The Matrix-Forest Theorem and Measuring Relations in Small Social Groups

Article *in* Automation and Remote Control · March 2006

2 authors, including:

Pavel Chebotarev

Institute of Control Sciences of the Russian Academy of sciences, Moscow, Russia

80 PUBLICATIONS   **1,185** CITATIONS

Some of the authors of this publication are also working on these related projects:

Graph Measures Investigation View project

THE MATRIX-FOREST THEOREM AND MEASURING RELATIONS IN SMALL SOCIAL GROUPS[1]

P. Yu. Chebotarev and E. V. Shamis   UDC 519.172

We propose a family of graph structural indices related to the matrix-forest theorem. The properties of the basic index that expresses the mutual connectivity of two vertices are studied in detail. The derivative indices that measure "dissociation," "solitariness," and "provinciality" of vertices are also considered. A nonstandard metric on the set of vertices is introduced, which is determined by their connectivity. The application of these indices in sociometry is discussed.

1. INTRODUCTION

Given a graph, how should one evaluate the proximity between its vertices? The standard distance function is the length of the shortest path. But is it not worth taking into account the *number* of paths between vertices? Which vertices can be considered central and which peripheral? Which graphs are dense, and which are sparse? Which are homogeneous? The choice of indices that express these and other structural properties of graphs depends on applications, more exactly, on the type of applications. This type should ideally be formulated in terms of axiomatic requirements on the structural indices or via modeling those concepts that should be evaluated by these indices. The applications are numerous; essentially, these are all applications of graph theory: transport, reliability, transmission of information, structural modeling, chemistry, molecular biology, epidemiology, etc.

The application we shall focus on is one of the most difficult to formalize. It is sociology, more precisely, sociometry where structural indices are usually chosen heuristically.

Sociometry studies the structure of small social groups on the basis of given relations on them. As a rule, these relations are binary; in some cases they are weighted. Small social groups are groups where public relations manifest themselves in the form of personal contacts or, simply stated, these groups are natural communities where everyone knows each other. The binary relations under study mainly result from sociometric interrogations. For example, each member of a group is asked to indicate those persons with whom she is in sympathy (or out of sympathy), or with whom she spends her spare time most often, or with whom she *would prefer* to cooperate in certain activities (work, rest, "exploration," etc.), or who, in her opinion, has certain characteristics. If a member i of a group indicates (among others) j, an arc from i to j is drawn in the digraph of the relationship. Nonoriented graphs are frequently included to represent objective information (contacts, collaborations, etc.). If a set of similar questions is asked or the respondents report their assumptions on the opinion of others (autosociometric data), multigraphs or multidigraphs can serve as the model. A similar approach is used in political studies where countries or parties involved in certain relationships are investigated.

A lot of various kinds of relations can be studied, each requiring its own properties of the structural indices, so it is problematic to construct the desirable axiomatics for every application. Another approach seems more realistic: to collect a "library" of structural indices with specified properties, and to use those indices whose features are most appropriate for the relations under study.

Traditional sociological indices are very simple. For instance, the *sociometric status* of the ith member of a group is the normalized in-degree (the number of entering arcs) of the ith vertex; the *psychological effusiveness* is the

---

normalized out-degree; the *reciprocity of the choice of i* is the normalized number of pairs of opposite arcs incident to i [1]. The *density* and the *cohesion* of a group result by averaging the sociometric status and the reciprocity over the group. The *heterogeneity* of a group is measured by the empirical variance of the sociometric status over the group. The imperfection of these elementary indices is caused by their local nature. In this connection, Paniotto [1] adduces examples of essentially dissimilar structures with the same values of the above indices and states that more sensitive indices that capture the topological structure as a whole and the part of each individual in this structure are desirable.

A family of more sensitive indices is based on evaluating the vertex status by the sum of the lengths (or reciprocal lengths) of the shortest paths that lead to this vertex from all other vertices [2]. Note, however, that such characteristics of a group member are frequently unchanged on altering the connections between other members, which may not conform with the interpretation of the model. Furthermore, when characterizing the proximity of two members of a group, it is often worth taking into account not only the length of the shortest path between them, but also the numbers of paths of various lengths.

One more idea employed in the construction of sociometric indices is to evaluate the group cohesion by the number of arcs (edges) in the minimum cutset, i.e., by the minimum number of connections whose removal breaks the connectedness of the corresponding graph. The normalized minimum number of members whose removal (together with their connections) makes the graph disconnected is sometimes called *group stability* (*vitality*). One problem (besides the computational one) with such indices is that the graph can be found to be disconnected from the very beginning. This situation still allows one to study the increase in the number of connected components. On the other hand, even for a connected graph, these indices are solely determined by its "bottlenecks," i.e., such characteristics are indifferent to the existence of joined subgroups with relatively poor connections between them.

In this paper, we propose a family of sensitive structural indices and study its properties. The definitions of the indices are based on the matrix-forest theorem (Section 2). Section 3 is devoted to the properties of the basic index of vertex proximity; Section 4 discusses it and introduces derivative indices.

The basic results of this work were stated in [3]. A topological interpretation of the vertex proximity index (implicitly used in [4]) was obtained in [5, 6]; an interesting further investigation of the matrix of these indices was undertaken in [7]; close ideas with reference to chemistry were developed in [8], where an important analogy with electrical networks was also formulated. In a subsequent paper, we are going to compare the structural indices proposed here with other ones known from the literature (see, for example, [9]).

2. THE MATRIX-FOREST THEOREM

The matrix-forest theorem is formulated for multigraphs and multidigraphs (which differ from graphs and digraphs by the possibility of multiple edges and arcs). A *subgraph* of a multigraph G is a multigraph all of whose vertices and edges belong to the vertex set and the edge set of G. A *spanning subgraph* of a multigraph G is a subgraph of G with the same vertex set as that of G. A *path* in a multigraph G is an alternating sequence of distinct vertices and edges, which starts and ends with vertices and has each edge situated between two vertices incident to it. Sometimes we consider a path as a subgraph of G. A *forest* is a cycleless graph. A *tree* is a connected forest. A *rooted tree* is a tree with one marked vertex, called a *root*. Formally, a rooted tree is a pair (T, r), where T is a tree and r is its vertex. A *component* of a multigraph G is any maximal (by inclusion) connected subgraph of G. Obviously, all components of a forest are trees.

A *rooted forest* is defined as a forest with one marked vertex in each component. A *directed path* in a multidigraph Γ is defined similarly to a path in a multigraph, but here each arc is directed from the previous vertex to the next one in the sequence. A digraph is called a *directed tree* (a *directed forest*) if the graph obtained from it by replacement of all its arcs with edges is a tree (a forest). The definitions of *directed rooted tree* and *directed rooted forest* are analogous to the definitions of rooted tree and rooted forest (we will omit the word "directed" while talking about subgraphs of Γ). A *diverging tree* is a directed rooted tree that contains directed paths from the root to all other vertices. A *diverging forest* is a directed rooted forest, all of whose components are diverging trees.

Suppose that G is a weighted multigraph with vertex set $V(G) = \{1, \ldots, n\}$ and edge set $E(G)$. Let $\varepsilon_{ij}^p \geq 0$ be the *weight* of the pth edge between vertices i and j in G. This weight will be also referred to as the *conductance* of the edge.

The *Kirchhoff matrix* of G is the $n \times n$ matrix $L = L(G) = \left(\ell_{ij}\right)$ with

$$\ell_{ij} = -\sum_{p=1}^{a_{ij}} \varepsilon_{ij}^p, \;\; j \neq i, \;\; i, j = 1, \ldots, n, \tag{1}$$

$$\ell_{ii} = -\sum_{j \neq i} \ell_{ij}, \quad i = 1, \ldots, n, \tag{2}$$

where a_{ij} is the number of edges between i and j. The product of the weights of all edges that belong to a subgraph H of a multigraph G will be referred to as the *weight* or *conductance* of H and denoted by $\varepsilon(H)$. The weight (conductance) of a subgraph without edges is set to be 1. For every nonempty set of subgraphs $\mathcal{G}$, its weight is defined as follows:

$$\varepsilon(\mathcal{G}) = \sum_{H \in \mathcal{G}} \varepsilon(H).$$

The weight of the empty set is zero.

The following matrix-forest lemmas are similar to the classical matrix-tree theorems, obtained by Kirchhoff and some other writers in the nineteenth century (for the history, see [10]). We shall formulate Tutte's generalization of the matrix-tree theorem to weighted multigraphs (see [11]).

Denote by L^{ij} the cofactor of ℓ_{ij} in L. Let $\mathcal{T}(G) = \mathcal{T}$ be the set of all spanning trees of multigraph G.

THEOREM 1 (matrix-tree theorem for weighted multigraphs). *For any weighted multigraph G and for any $i, j \in V(G)$, $L^{ij} = \varepsilon(\mathcal{T})$.*

Tutte also obtained an analogous result for weighted multidigraphs.

Let Γ be a multidigraph with vertex set $V(\Gamma) = \{1, \ldots, n\}$, and suppose that ε_{ij}^p is the *weight* (or the *conductance*) of the pth arc from i to j in Γ. The Kirchhoff matrix of Γ is the $n \times n$ matrix $L = L(\Gamma) = (\ell_{ij})$ with entries $\ell_{ij} = -\sum_{p=1}^{a_{ji}} \varepsilon_{ji}^p$, $j \neq i$, $i, j = 1, \ldots, n$, and $\ell_{ii} = -\sum_{j \neq i} \ell_{ij}$, $i = 1, \ldots, n$, where a_{ji} is the number of arcs from j to i in Γ. Observe that ℓ_{ii} is the total conductance of the arcs *converging* to i. The conductance (weight) of a subgraph of Γ and the weight of a set of multidigraphs are defined analogously to the case of multigraphs.

Suppose that $\mathcal{T}^i$ is the set of all spanning trees of Γ diverging from i, and L^{ij} is the cofactor of ℓ_{ij} in L, as before.

THEOREM 2 (matrix-tree theorem for weighted multidigraphs). *For any weighted multidigraph Γ and for any $i, j \in V(\Gamma)$, $L^{ij} = \varepsilon(\mathcal{T}^i)$.*

Observe that in the directed case, entries in different rows of L may have different cofactors, but all the entries of the same row have equal cofactors. For simplicity, Tutte formulates these theorems only for diagonal cofactors L^{ii}. The "directed" matrix-tree theorem concerning all L^{ij} is given in [12]. If the weights of all edges (arcs) are ones, Theorems 1 and 2 tell us about the *numbers* of the corresponding spanning trees.

We shall now formulate the matrix-forest lemmas and the matrix-forest theorem.

Consider the matrices

$$W(G) = I + L(G)$$

and

$$W(\Gamma) = I + L(\Gamma),$$

where I is the identity matrix. $W^{ij}(G)$ and $W^{ij}(\Gamma)$ will denote the cofactors of the (i, j)-entries of $W(G)$ and $W(\Gamma)$, respectively.

Suppose that $\mathcal{F}(G) = \mathcal{F}$ is the set of all spanning rooted forests of a weighted multigraph G and $\mathcal{F}^{ij}(G) = \mathcal{F}^{ij}$ is the set of those spanning rooted forests of G such that i and j belong to the same tree rooted at i. Let $W = W(G)$, $W^{ij} = W^{ij}(G)$.

LEMMA 1 (matrix-forest lemma for weighted multigraphs). *For any weighted multigraph G,*
(1) $\det W = \varepsilon(\mathcal{F})$;
(2) for any $i, j \in V(G)$, $W^{ij} = \varepsilon(\mathcal{F}^{ij})$.

Since the matrix W of a weighted multigraph is symmetric, item (2) of Lemma 1 remains true if we replace $\mathcal{F}^{ij}$ by $\mathcal{F}^{ji}$.

Suppose that $\mathcal{F}(\Gamma) = \mathcal{F}$ is the set of all spanning diverging forests of multidigraph Γ and $\mathcal{F}^{i \to j}(\Gamma) = \mathcal{F}^{i \to j}$ is the set of those spanning diverging forests of Γ such that i and j belong to the same tree diverging from i. Let $W = W(\Gamma)$, $W^{ij} = W^{ij}(\Gamma)$.

LEMMA 2 (matrix-forest lemma for weighted multidigraphs). *For any weighted multidigraph* Γ,
(1) $\det W = \varepsilon(\mathcal{F})$;
(2) *for any* $i, j \in V(\Gamma)$, $W^{ij} = \varepsilon(\mathcal{F}^{i \to j})$.

A similar lemma can be formulated for converging forests.
If the matrix W^{-1} exists, we will denote it by

$$Q = (q_{ij}) = W^{-1} = (I + L)^{-1} \tag{3}$$

(either for a weighted multigraph G or for a weighted multidigraph Γ). Then $Q = (\det W)^{-1} W^*$, where $W^* = \left(W^{ij}\right)^T$ is the adjugate matrix of W. The matrix-forest theorem [5, 3] follows from Lemmas 1 and 2.

THEOREM 3 (matrix-forest theorem).
1. *For any weighted multigraph* G, *the matrix* $Q = W^{-1}$ *exists and* $q_{ij} = \varepsilon(\mathcal{F}^{ji})/\varepsilon(\mathcal{F})$, $i, j = 1, \ldots, n$.
2. *For any weighted multidigraph* Γ, *the matrix* $Q = W^{-1}$ *exists and* $q_{ij} = \varepsilon(\mathcal{F}^{j \to i})/\varepsilon(\mathcal{F})$, $i, j = 1, \ldots, n$.

If the weights of all edges (arcs) are ones, the weights of sets of spanning forests in Lemmas 1 and 2 and Theorem 3 are equal to the *numbers* of the corresponding forests.

Lemma 2 can be derived in the shortest way from one version of Chaiken's result [13], namely, by putting $U = W = \varnothing$ and then $U = \{i\}$, $W = \{j\}$ in the first formula on page 328 (cf. [14, Theorem 3.1]). A longer inference results by the sequential application of results from [15–18]. This also provides an interpretation for the inverse Laplacian characteristic matrix of a multidigraph. An inference of Lemma 1 from Lemma 2 is given in the Appendix, as well as the proofs of the following results. Another complete (i.e., not exploiting any strong theorems) proof of Lemma 1 for the case of equal weights of edges is contained in [6].

The matrix-forest theorem allows us to consider the matrix $Q = W^{-1}$ as the matrix of *"relative forest accessibilities"* (in short, *accessibilities*) of the vertices of G (or Γ). These values can be used to measure the proximity between vertices (the "farther" i from j, the smaller is q_{ij}). This interpretation is validated by the properties presented in the following section. For simplicity, these properties are formulated for nonoriented multigraphs, although many of them have "oriented" counterparts which can be proved similarly.

3. PROPERTIES OF THE RELATIVE FOREST ACCESSIBILITIES

Suppose that G is a weighted multigraph with strictly positive weights of edges, and let

$$Q = (q_{ij}) = W^{-1}$$

be its matrix of relative forest accessibilities.

PROPOSITION 1. *For any* G, *matrix* Q *is symmetric.*

PROPOSITION 2. *For any* G, Q *is a doubly stochastic matrix, i.e.,*
(1) $q_{ij} \geq 0$, $i, j = 1, \ldots, n$;
(2) $\sum_{j=1}^{n} q_{ij} = 1$, $i = 1, \ldots, n$;
(3) $\sum_{i=1}^{n} q_{ij} = 1$, $j = 1, \ldots, n$.

According to this property, q_{ij} may be interpreted as the fraction of the connectivity of vertices i and j in the total connectivity of i with all vertices.

PROPOSITION 3. *For any* G *and for any* $i, j = 1, \ldots, n$ *such that* $j \neq i$, $q_{ii} > q_{ij}$.

This property has a natural interpretation, namely, each vertex is more "accessible" from itself than from any other vertex.

PROPOSITION 4 (triangle inequality for proximities). *For any* G *and for any* $i, j, k = 1, \ldots, n$, $q_{ij} + q_{ik} - q_{jk} \leq q_{ii}$. *If, in addition,* $i \neq j$ *and* $i \neq k$, *then* $q_{ij} + q_{ik} - q_{jk} < q_{ii}$.

Consider the index

$$d_{ij} = q_{ii} + q_{jj} - q_{ij} - q_{ji} = q_{ii} + q_{jj} - 2q_{ij}, \quad i, j = 1, \ldots, n. \tag{4}$$

1508

ASSERTION 1. $d(i,j) = d_{ij}$, $i,j = 1,\ldots,n$, is a distance function for multigraph vertices, i.e., it complies with the axioms of metric.

This assertion is easily proved using the above propositions (this is left to the reader). The triangle inequality for proximities turns out to be equivalent to the ordinary triangle inequality for metric d_{ij}, which justifies the name of the former inequality. In contrast to the standard graph distance, this metric considers all connections in a graph.

PROPOSITION 5. For any G and for any $i,j = 1,\ldots,n$, $q_{ij} = 0$ iff there exist no paths between i and j.

COROLLARY. (1) Matrix Q is reducible to a block-diagonal form, where all block entries are strictly positive and all off-block entries are zeros. Q is strictly positive iff multigraph G is connected.
(2) For any $i,j,k \in V(G)$, if $q_{ij} > 0$ and $q_{jk} > 0$, then $q_{ik} > 0$.

PROPOSITION 6. For any G and for any $i,k,t = 1,\ldots,n$,
(1) if there exists a path in G from i to k, $t \neq k$, and every path from i to t includes k, then $q_{ik} > q_{it}$.
(2) if $q_{ik} > q_{it}$ and $i \neq k$, then there exists a path from i to k, such that the difference $(q_{jk} - q_{jt})$ strictly increases as j progresses from i to k along the path.

PROPOSITION 7. Suppose that some edge weight ε_{kt}^{p} in G increases by $\Delta\varepsilon_{kt} > 0$ or an extra edge between k and t with a strictly positive weight $\Delta\varepsilon_{kt}$ is added to G. Let G' be the new graph and $W' = W(G')$, $Q' = Q(G')$. Then

(1) $\Delta Q = hR$, where $\Delta Q = Q' - Q$, $h = \dfrac{\Delta\varepsilon_{kt}}{\Delta\varepsilon_{kt}(q_{kk} + q_{tt} - 2q_{kt}) + 1} = (d_{kt} + 1/\Delta\varepsilon_{kt})^{-1}$, and $R = (r_{ij})$ is the $n \times n$ matrix with entries $r_{ij} = (q_{ik} - q_{it})(q_{jt} - q_{jk})$;
(2) (this item and the following three are corollaries from item (1)) all rows and all columns of ΔQ are proportional, i.e., $\mathrm{rank}\,\Delta Q = 1$;
(3) if $q_{ik} > q_{it}$, then $\Delta q_{ij} > 0$ iff $q_{jt} > q_{jk}$, and $\Delta q_{ij} < 0$ iff $q_{jk} > q_{jt}$;
(4) the signs of all increments Δq_{ij} do not depend on the absolute value of $\Delta\varepsilon_{kt}$, and the absolute values of nonzero Δq_{ij} strictly increase in $\Delta\varepsilon_{kt}$;
(5) for any $i,j \in V(G)$, $\Delta d_{ij} = -\frac{1}{4}(d_{ik} - d_{it} + d_{jt} - d_{jk})^2 (d_{kt} + 1/\Delta\varepsilon_{kt})^{-1}$, and therefore $d'_{ij} \leq d_{ij}$.

According to item (3), if the direct connection between k and t intensifies, then the relative accessibility of j from i increases if and only if i and j initially were "more strongly connected" with *different* vertices of the pair (k,t). Otherwise, it can be said that the connections in the multigraph intensify outside of most paths from i to j, thus the relative accessibility of j from i decreases.

Propositions 8 and 9 are corollaries of Proposition 7.

PROPOSITION 8. Suppose that some edge weight ε_{kt}^{p} in G increases or an extra edge between k and t with a positive weight is added to G. Then
(1) q_{kt} increases;
(2) for any $i = 1,\ldots,n$, if there exists a path from i to k and every path from i to t includes k, then $\Delta q_{it} > \Delta q_{ik}$;
(3) for any $i_1, i_2 = 1,\ldots,n$, if both i_1 and i_2 can be substituted for i in the hypothesis of item (2), then $q_{i_1 i_2}$ decreases;
(4) for any $i = 1,\ldots,n$, if $q_{ik} = q_{it}$, then q_{ij} do not alter for all $j = 1,\ldots,n$.

By item (3), the relative accessibility between a pair of vertices decreases when some "extraneous" connections appear or intensify in G.

Let D be a subset of vertex set $V(G)$. We say that D is a *macrovertex in G* if for all $i \in D$, $j \in D$, and $k \notin D$, $\ell_{ik} = \ell_{jk}$.

The following property is among to the most interesting ones. It provides a sufficient condition for the equality and stability of relative forest accessibilities.

PROPOSITION 9 (macrovertex independence). Suppose that D is a macrovertex in G and $i \in D$, $j \in D$, $k \notin D$. Then
(1) $q_{ik} = q_{jk}$;
(2) q_{ik} does not alter when any new edges appear or the weights of any existing edges change inside D.

Now we shall obtain an alternative topological interpretation of the matrix Q of relative forest accessibilities (the first interpretation is provided by Theorem 3). It will be demonstrated that q_{ij} are related to the weights of routes of various lengths between i and j in G. To be more precise, introduce the notion of route with drains.

A *route with drains* (RWD) is an alternating sequence of multigraph vertices and edges with the following features:

(1) the sequence starts and ends with vertices;

(2) the edge located between two *different* vertices in the sequence is incident to them. If the same vertex stands in the sequence before and after an edge, it is only required that it be incident to this edge, the second incident vertex being arbitrary. Such an edge is called a *drain*.

Routes with drains result from the usual routes by adding any number of one-edge offshoots (drains), which may, in particular, follow "forward" and "backward" along the original route.

The total number of edges in the sequence is called the *length* of the route with drains. Set, by definition, te fact that for any vertex i there exist one route of length 0 from i to i with 0 drains and no other routes of length 0.

The *weight* of a route with drains is defined as the product of the weights of all its edges (if an edge enters a route with drains k times, its weight is taken with exponent k). For any $i = 1, \ldots, n$, the weight of the 0-length RWD from i to i is set to be 1.

Let $a^* = \max\limits_{i,j \in V(G)} a_{ij}$ be the maximal number of multiple edges incident to any pair of vertices in G.

PROPOSITION 10. *For any weighted multigraph G with all weights of edges from the interval $\left(0, \left(2a^*(n-1)\right)^{-1}\right)$ and for any $i, j = 1, \ldots, n$,*

$$q_{ij} = \sum_{t=0}^{\infty} (U_{ij}^{(t)} - P_{ij}^{(t)}),$$

where $U_{ij}^{(t)}$ and $P_{ij}^{(t)}$ are the total weights of all routes of length t with even and odd number of drains between vertices i and j in G, respectively.

One more interpretation of Q can be obtained with the help of the Cayley–Hamilton theorem [8].

Instead of Q, one can use the matrices $Q_\alpha = (I + \alpha L)^{-1}$, $\alpha > 0$, which have the same properties as Q except for Proposition 10, where the factor α appears. The parameter $\alpha > 0$ specifies the proportions of accounting for long connections between vertices of G versus short ones.

4. ACCESSIBILITY AND DERIVATIVE STRUCTURAL INDICES

The foregoing properties of the relative forest accessibility demonstrate that it is an appropriate index of proximity (connectivity, accessibility) of graph (multigraph) vertices. A distinctive feature of this index is its normalization: the sum of the accessibilities of all vertices from a given one and the sum of the accessibilities of a given vertex from all vertices of a multigraph are equal to unity. Therefore, each ith row of the matrix Q can be treated as a probability distribution (or shares of a certain resource) somehow related to the vertex i. In which cases is such a normalization necessary? Consider two examples.

Suppose that the members of a group collect information from the environment and exchange it with each other, the intensity of the exchange being specified for each pair. Every participant transmits not only the information collected on her own, but also that received from the others. It is required to ascertain which fractions of the cumulative information received by the ith participant were initially collected by each member of the group. In this example, information can be replaced with, for example, influence or material resources. The principal feature is the distribution of some resource related to a certain vertex, over all vertices.

The second example is a variant of a children's "ring" game in which the ring may successively be passed many times, and this is done secretly, not before the players' eyes. If the pairwise transfer probabilities are specified for all players along with the temporal parameters of this random process (which is a Markov process in the simplest case), one can take an interest in the ring's location probabilities at every moment, provided that its starting location was at vertex (player) i. The main feature of this example is the presence of probability distributions related to each vertex.

In the above examples, if an adequate mathematical model is stated, the result is precise, not heuristic, and there is no need to select it being guided by "good properties," such as those given in the previous section. It turns out, however, that for both examples there are rather natural models (we intend to describe them and compare them with other models, e.g., [9, 19], in our next paper), which lead to relative forest accessibilities. This means, in turn, that even when there is no detailed model, only the intensities (or probabilities) of pairwise interactions being known, the relative forest accessibilities provide a comprehensible first approximation for the required values.

Now we turn to derivative structural indices. The value

$$1 - \sum_{j \neq i} q_{ij} = q_{ii}$$

can serve to measure the *solitariness* of the ith member of a group. Now a number of other indices can be constructed in the usual fashion. Specifically, the mean solitariness over a group,

$$\rho = \frac{1}{n} \sum_{i=1}^{n} q_{ii},$$

indicates the extent of its *dissociation*. The empirical variance of the solitariness evaluates the *heterogeneity* of a group. The ratio of q_{ii} to ρ (or their difference) measures the *provinciality* of the ith member of a group. Equation (4) introduces a specific distance between the members of a group (Assertion 1 in Section 3). The properties of all these indices are determined by those of the relative forest accessibilities studied above.

Notice, in conclusion, that there exists a certain relation between the problem of centrality (respectively, provinciality) evaluation and the problem of estimating the strength of players from incomplete tournaments. In the latter case, an "object–object" matrix is processed as well, but its entries express the results of *paired comparisons* (e.g., games or comparative preferences) rather than personal choices within a group. The problem of scoring from paired comparisons has been investigated a little bit better (but also insufficiently). It is worth noting, for example, that the work [20] was accepted as relevant in the literature on paired comparisons, though it was concerned with sociometric data. And conversely, sensitive scoring methods for preference aggregation can be considered with reference to sociometric data. A review of these methods can be found in [21].

APPENDIX

Proof of Lemma 1. Lemma 1 is reducible to Lemma 2, since for every multigraph G, the corresponding multidigraph Γ can be introduced by replacing every edge of G with a pair of opposite arcs with the same weight each. The matrix W (and thus Q) is the same for G and Γ, so the desired statements of Lemma 1 follow from the existence of a natural one-to-one correspondence between all spanning rooted forests in G and all spanning diverging forests in Γ.

Proposition 1 follows from the symmetry of W.

Proof of Proposition 2. Item (1) follows from Theorem 3 and the positiveness of edge weights.

Item (2) immediately follows from the fact that $W = Q^{-1}$ satisfies the same condition [8, 6]. Another easy proof is provided by Theorem 3 and the fact that for any $i_1, i_2, j \in V(G)$, $i_1 \neq i_2 \Rightarrow \mathcal{F}^{i_1 j} \cap \mathcal{F}^{i_2 j} = \varnothing$ and $\bigcup_{i=1}^{n} \mathcal{F}^{ij} = \mathcal{F}$.

Item (3) follows from item (2) and Proposition 1.

Proof of Proposition 3. Note that for any $i, j = 1, \ldots, n$ such that $j \neq i$ and for any $H \in \mathcal{F}$, if $H \in \mathcal{F}^{ij}$ then $H \in \mathcal{F}^{ii}$. Therefore, $\mathcal{F}^{ij} \subseteq \mathcal{F}^{ii}$. Let F_0 be a subgraph of G such that $V(F_0) = V(G)$ and $E(F_0) = \varnothing$. Then $F_0 \in \mathcal{F}^{ii} \smallsetminus \mathcal{F}^{ij}$ and $\varepsilon(F_0) = 1$, i.e., $\mathcal{F}^{ij} \subset \mathcal{F}^{ii}$ and $\varepsilon(\mathcal{F}^{ij}) < \varepsilon(\mathcal{F}^{ii})$. By Theorem 3, $q_{ii} > q_{ij}$.

Proof of Proposition 4. If $i = j$ or $i = k$ then, obviously,

$$q_{ij} + q_{ik} - q_{jk} = q_{ii}.$$

Assume that $i \neq j$ and $i \neq k$. In the same way as in the proof of Proposition 3, we have

$$\mathcal{F}^{ij} \cup \mathcal{F}^{ik} \subset \mathcal{F}^{ii},$$

and hence

$$\varepsilon(\mathcal{F}^{ij} \cup \mathcal{F}^{ik}) = \varepsilon(\mathcal{F}^{ij}) + \varepsilon(\mathcal{F}^{ik}) - \varepsilon(\mathcal{F}^{ij} \cap \mathcal{F}^{ik}) < \varepsilon(\mathcal{F}^{ii}). \tag{5}$$

Define $\mathcal{F}^{ijk}$ as $\mathcal{F}^{ij} \cap \mathcal{F}^{ik}$. Observe that $\mathcal{F}^{ijk}$ differs from $\mathcal{F}^{jik} = \mathcal{F}^{ji} \cap \mathcal{F}^{jk}$ only by the roots in the trees containing i, j, and k simultaneously. Therefore,

$$\varepsilon(\mathcal{F}^{ij} \cap \mathcal{F}^{ik}) = \varepsilon(\mathcal{F}^{ijk}) = \varepsilon(\mathcal{F}^{jik}) \leq \varepsilon(\mathcal{F}^{jk}). \tag{6}$$

Inequalities (5) and (6) imply

$$\varepsilon(\mathcal{F}^{ij}) + \varepsilon(\mathcal{F}^{ik}) - \varepsilon(\mathcal{F}^{jk}) < \varepsilon(\mathcal{F}^{ii}),$$

and, by Theorem 3,

$$q_{ij} + q_{ik} - q_{jk} < q_{ii} \, .$$

Proposition 5 follows directly from Theorem 3.

Proof of Proposition 6. Item (1). Note that $H \in \mathcal{F}^{it}$ implies $H \in \mathcal{F}^{ik}$. On the other hand, $\mathcal{F}^{ik} \smallsetminus \mathcal{F}^{it} \neq \varnothing$ and $\varepsilon(\mathcal{F}^{ik} \smallsetminus \mathcal{F}^{it}) > 0$. Hence by Theorem 3, $q_{ik} > q_{it}$.

Item (2). By virtue of Eq. (3),

$$(I + L)Q = I. \tag{7}$$

Rewrite (7) componentwise for entries ik and it of the matrix $(I + L)Q$. Using Eqs. (1) and (2), the notation $\varepsilon_{ij} = -\ell_{ij}$, and $i \neq t$, which follows from Proposition 3, we get

$$q_{ik} = \sum_{j \neq i} \varepsilon_{ij}(q_{jk} - q_{ik}),$$

$$q_{it} = \sum_{j \neq i} \varepsilon_{ij}(q_{jt} - q_{it}),$$

$$q_{ik} - q_{it} = \sum_{j \neq i} \varepsilon_{ij}[(q_{jk} - q_{jt}) - (q_{ik} - q_{it})].$$

Then, since $q_{ik} - q_{it} > 0$, there exists $j \neq i$ such that $\varepsilon_{ij} \neq 0$ (and thus $(ij) \in E(G)$) and $q_{jk} - q_{jt} > q_{ik} - q_{it}$ (recall that the case $\varepsilon_{ij} < 0$ is excluded).

Applying this argument to vertex j instead of i, and so forth, and taking into account that no vertex in the path thereby constituted may coincide with any previous one and that $i \neq k$, we finally obtain k as the terminal vertex of this path, as desired.

Proof of Proposition 7. Let $\Delta W = W' - W$. Note that $\Delta W = XY$, where $X = (x_{i1})$, $i = 1, \ldots, n$, is the column vector with entries $x_{k1} = 1$, $x_{t1} = -1$, and $x_{i1} = 0$ for all $i \neq k$, $i \neq t$; $Y = (y_{1j})$, $j = 1, \ldots, n$, is the row vector with entries $y_{1k} = \Delta\varepsilon_{kt}$, $y_{1t} = -\Delta\varepsilon_{kt}$, and $y_{1j} = 0$ for all $j \neq k$, $j \neq t$. According to [22, Sec. 0.7.4],

$$Q' = Q - \frac{1}{1 + YQX} QXYQ.$$

It is straightforward to verify that $(-\frac{1}{1+YQX}) = -h/\Delta\varepsilon_{kt}$ and $QXYQ = -\Delta\varepsilon_{kt} R$, and thereby item (1) is proved. Items (2) through (5) follow from item (1) and the nonnegativity of d_{kt} (see Proposition 3 or Assertion 1).

Proof of Proposition 8. Item (1). By Proposition 3, $q_{kk} > q_{kt}$ and $q_{tt} > q_{tk}$, and hence item (3) of Proposition 7 implies $\Delta q_{kt} > 0$.

Item (2). Setting $Q' = Q(G')$, by item (1) of Proposition 7 we have

$$\Delta q_{it} - \Delta q_{ik} = h(q_{ik} - q_{it})(q_{tt} - q_{tk}) - h(q_{ik} - q_{it})(q_{kt} - q_{kk})$$
$$= h(q_{ik} - q_{it})(q_{kk} + q_{tt} - q_{tk} - q_{kt}) = h(q_{ik} - q_{it})d_{kt} \, .$$

Now the desired inequality follows from item (1) of Proposition 6 together with Assertion 1.

Item (3). By item (1) of Proposition 6, $q_{i_1 k} > q_{i_1 t}$ and $q_{i_2 k} > q_{i_2 t}$, and by item (3) of Proposition 7, $\Delta q_{i_1 i_2} < 0$.

Item (4). By item (1) of Proposition 7 we have $\Delta q_{ij} = h(q_{ik} - q_{it})(q_{jt} - q_{jk}) = 0$.

Proof of Proposition 9. Consider the graph G' on the vertex set $V(G)$ such that

(1) $(ij) \in E(G')$ iff $i \neq j$ and $\ell_{ij} \neq 0$, and

(2) for every edge $(ij) \in E(G')$, $\varepsilon'_{ij} = -\ell_{ij}$.

Let $Q' = Q(G') = (q'_{ij})$. Obviously, D is a macrovertex in G' as well as in G. Let $S = V(G) \smallsetminus D$. First, we prove Proposition 9 for G'. Consider the graph G'' resulting from G' by deleting all edges inside D. Let $Q'' = Q(G'') = (q''_{ij})$. All vertices of D are symmetric in G''; therefore, $q''_{ik} = q''_{jk}$ for any $i, j \in D$, $k \in S$. Then using item (4) of Proposition 8, by means of induction we get $q'_{ik} = q''_{ik} = q''_{jk} = q'_{jk}$, for all $i, j \in D$ and $k \in S$. This proves Proposition 9, since $Q' = Q$.

Proof of Proposition 10. Expand $Q = (I - (-L))^{-1}$ as the sum of an infinitely decreasing geometric progression using the notation $M = (m_{ij}) = -L$:

$$Q = (I - M)^{-1} = I + M + M^2 + \ldots \quad . \tag{8}$$

This expansion is valid if and only if

$$|\lambda_1| < 1, \tag{9}$$

where $|\lambda_1|$ is the spectral radius of $M = -L$ [22, Corollary 5.6.16].

Consider the upper bound of $|\lambda|_{\max}$ provided by the Geršgorin theorem (see [22]):

$$|\lambda_1| \le \max_{1\le i\le n} \sum_{j=1}^{n} |\ell_{ij}|. \tag{10}$$

Let $\varepsilon_{\max} = \max_{1\le i\ne j\le n} \varepsilon_{ij}$, where $\varepsilon_{ij} = \sum_{p=1}^{a_{ij}} \varepsilon_{ij}^p = -\ell_{ij}$. Then by Eqs. (1) and (2),

$$\max_{1\le i\le n} \sum_{j=1}^{n} |\ell_{ij}| = 2 \max_{1\le i\le n} \sum_{j\ne i} |\ell_{ij}| \le 2 \max_{1\le i\le n} \sum_{j\ne i} a^* \varepsilon_{\max} = 2a^*(n-1)\varepsilon_{\max}. \tag{11}$$

Consequently, the fulfillment of (9), and therefore of (8) is assured by

$$\varepsilon_{\max} < \left(2a^*(n-1)\right)^{-1}.$$

By virtue of (8), it suffices to prove that

$$m_{ij}^{(k)} = U_{ij}^{(k)} - P_{ij}^{(k)}, \quad i, j = 1, \ldots, n, \quad k = 0, 1, 2, \ldots, \tag{12}$$

where $m_{ij}^{(k)}$, $i, j = 1, \ldots, n$, are the entries of M^k.

Let us apply induction on the length k of the roots with drains between i and j. The proof can be used with no change for the case of digraphs, because it does not use the symmetry of M.

1^0. $k = 0$. Equation (12) is valid because $M^0 = I$ and by the definition of root with drains, for every $i, j = 1, \ldots, n$, $j \ne i$, $U_{ii}^{(0)} = 1$ and $P_{ii}^{(0)} = P_{ij}^{(0)} = U_{ij}^{(0)} = 0$ hold.

2^0. Let (12) be valid for $k = v$. Prove it for $k = v+1$. Consider an arbitrary route μ of length $v+1$ with g drains between vertices i and j. Let t be the next to last vertex of μ. If $t \ne j$, then μ is representable as the combination of a route of length v with g drains from i to t and an edge (tj). Otherwise, $t = j$, and μ can be considered as the combination of a route between i and j with $g-1$ drains and an edge incident with j (this edge is the gth drain). Therefore,

$$U_{ij}^{(v+1)} = \sum_{t\ne j} U_{it}^{(v)} m_{tj} + \sum_{t\ne j} P_{ij}^{(v)} m_{jt},$$

$$P_{ij}^{(v+1)} = \sum_{t\ne j} P_{it}^{(v)} m_{tj} + \sum_{t\ne j} U_{ij}^{(v)} m_{jt}.$$

Then

$$U_{ij}^{(v+1)} - P_{ij}^{(v+1)} = \sum_{t\ne j} U_{it}^{(v)} m_{tj} + \sum_{t\ne j} P_{ij}^{(v)} m_{jt} - \sum_{t\ne j} P_{it}^{(v)} m_{tj} - \sum_{t\ne j} U_{ij}^{(v)} m_{jt}$$

$$= \sum_{t\ne j} (U_{it}^{(v)} - P_{it}^{(v)}) m_{tj} - \sum_{t\ne j} (U_{ij}^{(v)} - P_{ij}^{(v)}) m_{jt} \overset{\langle 1\rangle}{=} \sum_{t\ne j} m_{it}^{(v)} m_{tj} - m_{ij}^{(v)} \sum_{t\ne j} m_{jt}$$

$$\overset{\langle 2\rangle}{=} \sum_{t\ne j} m_{it}^{(v)} m_{tj} + m_{ij}^{(v)} m_{jj} = \sum_{t=1}^{n} m_{it}^{(v)} m_{tj} = m_{ij}^{(v+1)},$$

where transition $\langle 1 \rangle$ is carried out by the induction hypothesis, and $\langle 2 \rangle$ uses the equality $m_{jj} = -\sum_{t\ne j} m_{jt}$ which follows from Eq. (2) using $M = -L$. Proposition 10 is proved.

REFERENCES

1. V. I. Paniotto, "Analysis of the structure of interpersonal relations," in: *Mathematical Methods of the Analysis and Interpretation of Sociological Data* [in Russian], Nauka, Moscow (1989), pp. 121–162.

2. T. Høvik and N. P. Gleditsch, "Structural parameters of graphs: A theoretical investigation," in: H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, and V. Capecchi (eds.) *Quantitative Sociology. International Perspectives on Mathematical and Statistical Modeling*, Academic Press, New York (1975), pp. 203–223.

3. P. Yu. Chebotarev and E. Shamis, "On the proximity measure for graph vertices provided by the inverse Laplacian characteristic matrix," in: *5th Conference of the International Linear Algebra Society*, Georgia State University, Atlanta (1995), pp. 30–31.

4. P. Yu. Chebotarev, "Generalization of the row sum method for incomplete paired comparisons," *Automat. Remote Control*, **50**, No. 8, Part 2, 1103–1113 (1989).

5. E. Shamis, "Counting spanning converging forests," *Abstracts of Papers Presented to the American Mathematical Society*, **15**, 412–413 (1994).

6. E. Shamis, "Graph-theoretic interpretation of the generalized row sum method," *Math. Soc. Sci.*, **27**, 321–333 (1994).

7. R. Merris, "Doubly stochastic graph matrices," *Publikacije Elektrotehn. Fakulteta, Univerzitet U Beogradu*, **8**, 6–13 (1997).

8. V. E. Golender, V. V. Drboglav, and A. B. Rosenblit, "Graph potentials method and its application for chemical information processing," *J. Chem. Inf. Comput. Sci.*, **21**, 196–204 (1981).

9. N. E. Friedkin, "Theoretical foundations of centrality measures," *Amer. J. Sociology*, **96**, 1478–1504 (1991).

10. J. W. Moon, *Counting Labelled Trees*, Canad. Math. Congress, Montreal (1970).

11. W. T. Tutte, *Graph Theory*, Addison-Wesley, Reading, Massachusetts (1984).

12. F. Harary and E. M. Palmer, *Graphical Enumeration*, Academic Press, New York (1973).

13. S. Chaiken, "A combinatorial proof of the all minors matrix tree theorem," *SIAM J. Alg. Disc. Meth.*, **3**, 319–329 (1982).

14. J. W. Moon, "Some determinant expansions and the matrix-tree theorem," *Discrete Math.*, **124**, 163–171 (1994).

15. A. K. Kelmans, "On the properties of the characteristic polynomial of a graph," in: *Cybernetics Serves Communism* [in Russian], Vol. 4, Energiya, Moscow–Leningrad (1967), pp. 27–41.

16. A. K. Kelmans and V. M. Chelnokov, "A certain polynomial of a graph and graphs with an extremal number of trees," *J. Comb. Theory*, Ser. B, **16**, 197–214 (1974).

17. M. Fiedler and J. Sedláček, "O W-basích orientovaných grafů," *Časopis Pěst. Mat.*, **83**, 214–225 (1958).

18. J. S. Maybee, D. D. Olesky, P. van den Driessche, and G. Wiener, "Matrices, digraphs and determinants," *SIAM J. Matrix Anal. Appl.*, **10**, 500–519 (1989).

19. K. Yamaguchi, "The flow of information through social networks: diagonal-free measures of inefficiency and the structural determinants of inefficiency," *Social Networks*, **16**, 57–86 (1994).

20. L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, **18**, 39–43 (1953).

21. P. Yu. Chebotarev and E. Shamis, "Preference fusion when the number of alternatives exceeds two: Indirect scoring procedures," in: *Proceedings of Workshop on Foundations of Information / Decision Fusion*, ed. by N. S. V. Rao, V. Protopopescu, J. Bernen, and G. Seetharaman, Acadiana, Lafayette (1996), pp. 20–32.

22. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge (1986).