is straightforward to realise that for networks that contain no cycles, i.e. trees, both matrices coincide. However, the presence of cycles reduces the resistance distance in comparison with the topological distance. The semi-sum of all entries of the $\mathbf{D}$ matrix is known as the Wiener index $W(G)$. Then, the average path length $\bar{l}$ is given by $\bar{l} = 2W(G)/n(n-1)$. The analogue of the Wiener index in the context of the resistance distance matrix is known as the Kirchhoff index Kf and is defined as $Kf = \sum_{i<j} \Omega_{ij}$ [6, 25, 35, 36, 38]. It is known that Kf can be expressed in terms of the Laplacian eigenvalues as follows [35]:

$$Kf = n \sum_{j=2}^{n} \frac{1}{\lambda_j} = n \operatorname{Tr} \mathbf{L}^+. \tag{2.3}$$

The Wiener and Kirchhoff indices for the graph illustrated in Fig. 2.1 are 20.75 and 27, respectively.

2.3 Information Centrality

The *information centrality (IC)* was introduced by Stephenson and Zelen [29] as a measure of node centrality of social networks. It is based on information that can be transmitted between any two points in a connected network. The motivation for this measure comes from the theory of statistical estimation. Here a path connecting two nodes is considered as a "signal", while the "noise" in the transmission of the signal is measured by the variance of this signal. The information measure I_{ij} between two nodes is defined as the reciprocal of the topological distance d_{ij} between the corresponding nodes, $I_{ij} = 1/d_{ij}$. Stephenson and Zelen [29] proposed to define I_{ii} as infinite for computational purposes, which makes $1/I_{ii} = 0$. The information centrality of the node i is then defined by using the harmonic average:

$$IC(i) = \left[\frac{1}{n} \sum_j \frac{1}{I_{ij}} \right]^{-1}. \tag{2.4}$$

If $\mathbf{A}$ is the adjacency matrix of a network, $\mathbf{D}$ a diagonal matrix of the degree of each node and $\mathbf{J}$ a matrix with all its elements equal to one, then IC is defined by inverting the matrix $\mathbf{B} \equiv \mathbf{D} - \mathbf{A} + \mathbf{J} = \mathbf{L} + \mathbf{J}$, from which the information matrix is obtained as follows:

$$I_{ij}^{-1} = \left(\mathbf{B}^{-1} \right)_{ii} + \left(\mathbf{B}^{-1} \right)_{jj} - 2 \left(\mathbf{B}^{-1} \right)_{ij}. \tag{2.5}$$

The information centrality for the nodes of the graph illustrated in Fig. 2.1 is $IC(1) = IC(6) = 0.649$, $IC(2) = IC(5) = 1.143$ and $IC3 = IC(4) = 0.960$.

2.4 Vibrations in Complex Networks

We now introduce a recently proposed measure of node vulnerability, namely the node displacement [14, 15]. For the purpose, we regard the nodes of the complex network as balls of a common mass and the links as springs of a common spring constant k. We immerse this system of balls and springs in a thermal bath of inverse temperature β and observe the amplitude of thermal fluctuation of each ball. The thermal bath simulates an external stress to the network, such as economical crisis, social agitation, environmental pressure or physiological conditions. The amplitude of thermal fluctuation of a ball tells us how vulnerable the corresponding node is to such stresses.

The vibrational potential energy of the network can be expressed as

$$V(\mathbf{x}) = \frac{k}{2} \mathbf{x}^T \mathbf{L} \mathbf{x}, \tag{2.6}$$

where the ith component x_i of the vector $\mathbf{x}$ denotes the displacement of the node i from its static position due to thermal fluctuation and $\mathbf{L}$ is the same graph Laplacian as used in (2.1). The probability distribution of the displacement of the nodes may be given by the Boltzmann distribution according to the potential energy:

$$P(\mathbf{x}) = \frac{e^{-\beta V(\mathbf{x})}}{Z} = \frac{1}{Z} \exp\left(-\frac{\beta k}{2} \mathbf{x}^T \mathbf{L} \mathbf{x} \right), \tag{2.7}$$

where Z is the partition function of the network:

$$Z \equiv \int d\mathbf{x} \exp\left(-\frac{\beta k}{2} \mathbf{x}^T \mathbf{L} \mathbf{x}\right). \qquad (2.8)$$

The mean square displacement of a node i is given by

$$(\Delta x_i)^2 \equiv \langle x_i^2 \rangle = \int x_i^2 P(\mathbf{x}) \, d\mathbf{x} \qquad (2.9)$$

and the correlation between the displacements of nodes i and j is given by

$$\langle x_i x \rangle_j = \int x_i x_j P(\mathbf{x}) \, d\mathbf{x}, \qquad (2.10)$$

where $\langle \cdots \rangle$ denotes the average with respect to $P(\mathbf{x})$.

We can calculate these quantities by diagonalising the graph Laplacian $\mathbf{L}$. Here we should take care of the fact that the Laplacian of a connected network has a spectrum of the form $0 = \lambda_1 \leq \cdots \leq \lambda_n$; i.e. it has one zero eigenvalue apart from positive eigenvalues. In fact, we should not let the zero eigenvalue contribute in the calculation because the mode $\mu = 1$ represents the motion of the centre of mass and hence its vibrational energy is zero; see [14, 15] for details of the calculation. Here we simply list the results of the calculation. We can represent the results in the following unified form [14, 15]:

$$\langle x_i x \rangle_j = \sum_{\mu=2}^{n} \frac{(\boldsymbol{\psi}_\mu)_i (\boldsymbol{\psi}_\mu)_j}{\beta k \lambda_v} = \frac{1}{\beta k} (\mathbf{L}^+)_{ij}, \qquad (2.11)$$

where $\boldsymbol{\psi}_\mu$ is the eigenvector of the mode μ and $\mathbf{L}^+$ is again the Moore–Penrose generalised inverse of the graph Laplacian [35]. The case $i = j$ gives the mean square displacement $(\Delta x_i)^2 \equiv \langle x_i^2 \rangle$ in (2.11). This quantity is obviously related to the resistance distance defined by (2.11), which we will elucidate in the next section. Meanwhile, (2.11) is followed by the thermal average of the vibrational potential energy (2.6) in the form

$$\langle V(\mathbf{x}) \rangle = \frac{1}{2} \sum_{i=1}^{n} k_i \langle x_i^2 \rangle - \sum_{i,j \in E} \langle x_i x_j \rangle = \frac{1}{\beta k} \sum_{i=1}^{n} k_i (\mathbf{L}^+)_{ii} - \sum_{i,j \in E} (\mathbf{L}^+)_{ij}. \quad (2.12)$$

2.5 Node Displacements and Resistance Distance

Hereafter we set $\beta k \equiv 1$ for simplicity. By using (2.11) in (2.1), we have

$$\Omega_{ij} = \lfloor \langle x_i^2 \rangle + \langle x_j^2 \rangle - \langle x_i x_j \rangle - \langle x_j x_i \rangle \rfloor = \langle (x_i - x_j)^2 \rangle. \qquad (2.13)$$

Roughly speaking, the right-hand side of (2.13) is small if the nodes i and j vibrate coherently in the same direction and large if they move in the opposite directions.

More rigorously, let us focus on the mode $\mu = 2$ of the graph Laplacian $\mathbf{L}$. Then the corresponding eigenvector $\boldsymbol{\psi}_2$ is called the Fiedler vector [17]. This vector is

known to define a partitioning of the graph [17] in the following way. The nodes of a graph are partitioned into two sets $V_1 = \{i|(\boldsymbol{\psi}_2)_i < 0\}$ and $V_2 = \{i|(\boldsymbol{\psi}_2)_i \geq 0\}$. Therefore, two nodes in the same partition according to the Fiedler vector vibrate in the same direction and ones in different partitions vibrate in the opposite directions, when we restrict ourselves to the mode $\mu = 2$. Then, (2.13) gives us a plausible observation that two nodes i and j that are close in terms of the resistance distance Ω_{ij} tend to be in the same partition of the Fiedler vector, whereas ones that are far tend to be in different partitions.

We can also express the Kirchhoff index defined by (2.3) in terms of the node displacements as

$$Kf = n \sum_{i=1}^{n} (\Delta x_i)^2 = n^2 \overline{(\Delta x)^2}, \tag{2.14}$$

where the bar on the right-hand side denotes the average over the nodes. Equation (2.14) tells us that the Kirchhoff index of a molecular graph is proportional to the sum of the squared atomic displacements due to molecular vibrations. Since the Kirchhoff and Wiener indices are known to coincide for acyclic networks, i.e. trees, we also have

$$W(T) = n \sum_{i=1}^{n} (\Delta x_i)^2 = n^2 \overline{(\Delta x)^2}. \tag{2.15}$$

We now consider the average potential energy in (2.12). For this purpose, let us calculate the quantity

$$R_i = \sum_{j=1}^{n} \Omega_{ij}, \tag{2.16}$$

the sum of all resistance distances from atom i to any atoms in the molecule. By combining expression (2.1) with the general fact $\sum_{j=1}^{n} (\mathbf{L}^+)_{ij} = 0$, we have

$$R_i = n(\mathbf{L}^+)_{ii} + \mathrm{Tr}\,\mathbf{L}^+ = n(\Delta x_i)^2 + n\overline{(\Delta x)^2} = n(\Delta x_i)^2 + \frac{Kf}{n}. \tag{2.17}$$

This shows that $(\Delta x_i)^2$ and R_i are linearly related for all nodes of a given network. Using then (2.13), we also have

$$\Omega_{ij} = \frac{R_i + R_j}{n} - 2\frac{Kf}{n^2} - 2\langle x_i x_j \rangle. \tag{2.18}$$

The average potential energy is then given by

$$\langle V(\mathbf{x}) \rangle = \frac{1}{2n} \sum_{i=1}^{n} k_i R_i - \frac{1}{2n} \sum_{i,j \in E} (R_i + R_j - n\Omega_{ij}). \tag{2.19}$$

The first term on the right-hand side of (2.19) was first introduced by Estrada et al. [16] as a topological index for trees obtained from the quadratic form $\langle \mathbf{v}|\mathbf{D}|\mathbf{u} \rangle$, where $\mathbf{v}$ is a vector of node degrees, $\mathbf{D}$ is the distance matrix and $\mathbf{u}$ is a vector of ones of length equal to the number of nodes in the graph.

2.6 Node Displacement and Information Centrality

In the present section, we explore the relation between the node displacement and the information centrality (2.4). Let us first prove that the inverse of $\mathbf{B} = \mathbf{L} + \mathbf{J}$ exists and is given by

$$\mathbf{B}^{-1} = \mathbf{L}^+ + \frac{1}{n^2}\mathbf{J}. \tag{2.20}$$

Let $\boldsymbol{\psi}_\mu$ denote the μth eigenvector of the graph Laplacian $\mathbf{L}$, which has the spectrum $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$ for a connected network. Note here that $\boldsymbol{\psi}_1 = \frac{1}{\sqrt{n}}\mathbf{1}$. For $\mu \neq 1$, we have

$$\mathbf{B}\boldsymbol{\psi}_\mu = \mathbf{L}\boldsymbol{\psi}_\mu + \mathbf{J}\boldsymbol{\psi}_\mu = \mathbf{L}\boldsymbol{\psi}_\mu = \lambda_\mu\boldsymbol{\psi}_\mu \tag{2.21}$$

because

$$(\mathbf{J}\boldsymbol{\psi}_\mu)_j = \sum_{i=1}^n (\boldsymbol{\psi}_\mu)_i = \sqrt{n}\boldsymbol{\psi}_1 \cdot \boldsymbol{\psi}_\mu = 0 \quad \text{for } \mu \neq 1.$$

For $\mu = 1$, we have

$$\mathbf{B}\boldsymbol{\psi}_1 = \mathbf{L}\boldsymbol{\psi}_1 + \mathbf{J}\boldsymbol{\psi}_1 = n\boldsymbol{\psi}_1. \tag{2.22}$$

The above means that the eigenvalues of the matrix $\mathbf{B}$ are $n, \lambda_2, \lambda_3, \ldots, \lambda_n$, which are all positive. The matrix $\mathbf{B}$ is thereby invertible. Indeed, we can confirm (2.20) as

$$(\mathbf{L} + \mathbf{J})\left(\mathbf{L}^+ + \frac{1}{n^2}\mathbf{J}\right) = \mathbf{I} - \frac{1}{n}\mathbf{J} + \frac{n\mathbf{J}}{n^2} = \mathbf{I}, \tag{2.23}$$

because $\mathbf{L}\mathbf{L}^+ = \mathbf{L}^+\mathbf{L} = \mathbf{I} - \frac{1}{n}\mathbf{J}$, $\mathbf{L}\mathbf{J} = \mathbf{J}\mathbf{L} = \mathbf{L}^+\mathbf{J} = \mathbf{J}\mathbf{L}^+ = 0$, and $\mathbf{J}^2 = n\mathbf{J}$. This proves that the matrix in (2.20) is the inverse of $\mathbf{B}$.

Equation (2.20) then transforms (2.5) into the form

$$I_{ij}^{-1} = \left(\mathbf{B}^{-1}\right)_{ii} + \left(\mathbf{B}^{-1}\right)_{jj} - 2\left(\mathbf{B}^{-1}\right)_{ij} = \left(\mathbf{L}^+\right)_{ii} + \left(\mathbf{L}^+\right)_{jj} - 2\left(\mathbf{L}^+\right)_{ij} = \Omega_{ij}. \tag{2.24}$$

Therefore, the information centrality (2.4) is now given by

$$IC(i) = \left(\frac{1}{n}\sum_j \frac{1}{I_{ij}}\right)^{-1} = \left(\frac{1}{n}\sum_j \Omega_{ij}\right)^{-1} = \frac{n}{R_i} = \left((\Delta x_i)^2 + \overline{(\Delta x)^2}\right)^{-1}. \tag{2.25}$$

2.7 Node Displacement as a Measure of Node Vulnerability

Most of the studies on vulnerability of complex networks consider how resilient the whole network is to random failures and intentional attacks. In these studies, it is assumed that we can attack any node by simply removing it from the graph. The primary removal of these nodes can give rise to the secondary disconnection of

other nodes from the main connected component of the network. The most resilient network is the one that, after many removals, still keeps the functioning size of the main connected component.

Here we are interested in the vulnerability of a node rather than the vulnerability of the whole network to targeted attacks. Intuitively, a node is highly vulnerable if there are many other nodes whose individual removal disconnects the node in question from the main connected component of the network. The information centrality $IC(i)$ can be an index of the node vulnerability; we may be able to say that a node i with a larger information centrality is less vulnerable. The resistance distance can provide an equivalent index; as was shown in the previous section, the quantity R_i is inversely proportional to the information centrality. Hereafter, we will suggest that the node displacement Δx_i can be another index of the vulnerability and actually has a better resolution than the information centrality.

Let us first explain in terms of the Fiedler vector why the node displacement can measure the node vulnerability. Recall that we have ordered the eigenvalues of the Laplacian as $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$. The eigenvector of the second mode (the first non-zero mode) is the Fiedler vector [17]. Let us consider the particular case $\lambda_2 < \lambda_3$. Then (2.11) implies that the term $(\boldsymbol{\psi}_2)_i^2/\lambda_2$ of the Fiedler vector has the largest contribution to Δx_i. Among the nodes, a node with $(\boldsymbol{\psi}_2)_i$ close to zero does not strongly belong to either of the two partitions V_1 and V_2 defined by the Fiedler vector. Such nodes are located in between the partitions; in other words, they tend to have ties with many other nodes and hence may be less vulnerable to external stresses. Then a small value of $(\boldsymbol{\psi}_2)_i^2/\lambda_2$ can indeed indicate little node vulnerability.

In the seminal paper by Albert, Jeong and Barabási [2], they chose the nodes with the highest degree for their targets of the primary removals. This is based on an empirical observation that the nodes with the lowest degree are the most vulnerable. Consider the case where a node i has only one connection, i.e. $K_i = 1$. Then, we can isolate it from the network by removing the node to which i is connected. A measure of the node vulnerability in complex networks should be consistent in some way with the above observation that low-degree nodes are more vulnerable than high-degree ones. As we discussed above, the term $(\boldsymbol{\psi}_2)_i^2/\lambda_2$ indeed has such a property.

In fact, the node displacement (2.11) takes account of the higher modes, too. As the Fiedler vector defines a bipartition of the network, the eigenvectors of higher modes can define partitions into a larger number of groups. These partitions may identify the clusters and the nodes in between them in a more appropriate way. The use of the eigenvectors of higher modes also helps avoiding the problem that can arise when $\lambda_2 = \lambda_3$. This degeneracy can happen in square grids and complete graphs, for example. In this degenerate case, it has been reported that the convergence of partitioning algorithms can be poor.

Let us demonstrate that the node displacement can be indeed a measure of the node vulnerability in the sense that it tends to give higher vulnerability to low-degree nodes than to high-degree nodes. For the purpose, we use the trade network of miscellaneous manufactures of metal (MMM) among 80 countries in 1994. The data was compiled by de Nooy [8] and the reader is refereed to this work to obtain
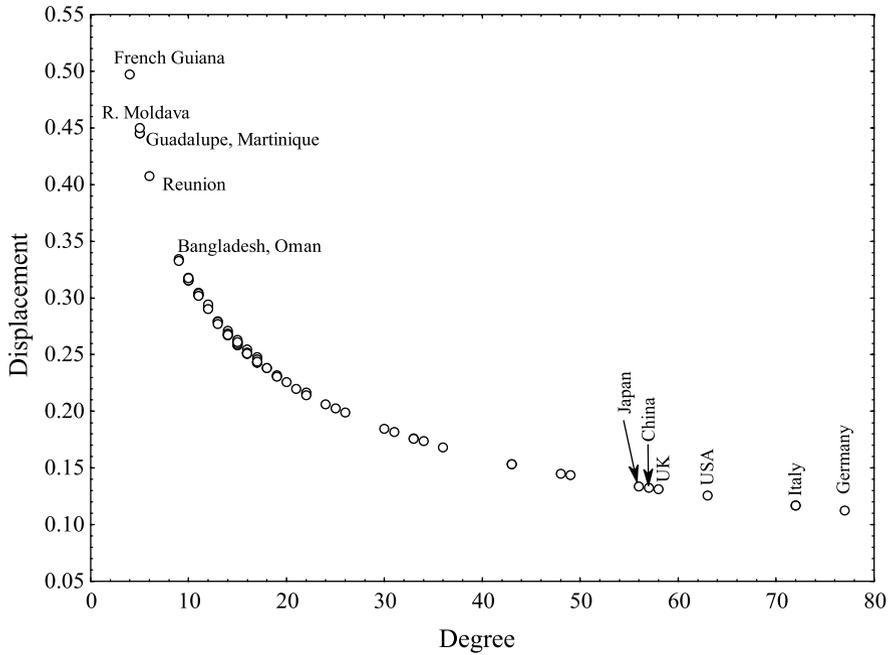
Fig. 2.2 Illustration of the relationship between the node displacement and the node degree for the trade network of miscellaneous manufactures of metal among 80 countries in 1994

the details of this dataset. We consider here only the undirected and unweighted version of this network. Here, the nodes represent the countries and a link exists between two countries if one of them imports MMM from the other.

We show in Fig. 2.2 the relation between the node degree and the node displacement for the above trade network. The node displacement decreases with the node degree under a power law. Interestingly, the countries with the largest node displacements are mostly the poorest countries in the network, whereas the richest countries are among the ones with the smallest node displacements. Choosing the node displacement as a measure of the node vulnerability, we can say that the poorest countries are the most vulnerable to changes in economical situations such as a financial crisis.

Having said this, we emphasise here that the node degree accounts only for the nearest neighbours of a node. In other words, the influence of more distant nodes is not taken into account if we use the node degree as a measure of the node vulnerability. This can be seen in the fact that many networks have several nodes with the same degree but with different values of the node displacement.

For an illustrative example, let us study two networks of sexual contacts collected by Lind et al. [26]. One of them is composed solely of heterosexual contacts among 82 people, which was extracted from the Cadham Provincial Laboratory during a period of six months. The other is formed by sexual contacts (mainly homosexual)

among 250 individuals collected from an HIV test study in Colorado Springs (USA). Note that the first network is bipartite but the second one is not. In Fig. 2.3, we show the correlations between the degree and the displacements for the nodes of these two networks. We can see that there are many nodes with the same degree that display a large variability of their node displacements. This demonstrates the fact that node vulnerability is a different characterisation of node vulnerability than the one provided by the node degree. In fact, node displacement takes into account a more global picture of the environment of a node than the node degree.

2.8 Topological Displacements in Protein Residue Networks

We next describe the application of the node displacement to molecular networks. We can represent proteins as complex networks by using information on their three-dimensional structures. One example of these representations is the residue network. The nodes of a spatial residue network correspond to the amino acid residues of the protein; as the spatial location of the residue, we use the coordinate of its β-carbon except that we use the α-carbon for glycine. We then determine the links of the residue network in terms of the spatial distance between the two residues; two nodes are connected if the spatial locations of the two residues are closer than a cutoff radius r_C [3]. In other words, we define the elements of the adjacency matrix of the residue network as

$$A_{ij} = \begin{cases} \Theta(r_C - r_{ij}) & \text{for } i \neq j, \\ 0 & \text{for } i = j, \end{cases}$$

where $\Theta(x > 0) = 1$ and $\Theta(x \leq 0) = 0$. We can thus represent a protein as a graph $G = (V, E)$, where V is the set of the amino acid residues and E is the set of the connections between them. The residue network of the protein with PDB code 1ash, for example, is shown in Fig. 2.4 [3, 12].

It is then natural to suppose that the node displacement Δx_i calculated for a node of the residue network displays linear correlation with an experimental measure of how much a residue oscillates or vibrates around its equilibrium position. One such experimental measure is the B-factor, or the temperature factor provided by X-ray experiments. It represents the reduction of coherent scattering of X-rays due to thermal motion of the atoms.

The B-factors are important for the study of protein structures as they contain valuable information on the dynamical behaviour of proteins. Several methods have been designed for the prediction of the B-factors [28]. Regions with large B-factors are known to be flexible and functionally important. Bahar et al. have used the atomic displacements to describe thermal fluctuations in proteins [4]. Note that we use here Bahar et al.'s representation of a residue network in the sense that we use the β-carbons instead of the α-carbon for the spatial locations of the amino acids.

In Fig. 2.5, we show the profiles of the normalised B-factors and the node displacements of the residue networks for the spinach ferredoxin reductase (top) at 1.7 angstroms resolution (1fnc) and for the human uracil-DNA glycosylase (1akz) at
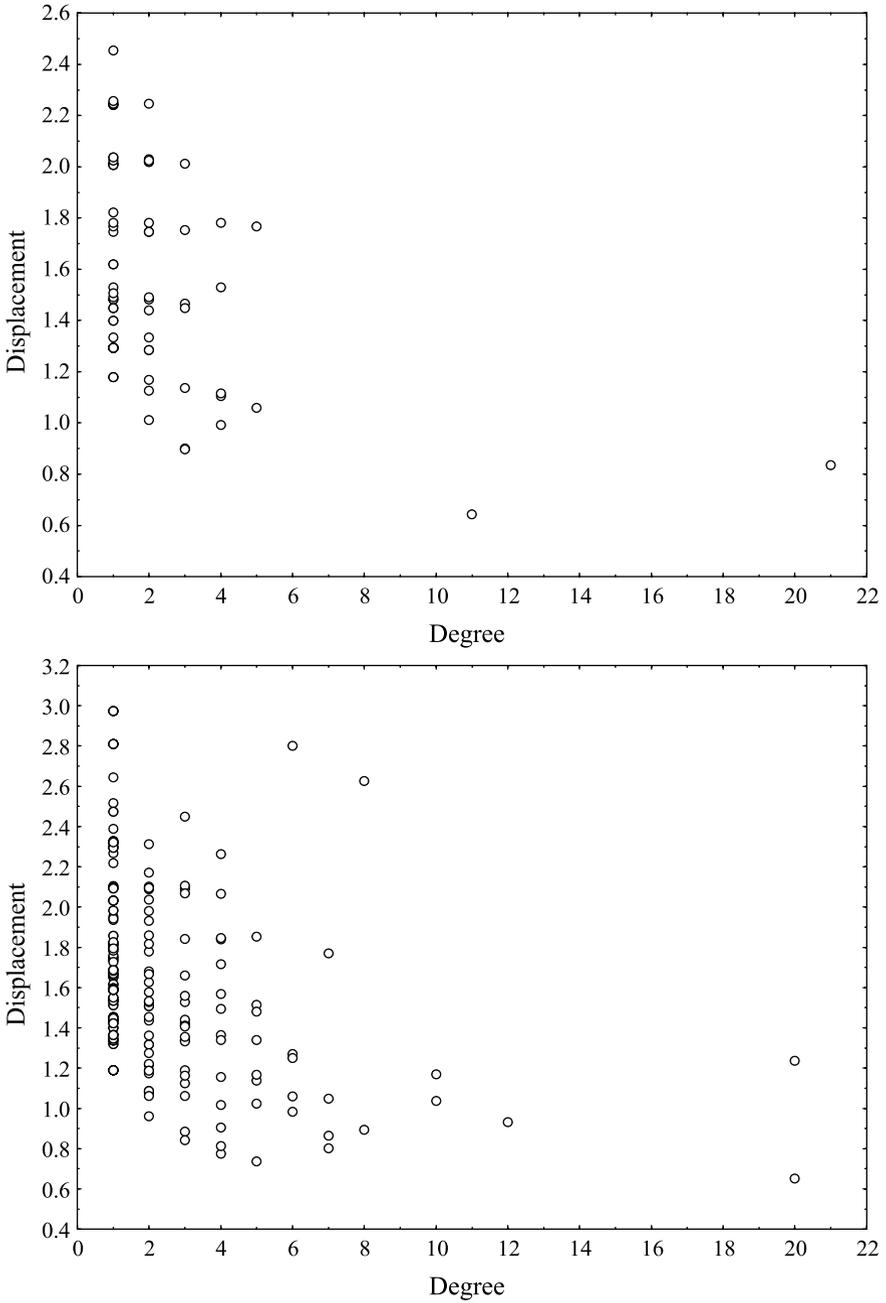
Fig. 2.3 Illustration of the relationship between the node displacement and the node degree for two networks of sexual contacts. (*Top*) A network formed solely by heterosexual contacts. (*Bottom*) Network of sexual contacts (mainly homosexual) among 250 individuals in Colorado Springs (USA)
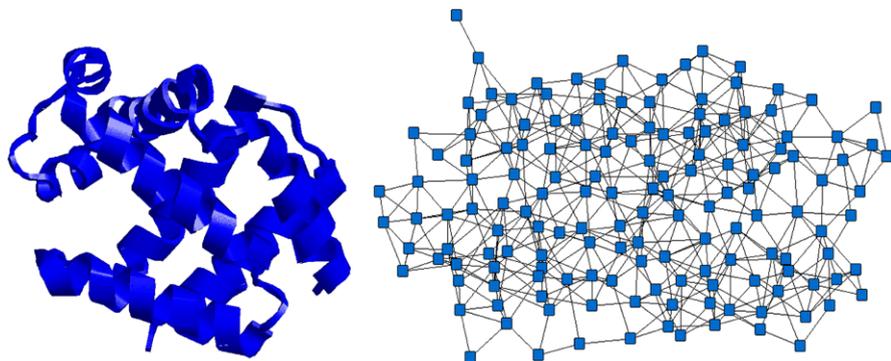
Fig. 2.4 3D representation of the structure of a protein (*left*) and the network representation of physical contacts between amino acids represented by their β-carbons (*right*)

1.57 angstroms (bottom). In both cases, the experimental profiles of the B-factors are very well reproduced by the node displacements of the β-carbons in the protein residue networks. The correlation coefficients between the B-factors and the node displacements are $r = 0.56$ and $r = 0.65$ for the proteins 1fnc and 1akz, respectively. For them, Yuan et al. [37] obtained $r = 0.48$ and $r = 0.72$, respectively, by a statistical method based on support vector regression. In short, the node displacements of a residue network are correlated with the B-factors obtained for the residue itself by X-ray crystallography in a similar way to other methods currently in use for this purpose.

2.9 Node Displacements for Temporal Change on Networks

Another interesting application of the node displacement is the analysis of the temporal change of a network. We can compare the node displacements of an evolving network at different times. For demonstration, here we use a dataset obtained by Kapferer for the social ties among 39 tailor shops in Zambia [23]. The friendship and socio-emotional relationship among the 39 tailors were under observation during a period of ten months. The dataset consists of two phases of the network recorded with an interval of seven months [23]; see http://vlado.fmf.uni-lj.si/pub/networks/data/Ucinet/UciData.htm.

After the first dataset was collected an abortive strike was reported [23]. After the collection of the second dataset, a successful strike took place.

We calculated the change in the node displacement between the two phases of the network. Let $\Delta x_i(t)$ denote the node displacement of the node i at t, where $t = 1, 2$. We then define the difference $\Delta \Delta x_i = \Delta x_i(2) - \Delta x_i(1)$. For comparison, we also calculated in a similar way the difference $\Delta IC(i)$ between the information centralities of the node i at the two phases.

In order to analyse the differences in the ranking of nodes in terms of $\Delta \Delta x_i$ and $\Delta IC(i)$, we use a nonparametric measure of correlation known as the *Kendall*
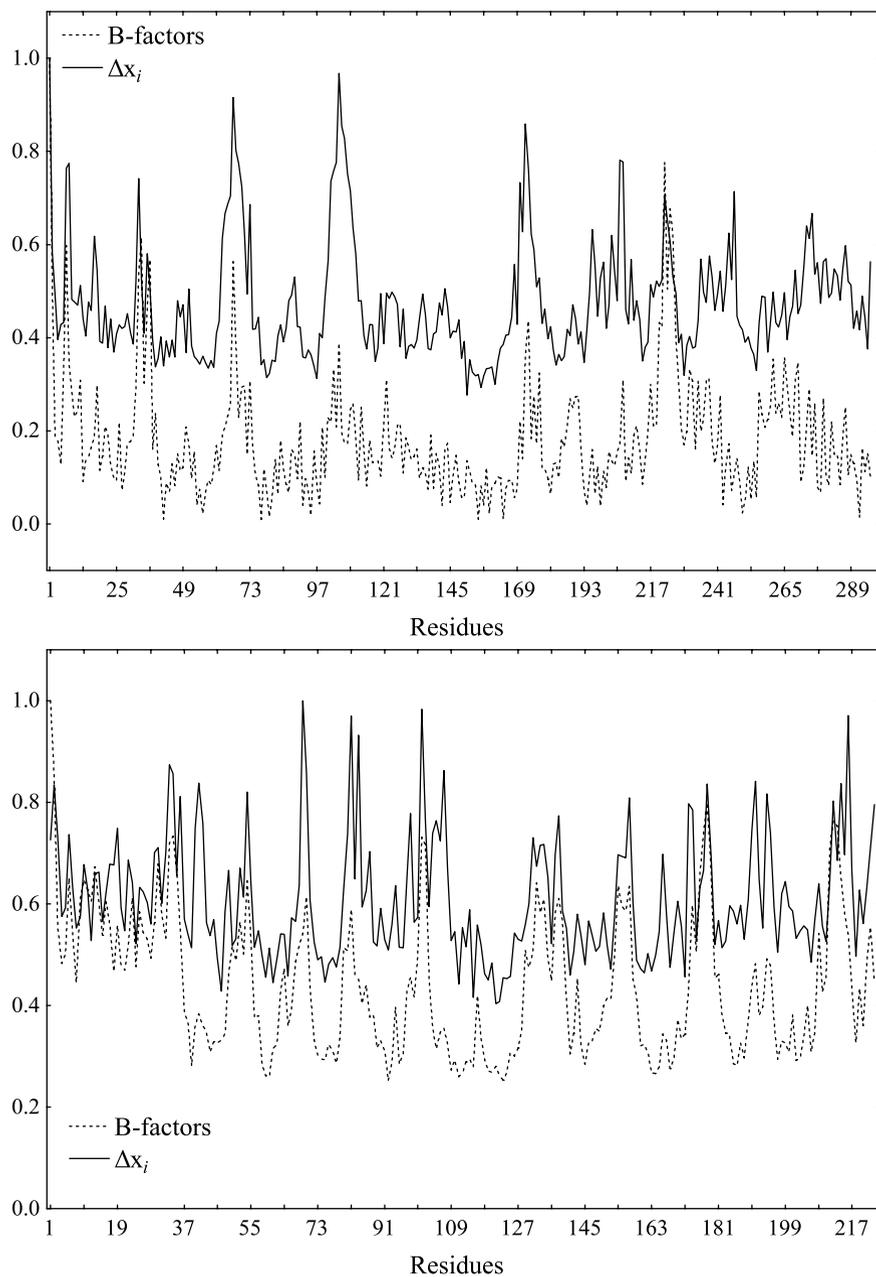
Fig. 2.5 Profiles of the experimental *B*-factors and the node displacements for the residues of the spinach ferredoxin reductase, PDB: 1fnc, (*top*) and for the human uracil-DNA glycosylase, PDB: 1akz, (*bottom*) represented by their residue networks
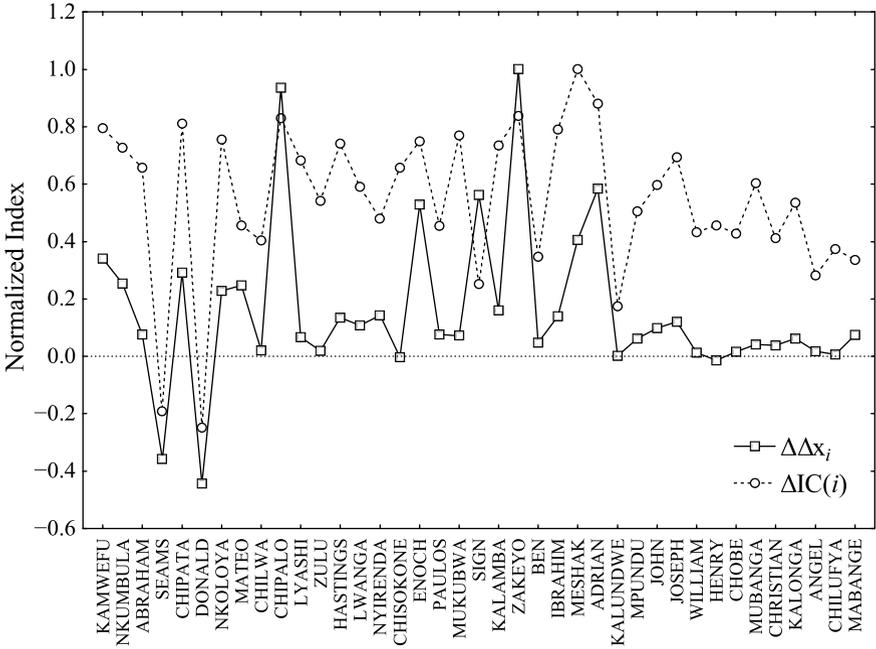
Fig. 2.6 Profiles of the normalised differences of the nodes displacements and the information centralities of the nodes in the tailor shop network observed at two different times

τ statistic [24]. This statistic represents the difference between the probability that the two datasets are in the same order and the probability that they are in different orders. Let p_c and p_d be the number of the concordant and discordant pairs of the data points, respectively, such that $p = p_c + p_d$. Then the Kendall τ index is defined as [24]

$$\tau = \frac{2(p_c - p_d)}{p(p-1)}. \tag{2.26}$$

The nonparametric correlation between $\Delta\Delta x_i$ and $\Delta IC(i)$ is $\tau = -0.56$, which indicates that both indices rank very differently and almost in a completely uncorrelated way. For instance, the ranking of top individuals in terms of $\Delta\Delta x_i$ is as follows: Zakeyo > Chipalo > Adrian > Sign > Enoch > Donald > Meshak > Seans > Kamwefu > Chipata. On the other hand, the ranking in terms of $\Delta IC(i)$ is as follows: Meshak > Adrian > Zakeyo > Chipalo > Chipata > Kamwefu > Ibrahim > Mukubwa > Nkoloya > Enoch. In Fig. 2.6, we show the profiles of the normalised values of $\Delta\Delta x_i$ and $\Delta IC(i)$ for all individuals in this network.

This striking difference between the information centrality and the node displacement can be traced back to (2.25). It tells us that an increase in the information centrality of a node can have two origins: a decrease in the local node displacement $(\Delta x_i)^2$ or a decrease in the global (average) node displacement $\overline{(\Delta x)^2}$ (or both).

More detailed analysis of the two rankings of the nodes tells us that for most nodes, the increase of the information centrality is caused by the decrease of the global node displacement, not by the local one. Many tailors had many social ties and were not very vulnerable in the first place. Their information centrality increased after the seven months mostly because the number of the links generally increased all over the network; on the other hand, their local node displacement (the vulnerability of each tailor) scarcely changed. For tailors such as Zakeyo and Chipalo, however, their local node displacements decreased greatly after the seven months because their own degrees increased dramatically. This demonstrates clearly that the node displacement has a better resolution of the time evolution than the information centrality.

In summary, despite the relation (2.25) between the information centrality and the node displacement, there is a fundamental difference between them. The information centrality can be seen as a composite index containing local information of a node as well as global topological information of the network, whereas with the node displacement we can separate it into the local information as $(\Delta x_i)^2$ and the global one as $\overline{(\Delta x)^2}$. This difference is very relevant when comparing nodes in different networks.

2.10 Outlook

It is a well known fact that there are several common features between very disparate complex systems arising in non-related areas of nature, society or technology. When these systems are represented by complex networks, some of these features are well documented in the scientific literature [1, 27, 30]. In order to discover these universal features, we need to carry out cross-comparative analysis of complex systems and their behaviours by using appropriate mathematical tools and physical concepts. Here we have introduced the concept of the node displacement as a measure of vulnerability of each node in a network. It is defined in terms of the amplitude of vibration caused by thermal fluctuation of a heat bath in which the network is immersed. This physical analogy simulates the situation in which the network in question is under a level of external stress. It is interesting that this fundamental physical concept is related to graph-theoretic invariants previously developed and used in very different scientific disciplines like Chemistry and Social Sciences. In this sense, we have seen in practise the unifying nature of physico-mathematical concepts across the boundaries of many disciplines. Then, we hope that this work contributes to the interdisciplinary search of more universal properties of complex systems that permit a better understanding of their structure and dynamics.

References

1. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Rev. Mod. Phys. **74**, 47–97 (2002)

2. Albert, R., Jeong, H., Barabási, A.L.: Error and attack tolerance of complex networks. Nature **406**, 378–382 (2000)
3. Atilgan, A.R., Akan, P., Baysal, C.: Small-world communication of residues and significance for protein dynamics. Biophys. J. **86**, 85–91 (2004)
4. Bahar, I., Atilgan, A.R., Erman, B.: Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold. Des. **2**, 173–181 (1997)
5. Barthelemy, M., Barrat, A., Vespignani, A.: The role of geography and traffic in the structure of complex networks. Adv. Complex Syst. **10**, 5–28 (2007)
6. Chen, H.Y., Zhang, F.J.: Resistance distance and the normalized Laplacian spectrum. Discrete Appl. Math. **155**, 654–661 (2007)
7. Choi, J., Barnett, G., Chou, B.: Comparing world city networks: a network analysis of Internet backbone and air transport intercity linkages. Glob. Netw. **6**, 81–99 (2006)
8. de Nooy, W., Mrvar, A., Batagelj, V.: Exploratory Social Network Analysis with Pajek. Cambridge University Press, Cambridge (2005)
9. del Rio, G., Koschutzki, D., Coello, G.: How to identify essential genes from molecular networks? BMC Syst. Biol. (2009). doi:10.1186/1752-0509-3-102
10. Dobrynin, A.A., Entringer, R., Gutman, I.: Wiener index of trees: Theory and applications. Acta Appl. Math. **66**, 211–249 (2001)
11. Estrada, E.: Virtual identification of essential proteins within the protein interaction network of yeast. Proteomics **6**, 35–40 (2006)
12. Estrada, E.: Universality in protein residue networks. Biophys. J. **98**, 890–900 (2010)
13. Estrada, E., Bodin, O.: Using network centrality measures to manage landscape connectivity. a short path for assessing habitat patch importance. Ecol. Appl. **18**, 1810–1825 (2008)
14. Estrada, E., Hatano, N.: A vibrational approach to node centrality and vulnerability in complex networks. arXiv:0912.4307
15. Estrada, E., Hatano, N.: Topological atomic displacements, Kirchhoff and Wiener indices of molecules. Chem. Phys. Lett. **486**, 166–170 (2010)
16. Estrada, E., Rodríguez, L., Gutiérrez, A.: Matrix algebraic manipulations of molecular graphs. 1. graph theoretical invariants based on distances and adjacency matrices. MATCH Commun. Math. Comput. Chem. **35**, 145–156 (1997)
17. Fiedler, M.: Algebraic connectivity of graphs. Czechoslov. Math. J. **23**, 298–305 (1973)
18. Freeman, L.C.: A set of measures of centrality based upon betweenness. Sociometry **40**, 35–41 (1977)
19. Freeman, L.C.: The Development of Social Network Analysis. Empirical Press, Vancouver (2004)
20. Harary, F. (ed.): Graph Theory and Theoretical Physics. Academic Press, London (1967)
21. Jeong, H., Mason, S.P., Barabási, A.L., Oltvai, Z.N.: Lethality and centrality in protein networks. Nature **411**(6833), 41–42 (2001)
22. Jordán, F.: Keystone species and food webs. Philos. Trans. R. Soc. Lond. B, Biol. Sci. **364**, 1733–1741 (2009)
23. Kapferer, B.: Strategy and Transaction in an African Factory. Manchester University Press, Manchester (1972)
24. Kendall, M.: A new measure of rank correlation. Biometrika **30**, 81–89 (1938)
25. Klein, D., Randić, M.: Resistance distance. J. Math. Chem. **12**, 81–95 (1993)
26. Lind, P., González, M., Herrmann, H.: Cycles and clustering in bipartite networks. Phys. Rev. E **72**, 056127 (2005)
27. Newman, M.E.J.: The structure and function of complex networks. SIAM Rev. **45**(2), 167–256 (2003)
28. Soheilifard, R., Makarov, D., Rodin, G.: Critical evaluation of simple network models of protein dynamics and their comparison with crystallographic B-factors. Phys. Biol. **5**, 1–13 (2008)
29. Stephenson, K., Zelen, M.: Rethinking centrality: methods and examples. Soc. Netw. **11**, 1–37 (1989)
30. Strogatz, S.H.: Exploring complex networks. Nature **410**, 268–276 (2001)

31. Trinajstić, N.: Chemical Graph Theory. CRC Press, Boca Raton (1992)
32. Wasserman, S., Faust, K.: Social Network Analysis. Cambridge University Press, Cambridge (1994)
33. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**, 440–442 (1998)
34. Wiener, H.: Structural determination of paraffin boiling points. J. Am. Chem. Soc. **69**, 17–20 (1947)
35. Xiao, W., Gutman, I.: Resistance distance and Laplacian spectrum. Theor. Chem. Acc. **110**, 284–289 (2003)
36. Yang, Y.J., Zhang, H.P.: Some rules on resistance distance with applications. J. Phys. A, Math. Theor. **41**, 445203 (2008)
37. Yuan, Z., Bailey, T., Teasdale, R.: Prediction of protein B-factor profiles. Proteins **58**, 905–912 (2005)
38. Zhou, B., Trinajstić, N.: On resistance distance and Kirchhoff index. J. Math. Chem. **46**, 283–289 (2009)