

Available online at www.sciencedirect.com**SciVerse ScienceDirect**

Procedia Computer Science 17 (2013) 1134 – 1141

Procedia
 Computer Science

Information Technology and Quantitative Management , ITQM 2013

Matrices of forests, analysis of networks, and ranking problems

Pavel Chebotarev^{a,*}, Rafiq Agaev^a^a*Institute of Control Sciences of the Russian Academy of Sciences, 65 Profsoyuznaya str., Moscow 117997, Russia*

Abstract

The matrices of spanning rooted forests are studied as a tool for analysing the structure of networks and measuring their properties. The problems of revealing the basic bicomponents, measuring vertex proximity, and ranking from preference relations / sports competitions are considered. It is shown that the vertex accessibility measure based on spanning forests has a number of desirable properties. An interpretation for the stochastic matrix of out-forests in terms of information dissemination is given.

© 2013 The Authors. Published by Elsevier B.V. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Selection and peer-review under responsibility of the organizers of the 2013 International Conference on Information Technology and Quantitative Management

Keywords: Laplacian matrix; spanning forest; matrix-forest theorem; proximity measure; bicomponent; ranking; incomplete tournament; paired comparisons

2010 MSC: 05C50, 05C05, 91B10, 62J15, 90B15, 60J10

1. Introduction

The matrices of walks between vertices are useful to analyse the structure of networks (see, e.g., [1] and the references therein). These matrices are the powers of the adjacency matrix. In this paper, we consider the matrices of spanning rooted forests as an alternative tool for analyzing networks (cf. [2]). We show how they can be used for measuring vertex proximity (Section 4) and for ranking on the base of preference relations / sports competitions (Section 5). In the first sections of the paper, we introduce the necessary notation (Section 2) and list some properties of spanning rooted forests and forest matrices (Section 3).

Three features that distinguish the matrices of forests from the matrices of walks are notable. First, all column sums (or row sums) of the forest matrices are the same, therefore, these matrices can be considered as matrices of *relative* accessibility. Second, there are matrices of “out-forests” and matrices of “in-forests”, enabling one to distinguish “out-accessibility” from “in-accessibility”, which is intuitively justifiable. Third, the total weights of maximum spanning forests are closely related to the Cesáro limiting probabilities of Markov chains determined by the network under consideration.

2. Notation and simple facts

2.1. Networks, components, and bases

Suppose that Γ is a weighted digraph (= network) without loops, $V(\Gamma) = \{1, \dots, n\}$, $n > 1$, is its set of vertices and $E(\Gamma)$ the set of arcs. Let $W = (w_{ij})$ be the matrix of arc weights. Its entry w_{ij} is zero if there is no arc from

*Corresponding author. Tel.: +7-495-334-8869; fax: +7-495-420-2016.

E-mail address: upi@ipu.ru; pavel4e@gmail.com.

vertex i to vertex j in Γ ; otherwise w_{ij} is strictly positive. In what follows, Γ is fixed, unless otherwise specified. If Γ' is a subgraph of Γ , then the weight of Γ' , $w(\Gamma')$, is the product of the weights of all its arcs; if $E(\Gamma') = \emptyset$, then $w(\Gamma') = 1$ by definition. The weight of a nonempty set of digraphs \mathcal{G} is

$$w(\mathcal{G}) = \sum_{H \in \mathcal{G}} w(H); \quad w(\emptyset) = 0. \quad (1)$$

A *spanning subgraph* of Γ is a subgraph with vertex set $V(\Gamma)$. The *indegree* $\text{id}(v)$ and *outdegree* $\text{od}(v)$ of a vertex v are the number of arcs that come *in* v and *out of* v , respectively. A vertex v is called a *source* if $\text{id}(v) = 0$. A vertex v is *isolated* if $\text{id}(v) = \text{od}(v) = 0$. A *walk* (*semiwalk*) is an alternating sequence of vertices and arcs $v_0, e_1, v_1, \dots, e_k, v_k$ with every arc e_i being (v_{i-1}, v_i) (resp., either (v_{i-1}, v_i) or (v_i, v_{i-1})). A *path* is a walk with distinct vertices. A *circuit* is a walk with $v_0 = v_k$, the other vertices being distinct and different from v_0 . Vertex v is *reachable* from vertex z in Γ if $v = z$ or Γ contains a path from z to v .

A digraph is *strongly connected* (or *strong*) if all of its vertices are mutually reachable and *weakly connected* if any two different vertices are connected by a semiwalk. Any maximal strongly connected (weakly connected) subgraph of Γ is a *strong component*, or a *bicomponent* (resp., a *weak component*) of Γ . Let $\Gamma_1, \dots, \Gamma_r$ be all the strong components of Γ . The *condensation* (or *factorgraph*, or *leaf composition*, or *Hertz graph*) Γ° of digraph Γ is the digraph with vertex set $\{\Gamma_1, \dots, \Gamma_r\}$, where arc (Γ_i, Γ_j) belongs to $E(\Gamma^\circ)$ iff $E(\Gamma)$ contains at least one arc from a vertex of Γ_i to a vertex of Γ_j . The condensation of any digraph Γ obviously contains no circuits.

A *vertex basis* of a digraph Γ is any minimal (by inclusion) collection of vertices such that every vertex of Γ is reachable from at least one vertex of the collection. If a digraph does not contain circuits, then its vertex basis is obviously unique and coincides with the set of all sources [3, 4]. That is why the bicomponents of Γ that correspond to the sources of Γ° are called the *basic bicomponents* [4] or *source bicomponents* of Γ . In this paper, the term *source knot* of Γ will stand for the set of vertices of any source bicomponent of Γ . In [5], source knots are called *W-bases*.

The following statement [3, 4] characterizes all the vertex bases of a digraph.

Proposition 1. *A set $U \subseteq V(\Gamma)$ is a vertex basis of Γ if and only if U contains exactly one vertex from each source knot of Γ and no other vertices.*

Schwartz [6] referred to the source knots of a digraph as the *minimum P -undominated sets*. According to his *Generalized Optimal Choice Axiom* (GOCHA), if a digraph represents a preference relation on a set of alternatives, then the *choice* should be the union of its minimum P -undominated sets.¹ This choice is interpreted as the set of “best” alternatives. A review of choice rules of this kind can be found in [7]; for “fuzzy” extensions, see [8].

2.2. Matrices of forests

A *diverging tree* is a weakly connected digraph in which one vertex (called the *root*) has indegree zero and the remaining vertices have indegree one. A diverging tree is said to *diverge from* its root. Spanning diverging trees are sometimes called *out-arborescences*. A *diverging forest* (or *diverging branching*) is a digraph all of whose weak components are diverging trees. The roots of these trees are called the roots of the diverging forest. A *converging tree* (*converging forest*) is a digraph that can be obtained from a diverging tree (resp., diverging forest) by the reversal of all arcs. The roots of a converging forest are its vertices that have outdegree zero. In what follows, spanning diverging forests in Γ will be called *out-forests* of Γ ; spanning converging forests in Γ will be called *in-forests* of Γ .

Definition 1. An out-forest F of a digraph Γ is called a *maximum out-forest* of Γ if Γ has no out-forest with a greater number of arcs than in F .

It is easily seen that every maximum out-forest of Γ has the minimum possible number of diverging trees; this number will be called the *out-forest dimension* of Γ and denoted by d' . It can be easily shown that the number of arcs in any maximum out-forest is $n - d'$; in general, the number of weak components in a forest with k arcs is $n - k$.

¹This union is also called the *top cycle* and the *strong basis* of Γ .

By $\mathcal{F}^{*\rightarrow}(\Gamma) = \mathcal{F}^{*\rightarrow}$ and $\mathcal{F}_k^{*\rightarrow}(\Gamma) = \mathcal{F}_k^{*\rightarrow}$ we denote the set of all out-forests of Γ and the set of all out-forests of Γ with k arcs, respectively; $\mathcal{F}_k^{i*\rightarrow j}$ will designate the set of all out-forests with k arcs where j belongs to a tree diverging from i ; $\mathcal{F}^{i*\rightarrow j} = \bigcup_{k=0}^{n-d'} \mathcal{F}_k^{i*\rightarrow j}$ is the set of such out-forests with all possible numbers of arcs. The notation like $\mathcal{F}_{(k)}^{*\rightarrow}$ will be used for sets of out-forests that consist of k trees, so $\mathcal{F}_{(k)}^{*\rightarrow} = \mathcal{F}_{n-k}^{*\rightarrow}$, $k = 1, \dots, n$. Thus, the $*\rightarrow$ sign relates to out-forests; the corresponding notation with $\rightarrow*$, such as $\mathcal{F}^{\rightarrow*}$, relates to in-forests, i.e., $*$ images the root(s).

Let

$$\sigma_k = w(\mathcal{F}_k^{*\rightarrow}), \quad k = 0, 1, \dots; \quad \sigma = w(\mathcal{F}^{*\rightarrow}) = \sum_{k=0}^{n-d'} \sigma_k. \tag{2}$$

By (2) and (1), $\sigma_k = 0$ whenever $k > n - d'$; $\sigma_0 = 1$.

We also introduce the parametric value

$$\sigma(\tau) = \sum_{k=0}^{n-d'} w(\mathcal{F}_k^{*\rightarrow}) \tau^k = \sum_{k=0}^{n-d'} \sigma_k \tau^k, \quad \tau > 0, \tag{3}$$

which is the total weight of out-forests in Γ provided that all arc weights are multiplied by τ .

Consider the matrices $Q_k = (q_{ij}^k)$, $k = 0, 1, \dots$, of out-forests of Γ with k arcs: the entries of Q_k are

$$q_{ij}^k = w(\mathcal{F}_k^{i*\rightarrow j}). \tag{4}$$

By (4) and (1), $Q_k = 0$ whenever $k > n - d'$; $Q_0 = I$.

The matrix of all out-forests is

$$Q = (q_{ij}) = \sum_{k=0}^{n-d'} Q_k \quad \text{with entries } q_{ij} = w(\mathcal{F}^{i*\rightarrow j}). \tag{5}$$

We will also consider the stochastic matrices of out-forests:

$$J_k = \sigma_k^{-1} Q_k, \quad k = 0, \dots, n - d'; \quad J = (J_{ij}) = \sigma^{-1} Q \tag{6}$$

and the parametric matrices

$$Q(\tau) = \sum_{k=0}^{n-d'} Q_k \tau^k \quad \text{and } J(\tau) = \sigma^{-1}(\tau) Q(\tau), \quad \tau > 0, \tag{7}$$

where σ_k , σ , and $\sigma(\tau)$ are defined by (2) and (3).

The stochastic matrix of maximum out-forests $J_{n-d'}$ will also be denoted by $\tilde{J} = (\tilde{J}_{ij})$:

$$\tilde{J} = J_{n-d'}. \tag{8}$$

The matrices of forests can be found by means of matrix analysis [9, Section 5].

3. Properties of the forest matrices

A number of results on the forest matrices are presented in [9]. Some of them are collected in the following theorem.

Theorem 1. [9]. 1. Matrices J_k , $k = 0, \dots, n - d'$, J , and $J(\tau)$ are column stochastic.

2. For any $\tau > 0$, $Q(\tau) = \text{adj}(I + \tau L)$ and $\sigma(\tau) = \det(I + \tau L)$ hold, whence, $J(\tau) = (I + \tau L)^{-1}$.

3. $L\tilde{J} = \tilde{J}L = 0$.

4. \tilde{J} is idempotent: $\tilde{J}^2 = \tilde{J}$.

5. $\tilde{J} = \lim_{\tau \rightarrow \infty} J(\tau) = \lim_{\tau \rightarrow \infty} (I + \tau L)^{-1}$.

6. $\text{rank } \tilde{J} = d'$; $\text{rank } L = n - d'$.

7. $Q_k = \sum_{i=0}^k \sigma_{k-i} (-L)^i$, $k = 0, 1, \dots$.

8. \tilde{J} is the eigenprojection of L .

Item 2 of Theorem 1 is a parametric version of the *matrix-forest theorem* [10].

To formulate the topological properties of the matrix \tilde{J} , the following notation is needed.

Let $\tilde{K} = \cup_i K_i$, where K_i are all the source knots of Γ ; let K_i^+ be the set of all vertices reachable from K_i and unreachable from the other source knots. For any $k \in \tilde{K}$, $K(k)$ will designate the source knot that contains k . For any source knot K of Γ , denote by Γ_K the restriction of Γ to K and by Γ_{-K} the subgraph with vertex set $V(\Gamma)$ and arc set $E(\Gamma) \setminus E(\Gamma_K)$. For a fixed K , \mathcal{T} will designate the set of all spanning diverging trees of Γ_K , and \mathcal{P} the set of all maximum out-forests of Γ_{-K} . By \mathcal{T}^k , $k \in K$, we denote the subset of \mathcal{T} consisting of all trees that diverge from k , and by $\mathcal{P}^{K^* \rightarrow j}$, $j \in V(\Gamma)$, the set of all maximum out-forests of Γ_{-K} such that j is reachable from some vertex that belongs to K in these forests. $\tilde{J}_{k\bullet}$ is the k th row of \tilde{J} .

Theorem 2. [11]. *Let K be a source knot in Γ . Then the following statements hold.*

1. $\tilde{J}_{ij} \neq 0 \Leftrightarrow (i \in \tilde{K} \text{ and } j \text{ is reachable from } i \text{ in } \Gamma)$.
2. *Let $k \in K$. For any $j \in V(\Gamma)$, $\tilde{J}_{kj} = w(\mathcal{T}^k)w(\mathcal{P}^{K^* \rightarrow j})/w(\mathcal{F}_{(d^*)}^{* \rightarrow})$. Furthermore, if $j \in K^+$, then $\tilde{J}_{kj} = \tilde{J}_{kk} = w(\mathcal{T}^k)/w(\mathcal{T})$.*
3. $\sum_{k \in K} \tilde{J}_{kk} = 1$. *In particular, if k is a source, then $\tilde{J}_{kk} = 1$.*
4. *For any $k_1, k_2 \in K$, $\tilde{J}_{k_2\bullet} = (w(\mathcal{T}^{k_2})/w(\mathcal{T}^{k_1}))\tilde{J}_{k_1\bullet}$ holds, i.e., the rows k_1 and k_2 of \tilde{J} are proportional.*

We say that a weighted digraph Γ and a finite homogeneous Markov chain with transition probability matrix P *inversely correspond* to each other if

$$I - P = \alpha L^T, \tag{9}$$

where α is any nonzero real number.

If a Markov chain inversely corresponds to Γ , then the probability of transition from j to $i \neq j$ is proportional to the weight of arc (i, j) in Γ and is 0 if $E(\Gamma)$ does not contain (i, j) . We consider such an *inverse* correspondence in order to model preference digraphs in Section 5: in this case, the transitions in the Markov chain are performed from “worse” objects to “better” ones, so the Markov chain stochastically “searches the leaders.”

Theorem 3. *For any finite Markov chain, its matrix of Cesáro limiting probabilities coincides with the matrix \tilde{J} of any digraph inversely corresponding to this Markov chain.*

Theorem 3 follows from the *Markov chain tree theorem* first proved by Wentzell and Freidlin [12] and rediscovered in [13, 14], which, in turn, can be immediately proved using item 8 of Theorem 1 and a result of [15] (see [9]). Another proof of Theorem 3 can be found in [16]. A review on forest representations of Markov chain probabilities is given in [17]. For an interpretation of $J(\tau)$ in terms of Markov chains we refer to [18].

4. Forest based accessibility measures

Formally, by an *accessibility measure* for digraph vertices we mean any function that assigns a matrix $P = (p_{ij})_{n \times n}$ to every weighted digraph Γ , where $n = |V(\Gamma)|$. Entry p_{ij} is interpreted as the accessibility (or connectivity, relatedness, proximity, etc.) of j from i .

Consider the accessibility measures $P_\tau^{\text{out}} = J(\tau)$, where $J(\tau)$ is defined by (7), and $P_\tau^{\text{in}} = (p_{ij}^{\text{in}})$ with $p_{ij}^{\text{in}} = w(\mathcal{F}^{i \rightarrow *j}(\tau))/w(\mathcal{F}^{* \rightarrow}(\tau))$, where $\mathcal{F}^{i \rightarrow *j}(\tau)$ and $\mathcal{F}^{* \rightarrow}(\tau)$ are, respectively, the $\mathcal{F}^{i \rightarrow *j}$ and $\mathcal{F}^{* \rightarrow}$ for the digraph $\Gamma(\tau)$ obtained from Γ by the multiplication of all arc weights by τ . Parameter τ specifies the relative weight of short and long ties in Γ .

Definition 2. Accessibility measures $P^{(1)}$ and $P^{(2)}$ are *dual* if for every Γ and every $i, j \in V(\Gamma)$, $p_{ij}^{(1)}(\Gamma) = p_{ji}^{(2)}(\Gamma')$, where Γ' is obtained from Γ by the reversal of all arcs (preserving their weights).

The following proposition results from the fact that the reversal of all arcs in Γ transforms all out-forests into in-forests and vice versa.

Proposition 2. *For every $\tau > 0$, the measures P_τ^{out} and P_τ^{in} are dual.*

What is the difference in interpretation between P_{τ}^{out} and P_{τ}^{in} ? A partial answer is as follows. P_{τ}^{out} can be interpreted as the relative weight of $i \rightarrow j$ connections among the out-connections of i , whereas P_{τ}^{in} is the relative weight of $i \rightarrow j$ connections among the in-connections of j . Naturally, these relative weights need not coincide. For example, a connection between an average man and a celebrity is usually more important for the average man. This example demonstrates that self-duality is not an imperative requirement to accessibility measures. The properties of several self-dual measures have been studied in [19].

The following conditions some of which were proposed in [19] can be considered as desirable properties of vertex accessibility measures.

Nonnegativity. $p_{ij} \geq 0, i, j \in V(\Gamma)$.

Reachability condition. For any $i, j \in V(\Gamma)$, $(p_{ij} = 0 \Leftrightarrow j \text{ is unreachable from } i)$.

Self-accessibility condition. For any distinct $i, j \in V(\Gamma)$, (A) $p_{ii} > p_{ij}$ and (B) $p_{ii} > p_{ji}$ hold.

Triangle inequalities for proximities. For any $i, k, t \in V(\Gamma)$, (A) $p_{ki} - p_{ti} \leq p_{kk} - p_{tk}$ and (B) $p_{ik} - p_{it} \leq p_{kk} - p_{kt}$ hold.

The triangle inequalities for proximities is a counterpart of the ordinary triangle inequality which characterizes distances (cf. [20]).

Let $k, i, t \in V(\Gamma)$. We say that k mediates between i and t if Γ contains a path from i to t , $i \neq k \neq t$, and every path from i to t includes k .

Transit property. If k mediates between i and t , then (A) $p_{ik} > p_{it}$ and (B) $p_{kt} > p_{it}$.

Monotonicity. Suppose that the weight w_{kt} of some arc (k, t) is increased or a new (k, t) arc is added to Γ , and $\Delta p_{ij}, i, j \in V(\Gamma)$, are the resulting increments of the accessibilities. Then:

- (1) $\Delta p_{kt} > 0$;
- (2) If t mediates between k and i , then $\Delta p_{ki} > \Delta p_{it}$; if k mediates between i and t then $\Delta p_{it} > \Delta p_{ik}$;
- (3) (A) If t mediates between k and i , then $\Delta p_{kt} > \Delta p_{ki}$;
- (B) If k mediates between i and t , then $\Delta p_{kt} > \Delta p_{it}$.

Convexity. (A) If $p_{ki} > p_{ti}$ and $i \neq k$, then there exists a k to i path such that the difference $p_{kj} - p_{tj}$ strictly decreases as j advances from k to i along this path. (B) If $p_{ik} > p_{it}$ and $i \neq k$, then there exists an i to k path such that the difference $p_{jk} - p_{jt}$ strictly increases as j advances from i to k along this path.

The results of testing P_{τ}^{out} and P_{τ}^{in} are collected in

Theorem 4. *The measures P_{τ}^{out} and P_{τ}^{in} satisfy all the above conditions not partitioned into (A) and (B). Furthermore, P_{τ}^{out} obeys all (A) conditions and P_{τ}^{in} all (B) conditions.*

Consider now the accessibility measures $\tilde{P}^{\text{out}} = (p_{ij}) = \tilde{J} = \lim_{\tau \rightarrow \infty} P_{\tau}^{\text{out}}$ and $\tilde{P}^{\text{in}} = \lim_{\tau \rightarrow \infty} P_{\tau}^{\text{in}}$. Having in mind Theorem 3, we call \tilde{J}_{ij} the limiting out-accessibility of j from i .

Let us say that a condition is satisfied in the nonstrict form if it is not generally satisfied, but it becomes true after the substitution of \geq for $>$, \leq for $<$ and “nonstrictly” for “strictly” in the conclusion of this condition.

Similarly to Proposition 2 we have

Proposition 3. *The accessibility measures \tilde{P}^{out} and \tilde{P}^{in} are dual.*

The results of testing \tilde{P}^{out} and \tilde{P}^{in} are collected in

Theorem 5. *The accessibility measures \tilde{P}^{out} and \tilde{P}^{in} satisfy nonnegativity and the “ \Leftarrow ” part of reachability condition, but they violate the “ \Rightarrow ” part of reachability condition. Moreover, \tilde{P}^{out} satisfies, in the nonstrict form, items (A) of self-accessibility condition, transit property, monotonicity, and convexity, whereas \tilde{P}^{in} satisfies in the nonstrict form items (B) of these conditions. \tilde{P}^{out} satisfies (A) and \tilde{P}^{in} satisfies (B) of triangle inequality for proximities.*

By virtue of Theorem 5, the limiting accessibility measures only “marginally” correspond to the conception of accessibility that underlies the above conditions.

Proof. The nonstrict satisfaction of the conditions listed in the theorem follows from Theorem 4, Proposition 3 and item 5 of Theorem 1. To prove that the strict forms of these conditions and the “ \Rightarrow ” part of reachability condition are violated, it suffices to consider the digraph Γ with $n \geq 3$, $E(\Gamma) = \{(1, 2), (2, 3)\}$, and $w_{12} = w_{23} = 1$. \square

Let us mention one more class of accessibility measures: those of the form $(I + \alpha \tilde{J})^{-1}$, $0 < \alpha < \sigma_{(d')} / \sigma_{(d'+1)}$. These measures are “intermediate” between P_{τ}^{out} and \tilde{P}^{out} , because they are positive linear combination of $J_{(d')}$ and $J_{(d'+1)}$ [18]. That is why we termed them the *matrices of dense out-forests*. In the terminology of [21, p. 152], $(I + \alpha \tilde{J})^{-1}$ with various sufficiently small $\alpha > 0$ make up a class of *nonnegative nonsingular commuting weak inverses* for L . These measures and the dual measures have been studied in [18] (see also [9, p. 270–271]). Other interesting related topics are the forest distances [22] and the forest based centrality measures [10].

5. Rooted forests and the problem of leaders

Ranking from tournaments or irregular pairwise contests is an old, but still intriguing problem. Its statistical version is ranking objects on the basis of paired comparisons [23]. Analogous problems of the analysis of individual and collective preferences arise in the contexts of policy, economics, management science, sociology, psychology, etc. Hundreds of methods have been proposed for handling these problems (for a review, see, e.g., [23, 24, 25, 26, 27, 28, 29, 30, 31]).

In this section, we consider a weighted digraph Γ that represents a competition (which need not be a round robin tournament, i.e., can be “incomplete”) with weighted pairwise results. The digraph can also represent an arbitrary weighted preference relation. The result we present below can be easily extended to multidigraphs.

One of the popular exquisite methods for assigning scores to the participants in a tournament was independently proposed by Daniels [32], Moon and Pullman [33, 34], and Ushakov [35, 36] and reduces to finding nonzero and nonnegative solutions to the system of equations

$$Lx = 0. \tag{10}$$

Entry x_i of a solution vector $x = (x_1, \dots, x_n)$ is considered as a sophisticated “score” attached to vertex i . This method was multiply rediscovered with different motivations (some references are given in [28]). As Berman [37] noticed (although, in other contexts, similar results had been obtained by Maxwell [38] and other writers, see [39]), if a digraph is strong, then the general solution to (10) is provided by the vectors proportional to $t = (t_1, \dots, t_n)^T$, where t_j is the weight of the set of spanning trees (out-arborescences) diverging from j . This fact can be easily proved as follows. By the matrix-tree theorem for digraphs (see, e.g., [3]), t_j is the cofactor of any entry in the j th column of L . Then for every $i \in V(\Gamma)$, $\sum_{j=1}^n \ell_{ij} t_j = \det L$ (the row expansion of $\det L$) and, since $\det L = 0$, t is a solution to (10). As $\text{rank } L = n - 1$ (since the cofactors of L are nonzero), any solution to (10) is proportional to t .

Berman [37] and Berman and Liu [40] asserted that this result is sufficient to rank the players in an arbitrary competition, since the strong components of the corresponding digraph supposedly “can be ranked such that every player in a component of higher rank defeats every player in a component of lower rank. Now by ranking the players in each component we obtain a ranking of all the players.” While the statement about the existence of a natural order of the strong components is correct in the case of round-robin tournaments, it need not be true for arbitrary digraphs that may have, for instance, several source knots. That is why, the solution devised for strong digraphs does not enable one to rank the vertices of an arbitrary digraph.

Let us consider the problem of interpreting, in terms of forests, the general solution to (10) and the problem of choosing a particular solution that could serve as a reasonable score vector in the case of arbitrary digraph Γ .

If Γ contains more than one source knot, there is no spanning diverging tree in Γ . Recall that $K_1, \dots, K_{d'}$ are the source knots of Γ , where d' is the out-forest dimension of Γ , and $\tilde{K} = \bigcup_{s=1}^{d'} K_s$.

Suppose, without loss of generality, that the vertices of Γ are numbered as follows. The smallest numbers are attached to the vertices in K_1 , the following numbers to the vertices in K_2 , etc., and the largest numbers to the vertices in $V(\Gamma) \setminus \tilde{K}$. Such a numeration we call *standard*.

Theorem 6. *Any column of \tilde{J} is a solution to (10). Suppose that the numeration of vertices is standard and $j_1 \in K_1, \dots, j_{d'} \in K_{d'}$. Then the columns $\tilde{J}_{\bullet j_1}, \dots, \tilde{J}_{\bullet j_{d'}}$ of Γ make up an orthogonal basis in the space of solutions*

to (10) and $\tilde{J}_{\bullet j_s} = w^{-1}(\mathcal{T}_s)(0, \dots, 0, w(\mathcal{T}_s^{i_s+1}), \dots, w(\mathcal{T}_s^{i_s+k_s}), 0, \dots, 0)^T$, where $\{i_s+1, \dots, i_s+k_s\} = K_s$ and \mathcal{T}_s is the set of out-arborescences of K_s , $s = 1, \dots, d'$.

By virtue of Theorem 6, the general solution to (10) is the set of all linear combinations of partial solutions that correspond to each source knot of Γ .

Proof. The first statement follows from $L\tilde{J} = 0$ (item 3 of Theorem 1). By item 6 of Theorem 1, $\text{rank } \tilde{J} = d'$ and $\text{rank } L = n - d'$. Hence, d' is the dimension of the space of solutions to (10). Let $j_s \in K_s$, $s = 1, \dots, d'$. Then, by items 1 and 2 of Theorem 2,

$$\tilde{J}_{\bullet j_s} = w^{-1}(\mathcal{T}_s)(0, \dots, 0, w(\mathcal{T}_s^{i_s+1}), \dots, w(\mathcal{T}_s^{i_s+k_s}), 0, \dots, 0)^T.$$

These d' solutions to (10) are orthogonal and thus, linearly independent. □

As a reasonable ultimate score vector, the arithmetic mean $x = \frac{1}{n}\tilde{J} \cdot (1, \dots, 1)^T$ of the columns of \tilde{J} can be considered. A nice interpretation of this vector is given by

Corollary 1. (of Theorem 3). *For any Markov chain inversely corresponding to Γ , $x = \frac{1}{n}\tilde{J} \cdot (1, \dots, 1)^T$ is the limiting state distribution, provided that the initial state distribution is uniform.*

It can be mentioned, however, that the ranking method based on \tilde{J} takes into account long paths in Γ only. That is why, in any solutions to (10), the vertices that are not in the source knots are assigned zero scores, which is questionable. The estimates based on the matrices $Q(\tau)$, instead of \tilde{J} , are free of this feature. On the other hand, both methods violate the *self-consistent monotonicity* axiom [28], and so do the methods that count the *walks* between vertices. This axiom is satisfied by the *generalized Borda method* [41, 42] that produces the score vectors $J'(\tau) \cdot (\text{od}(1) - \text{id}(1), \dots, \text{od}(n) - \text{id}(n))^T$, where $J'(\tau)$ is the matrix $J(\tau)$ of the undirected graph corresponding to Γ [43]. In our opinion, the latter method can be recommended as a well-grounded approach to scoring objects on the base of arbitrary weighted preference relations, incomplete tournaments, irregular pairwise contests, etc.

A concluding remark: a communicatory interpretation of some forest matrices

In closing, let us mention an interpretation of forest matrices in terms of information dissemination. Consider the following metaphorical model. First, a plan of information transmission along a digraph is chosen. Such a plan is a diverging forest $F \in \mathcal{F}^{* \rightarrow}$: the information is injected into the roots of F ; then it ought to come to the other vertices along the arcs of F . Suppose that $w_{ij} \in]0, 1]$ is the probability of successful information transmission along the (i, j) arc, $i, j \in V(\Gamma)$, and that the transmission processes in different arcs are statistically independent. Then $w(F)$ is the probability that plan F is successfully realized. Suppose now that each plan is selected with the same probability $|\mathcal{F}^{* \rightarrow}|^{-1}$. Then J_{ij} (see (6)) is the probability that the information came to j from root i , provided that the transmission was successful. As a result, if one knows that the information was corrupted at root i and the transmission was successful, then J_{ij} is the probability that this corrupted information came to j .

Similarly, interpretations of this kind can be given to other stochastic forest matrices. This model is compatible with that of centered partitions [44] and comparable with some models of [45].

References

- [1] P. Chebotarev, The walk distances in graphs, *Discrete Applied Mathematics* 160 (10–11) (2012) 1484–1500.
- [2] M. Senelle, S. Garcia-Diez, A. Mantrach, M. Shimbo, M. Saerens, F. Fouss, The sum-over-forests density index: Identifying dense regions in a graph, arXiv preprint cs.LG/arXiv:1301.0725v1, submitted for publication (2013).
- [3] F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA, 1969.
- [4] A. A. Zykov, *Teoriya konechnykh Grafov (Theory of Finite Graphs)*, Nauka, Novosibirsk, 1969, in Russian.
- [5] M. Fiedler, J. Sedláček, O W -basích orientovaných grafů, *Časopis pro Pěstování Matematiky* 83 (1958) 214–225.
- [6] T. Schwartz, *The Logic of Collective Choice*, Columbia University Press, New York, 1986.
- [7] V. I. Vol'skii, Choice of best alternatives on directed graphs and tournaments, *Automation and Remote Control* 49 (3) (1988) 267–278.
- [8] M. Roubens, Choice procedures in fuzzy multicriteria decision analysis based on pairwise comparisons, *Fuzzy Sets and Systems* 84 (2) (1996) 135–142.
- [9] P. Chebotarev, R. Agaev, Forest matrices around the Laplacian matrix, *Linear Algebra and its Applications* 356 (2002) 253–274.

- [10] P. Y. Chebotarev, E. V. Shamis, The matrix-forest theorem and measuring relations in small social groups, *Automation and Remote Control* 58 (9) (1997) 1505–1514.
- [11] R. P. Agaev, P. Y. Chebotarev, The matrix of maximum out forests of a digraph and its applications, *Automation and Remote Control* 61 (9) (2000) 1424–1450.
- [12] A. D. Wentzell, M. I. Freidlin, On small random perturbations of dynamical systems, *Russian Mathematical Surveys* 25 (1) (1970) 1–55.
- [13] T. Leighton, R. L. Rivest, The Markov chain tree theorem, *Computer Science Technical Report MIT/LCS/TM-249*, Laboratory of Computer Science, MIT, Cambridge, Mass. (1983).
- [14] T. Leighton, R. L. Rivest, Estimating a probability using finite memory, *IEEE Transactions on Information Theory* 32 (6) (1986) 733–742.
- [15] U. G. Rothblum, Computation of the eigenprojection of a nonnegative matrix at its spectral radius, *Mathematical Programming Study* 6 (1976) 188–201.
- [16] P. Chebotarev, Spanning forests of digraphs and limiting probabilities of Markov chains, *Electronic Notes in Discrete Mathematics* 11 (2002) 108–116.
- [17] P. Chebotarev, Forest representations of Markov chain probabilities, unpublished (2004).
- [18] R. P. Agaev, P. Y. Chebotarev, Spanning forests of a digraph and their applications, *Automation and Remote Control* 62 (3) (2001) 443–466.
- [19] P. Y. Chebotarev, E. V. Shamis, On proximity measures for graph vertices, *Automation and Remote Control* 59 (10) (1998) 1443–1459.
- [20] P. Y. Chebotarev, E. V. Shamis, On a duality between metrics and Σ -proximities, *Automation and Remote Control* 59 (4) (1998) 608–612.
- [21] C. D. Meyer, Jr., M. W. Stadelmaier, Singular M-matrices and inverse positivity, *Linear Algebra and its Applications* 22 (1978) 139–156.
- [22] P. Chebotarev, E. Shamis, The forest metrics for graph vertices, *Electronic Notes in Discrete Mathematics* 11 (2002) 98–107.
- [23] H. A. David, *The Method of Paired Comparisons*, 2nd Edition, Griffin, London, 1988.
- [24] H. A. David, D. M. Andrews, Nonparametric methods of ranking from paired comparisons, in: M. A. Fligner, J. S. Verducci (Eds.), *Probability Models and Statistical Analyses for Ranking Data*, Springer, New York, 1993, pp. 20–36.
- [25] W. D. Cook, M. Kress, *Ordinal Information and Preference Structures: Decision Models and Applications*, Prentice-Hall, Englewood Cliffs, 1992.
- [26] A. R. Belkin, M. S. Levin, Prinyatie reshenii: kombinatornye modeli approksimatsii informatsii (*Decision Making: Combinatorial Models of Information Approximation*), Nauka, Moscow, 1990, in Russian.
- [27] P. Y. Chebotarev, E. Shamis, Constructing an objective function for aggregating incomplete preferences, in: A. Tangian, J. Gruber (Eds.), *Econometric Decision Models, Lecture Notes in Economics and Mathematical Systems*, Vol. 453, Springer, Berlin, 1997, pp. 100–124.
- [28] P. Y. Chebotarev, E. V. Shamis, Preference fusion when the number of alternatives exceeds two: Indirect scoring procedures, *Journal of the Franklin Institute* 336 (1999) 205–226, erratum, *J. Franklin Inst.*, 1999, vol. 336, pp. 747–748.
- [29] J.-F. Laslier, *Tournament Solutions and Majority Voting*, Springer, Berlin, 1997.
- [30] J. Gonzalez Diaz, R. Hendrickx, E. Lohmann, Paired comparisons analysis: An axiomatic approach to ranking methods, *Social Choice and Welfare*. To appear (2013) 1–31. DOI: 10.1007/s00355-013-0726-2.
- [31] L. Csató, Ranking by pairwise comparisons for Swiss-system tournaments, *Central European Journal of Operations Research*. To appear (2013) 1–21. DOI: 10.1007/s10100-012-0261-8.
- [32] H. E. Daniels, Round-robin tournament scores, *Biometrika* 56 (2) (1969) 295–299.
- [33] J. W. Moon, N. J. Pullman, Tournaments and handicaps, in: *Information Processing 68 (Proc. IFIP Congress, Edinburgh, 1968)*, North-Holland, Amsterdam, 1969, pp. 219–223, vol. 1: Mathematics, Software.
- [34] J. W. Moon, N. J. Pullman, On generalized tournament matrices, *SIAM Review* 12 (1970) 384–399.
- [35] I. A. Ushakov, The problem of choosing the preferable object, *Izv. Akad. Nauk SSSR. Tekhn. Kibernet.* No. 4 (1971) 3–7, in Russian.
- [36] I. A. Ushakov, The problem of choosing the preferred element: An application to sport games, in: R. E. Machol, S. P. Ladany, D. G. Morrison (Eds.), *Management Science in Sports*, North-Holland, Amsterdam, 1976, pp. 153–161.
- [37] K. A. Berman, A graph theoretical approach to handicap ranking of tournaments and paired comparisons, *SIAM Journal on Algebraic and Discrete Methods* 1 (3) (1980) 359–361.
- [38] J. C. Maxwell, *Electricity and Magnetism*, 3rd Edition, Oxford Univ. Press, London/New York, 1892, vol. 1, part. II, p. 409.
- [39] S. R. Caplan, D. Zeilberger, T.L. Hill's graphical method for solving linear equations, *Advances in Applied Mathematics* 3 (1982) 377–383.
- [40] K. A. Berman, Y. Liu, Generalized bicycles, *Discrete Applied Mathematics* 78 (1997) 27–40.
- [41] P. Y. Chebotarev, Generalization of the row sum method for incomplete paired comparisons, *Automation and Remote Control* 50 (8, Part 2) (1989) 1103–1113.
- [42] P. Y. Chebotarev, Aggregation of preferences by the generalized row sum method, *Mathematical Social Sciences* 27 (1994) 293–320.
- [43] E. Shamis, Graph-theoretic interpretation of the generalized row sum method, *Mathematical Social Sciences* 27 (1994) 321–333.
- [44] C. Lenart, A generalized distance in graphs and centered partitions, *SIAM Journal on Discrete Mathematics* 11 (2) (1998) 293–304.
- [45] Y. L. Pavlov, *Random Forests*, VSP, Utrecht, 2000.