Matrices of forests, analysis of networks, and ranking problems

Pavel Chebotarev\textsuperscript{a,}\textsuperscript{*}, Rafig Agaev\textsuperscript{a}

\textsuperscript{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.

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., \cite{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. \cite{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 Cesàro limiting probabilities of Markov chains determined by the network under consideration.

2. Notation and simple facts
2.1. Networks, components, and bases
Suppose that $\Gamma$ is a weighted digraph (= network) without loops, $V(\Gamma) = \{1, \ldots, n\}$, $n > 1$, 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 vertex $i$ to vertex $j$ in $\Gamma$; otherwise $w_{ij}$ is strictly positive.

\textsuperscript{*}Corresponding author. Tel.: +7-495-334-8869; fax: +7-495-420-2016.
E-mail address: upi@ipu.ru; pavel4e@gmail.com.
The column Laplacian matrix of $\Gamma$ is the $n \times n$ matrix $L = L(\Gamma) = (l_{ij})$ with entries $l_{ij} = -w_{ij}$ whenever $j \neq i$ and $l_{ii} = -\sum_{j \neq i} w_{ij}, i = 1, \ldots, n$. This matrix has zero column sums. The column Laplacian matrices are singular $M$-matrices (see, e.g., [3, p. 258]). Their index is 1 [3, Proposition 12].

If $\Gamma'$ is a subgraph of $\Gamma$, then the weight of $\Gamma'$, $w(\Gamma')$, 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

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

A spanning subgraph of $\Gamma$ 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 $v_0, e_1, v_1, \ldots, e_k, v_k$ 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 $\Gamma$ if $v = z$ or $\Gamma$ 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 $\Gamma$ is a strong component, or a bicomponent (resp., a weak component) of $\Gamma$. Let $\Gamma_1, \ldots, \Gamma_r$ be all the strong components of $\Gamma$. The condensation (or factorgraph, or leaf composition, or Hertz graph) $\Gamma'$ of digraph $\Gamma$ is the digraph with vertex set $\{\Gamma_1, \ldots, \Gamma_r\}$, where arc $(\Gamma_i, \Gamma_j)$ belongs to $E(\Gamma')$ if $E(\Gamma)$ contains at least one arc from a vertex of $\Gamma_i$ to a vertex of $\Gamma_j$. The condensation of any digraph $\Gamma$ obviously contains no circuits.

A vertex basis of a digraph $\Gamma$ is any minimal (by inclusion) collection of vertices such that every vertex of $\Gamma$ 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 [4, 5]. That is why the bicomponents of $\Gamma$ that correspond to the sources of $\Gamma'$ are called the basic bicomponents [5] or source bicomponents of $\Gamma$. In this paper, the term source knot of $\Gamma$ will stand for the set of vertices of any source bicomponent of $\Gamma$. In [4], source knots are called W-bases.

The following statement [4, 5] characterizes all the vertex bases of a digraph.

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

Schwartz [7] 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. This choice is interpreted as the set of “best” alternatives. A review of choice rules of this kind can be found in [8]; for “fuzzy” extensions, see [9].

### 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 $\Gamma$ will be called out-forests of $\Gamma$; spanning converging forests in $\Gamma$ will be called in-forests of $\Gamma$.

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

\[\text{This union is also called the top cycle and the strong basis of } \Gamma.\]
It is easily seen that every maximum out-forest of $\Gamma$ has the minimum possible number of diverging trees; this number will be called the \textit{out-forest dimension} of $\Gamma$ and denoted by $d'$. It can be easily shown that the number of arcs in any maximum out-forest is $n - d'$; the number of weak components in a forest with $k$ arcs is $n - k$.

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

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

By (2) and (1), $\sigma_k = 0$ whenever $k > n - d'$; $\sigma_0 = 1$. We also introduce the parametric value

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

which is the total weight of out-forests in $\Gamma$ provided that all arc weights are multiplied by $\tau$.

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

$$q_{ij}^k = w(\mathcal{F}_{k}^{i j}). \quad (4)$$

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

The matrix of all out-forests is

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

We will also consider the \textit{stochastic matrices of out-forests}:

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

and the parametric matrices

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

where $\sigma_k$, $\sigma$, and $\sigma(\tau)$ are defined by (2) and (3). $J_k$ are column stochastic by Theorem 1 below.

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

$$\tilde{J} = J_{n-d'}. \quad (8)$$

The matrices of forests can be found by means of matrix analysis [3, Section 5].

3. Properties of the forest matrices

A number of results on the forest matrices are presented in [3]. Some of them are collected in the following theorem.

\textbf{Theorem 1.} [3, 4, 10].

1. Matrices $J_k$, $k = 0, \ldots, n - d'$, $J$, and $J(\tau)$ are column stochastic.
2. The matrix-forest theorem: for any $\tau > 0$, $Q(\tau) = \text{adj}(I + \tau L)$ and $\sigma(\tau) = \det(I + \tau L)$, whence, $J(\tau) = (I + \tau L)^{-1}$.
3. $IJ = JI = 0$.
4. $\tilde{J}$ is idempotent: $\tilde{J}^2 = \tilde{J}$.
5. $\tilde{J} = \lim_{\tau \to 0} J(\tau) = \lim_{\tau \to 0}(I + \tau L)^{-1}$.
6. $\text{rank} \tilde{J} = d'$; $\text{rank} L = n - d'$.
7. $Q_k = \sum_{j=0}^{k} \sigma_{i j}^k (-L)^j, \quad k = 0, 1, \ldots$.
8. $\tilde{J}$ is the eigenprojection of $L$. 

Pavel Chebotarev and Rafig Agaev / Procedia Computer Science 17 (2013) 1134–1141
To formulate the topological properties of the matrix $\tilde{J}$, the following notation is needed.

Let $\tilde{K} = \bigcup_{k=1}^n K_i$, where $K_i$ are all the source knots of $\Gamma$; let $K^*$ 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 $\Gamma$, denote by $\Gamma_K$ the restriction of $\Gamma$ to $K$ and by $\Gamma_{-K}$ 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 $\Gamma_K$, and $\mathcal{F}$ the set of all maximum out-forests of $\Gamma_K$. By $\mathcal{T}^k$, $k \in K$, we denote the subset of $\mathcal{T}$ consisting of all trees that diverge from $k$, and by $\mathcal{F}^{K,-i}$, $j \in V(\Gamma)$, the set of all maximum out-forests of $\Gamma_{-K}$ such that $j$ is reachable from some vertex that belongs to $K$ in these forests. $J_{kk}$ is the $k$th row of $\tilde{J}$.

**Theorem 2.** [11]. Let $K$ be a source knot in $\Gamma$. Then the following statements hold.
1. $\tilde{J}_{ij} \neq 0 \iff (i \in \tilde{K}$ and $j$ is reachable from $i$ in $\Gamma)$.
2. Let $k \in K$. For any $j \in V(\Gamma)$, $\tilde{J}_{kj} = w(\mathcal{T}^k)w(\mathcal{F}^{K,-j})/w(\mathcal{F}^{K,-i})$. 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 $k_1, k_2 \in K$, $\tilde{J}_{k_1} = (w(\mathcal{T}^{k_2})/w(\mathcal{T}^{k_1}))\tilde{J}_{k_1}$ holds, i.e., the rows $k_1$ and $k_2$ of $\tilde{J}$ are proportional.

Thus, knowing $\tilde{J}$ enables one to reveal basic bicomponents and the digraph reachability structure (cf. [11, 12]).

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

$$I - P = \alpha L^\tau,$$

where $\alpha$ is any nonzero real number.

If a Markov chain inversely corresponds to $\Gamma$, then the probability of transition from $j$ to $i \neq j$ is proportional to the weight of arc $(i, j)$ in $\Gamma$ 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 Cesàro 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 [13] and rediscovered in [14, 15], which, in turn, can be immediately proved using item 8 of Theorem 1 and a result of [16] (see [12]). Another proof of Theorem 3 can be found in [17]. A review on forest representations of Markov chain probabilities is given in [18]. For an interpretation of $J(\tau)$ in terms of Markov chains we refer to [19]. Theorem 3 will be used in Section 5.

4. Forest based accessibility measures

Formally, by an accessibility measure for digraph vertices we mean any function that assigns a matrix $P = (p_{ij})_{i,j,\in V(\Gamma)}$ to every weighted digraph $\Gamma$, 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 $P^{\text{out}}_\tau = J(\tau)$, where $J(\tau)$ is defined by (7), and $P^{\text{in}}_\tau = (p^{\text{in}}_{ij})$ with $p^{\text{in}}_{ij} = w(\mathcal{F}^{i \rightarrow j}(\tau))/w(\mathcal{F}^{i \rightarrow j}(\tau))$, where $\mathcal{F}^{i \rightarrow j}(\tau)$ and $\mathcal{F}^{i \rightarrow j}(\tau)$ are, respectively, the $\mathcal{F}^{i \rightarrow j}$ and $\mathcal{F}^{i \rightarrow j}$ for the digraph $\Gamma(\tau)$ obtained from $\Gamma$ by the multiplication of all arc weights by $\tau$. Parameter $\tau$ specifies the relative weight of short and long connections in $\Gamma$.

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

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

**Proposition 2.** For every $\tau > 0$, the measures $P^{\text{out}}_\tau$ and $P^{\text{in}}_\tau$ are dual.
What is the difference in interpretation between $P_{\text{out}}^\tau$ and $P_{\tau}^\text{in}$? A partial answer is as follows. $P_{\text{out}}^\tau$ can be interpreted as the relative weight of $i \rightarrow j$ connections among the out-connections of $i$, whereas $P_{\tau}^\text{in}$ 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 significant for the average man. This example demonstrates that self-duality is not an imperative requirement to accessibility measures. Still, self-dual measures are quite common. Their properties have been studied in [20].

The following conditions some of which were proposed in [20] can be considered as desirable properties of vertex accessibility measures.

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

**Reachability condition.** For any $i, j \in V(\Gamma)$, $(p_{ij} = 0 \Leftrightarrow j$ 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.** If $i, k, t \in V(\Gamma)$, then (A) $p_{kt} - p_{ti} \leq p_{kt} - p_{it}$ and (B) $p_{ik} - p_{it} \leq p_{ik} - p_{kt}$.

The triangle inequalities for proximities is a counterpart of the ordinary triangle inequality which characterizes distances (cf. [21]).

Once $k$ is a cutpoint between $i$ and $t$ if $\Gamma$ contains a path from $i$ to $t$, $i \neq k \neq t$, and every path from $i$ to $t$ includes $k$.

**Transit property.** If $k$ is a cutpoint between $i$ and $t$, then (A) $p_{ik} > p_{it}$ and (B) $p_{it} > p_{ik}$.

**Monotonicity.** Suppose that the weight $\omega_{it}$ of some arc $(k, i)$ is increased or a new $(k, i)$ arc is added to $\Gamma$, and $\Delta p_{ij}$, $i, j \in V(\Gamma)$, are the resulting increments of the accessibility values. Then:
1. (1) $\Delta p_{ii} > 0$;
2. (2) If $i$ is a cutpoint between $k$ and $i$, then $\Delta p_{ik} > \Delta p_{ii}$; if $k$ is a cutpoint between $i$ and $t$ then $\Delta p_{it} > \Delta p_{ik}$;
3. (3) (A) If $i$ is a cutpoint between $k$ and $i$, then $\Delta p_{it} > \Delta p_{kt}$;
   (B) If $k$ is a cutpoint between $i$ and $t$, then $\Delta p_{it} > \Delta p_{it}$.

**Convexity.** (A) If $p_{ij} > p_{ik}$ and $i \neq k$, then there exists a $k$ to $i$ path such that the difference $p_{kj} - p_{ij}$ 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 $p_{jk} - p_{ij}$ strictly increases as $j$ advances from $i$ to $k$ along this path.

The results of testing $P_{\text{out}}^\tau$ and $P_{\tau}^\text{in}$ are collected in

**Theorem 4.** [12]. The measures $P_{\text{out}}^\tau$ and $P_{\tau}^\text{in}$ satisfy all the above conditions not partitioned into (A) and (B). Furthermore, $P_{\text{out}}^\tau$ obeys all (A) conditions and $P_{\tau}^\text{in}$ all (B) conditions.

Consider now the accessibility measures $\tilde{P}_{\text{out}} = (p_{ij}) = \tilde{J} = \lim_{\tau \to \infty} P_{\text{out}}^\tau$ and $\tilde{P}_{\tau}^\text{in} = \lim_{\tau \to \infty} P_{\tau}^\text{in}$. Having in mind Theorem 3 we call $\tilde{J}$ 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 $P_{\text{out}}$ and $\tilde{P}_{\text{in}}$ are dual.

The results of testing $\tilde{P}_{\text{out}}$ and $\tilde{P}_{\text{in}}$ are collected in

**Theorem 5.** The accessibility measures $\tilde{P}_{\text{out}}$ and $\tilde{P}_{\tau}^\text{in}$ satisfy nonnegativity and the “$\geq$” part of reachability condition, but they violate the “$\rightarrow$” part of the reachability condition. Moreover, $\tilde{P}_{\text{out}}$ satisfies, in the nonstrict form, items (A) of self-accessibility condition, transit property, monotonicity, and convexity, whereas $\tilde{P}_{\tau}^\text{in}$ satisfies, in the nonstrict form, items (B) of these conditions. The corresponding strict forms are violated. $\tilde{P}_{\text{out}}$ satisfies (A) and $\tilde{P}_{\tau}^\text{in}$ satisfies (B) of triangle inequality for proximities.

By virtue of Theorem 5, the limiting accessibility measures only “marginally” correspond to the concept of accessibility underlying the above conditions.
Proof. The nonstrict satisfaction of the conditions listed in the theorem follows from Theorem\[1\] Proposition\[3\] and item \(5\) of Theorem\[1\]. To prove that the strict forms of these conditions and the “⇒” part of reachability condition are violated, it suffices to consider the digraph \(\Gamma\) with \(n \geq 3\), \(E(\Gamma) = \{(1, 2), (2, 3)\}\), and \(w_{12} = w_{23} = 1\).

Let us mention one more class of accessibility measures: those of the form \((I + \alpha \tilde{J})^{-1}\), \(0 < \alpha < \sigma_{\max}(\tilde{P}^\text{out})/\sigma_{\min}(\tilde{P}^\text{out})\). These measures are “intermediate” between \(P^\text{out}\) and \(\tilde{P}^\text{out}\), because they are positive linear combination of \(J^{d_1}_{d+1}\) and \(J^{d+1}_{d+1}\). That is why we termed them the matrices of dense out-forests. In the terminology of \[22\] 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 \[19\] (see also \[3\], p. 270–271). Other interesting related topics are the forest distances \[23\] and the forest based centrality measures \[10\].

Thus, we can conclude that forest based accessibility measures may be useful for the analysis of various networks.

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 \[24\]. 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., \[24, 25, 26, 27, 28, 29, 30, 31, 32\]).

In this section, we consider a weighted digraph \(\Gamma\) 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 \[33\], Moon and Pullman \[34, 35\], and Ushakov \[36, 37\] (for its axiomatization, see \[38\]) and reduces to finding nonzero and nonnegative solutions to the system of equations

\[
Lx = 0.
\]

(10)

Entry \(x_i\) of a solution vector \(x = (x_1, \ldots, x_n)^T\) is considered as a sophisticated “score” attached to vertex \(i\). This method was multiply rediscovered with different motivations (some references are given in \[29\]). As Berman \[19\] noticed (although, in other contexts, similar results had been obtained by Maxwell \[40\] and other writers, see \[41\]), if a digraph is strong, then the general solution to (10) is provided by the vectors proportional to \(t = (t_1, \ldots, t_n)^T\), 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., \[44\]), \(t_j\) is the cofactor of any entry in the \(j\)th column of \(L\). Then for every \(i \in V(\Gamma)\), \(\sum_{j=1}^n t_j = \det L\) (the row expansion of \(\det L\)) and, since \(\det L = 0\), \(t\) is a solution to (10). As rank \(L = n - 1\) (since the cofactors of \(L\) are nonzero), any solution to (10) is proportional to \(t\).

Berman \[39\] and Berman and Liu \[42\] 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 \(\Gamma\).

If \(\Gamma\) contains more than one source knot, there is no spanning diverging tree in \(\Gamma\). Recall that \(K_1, \ldots, K_p\) are the source knots of \(\Gamma\), where \(d\) is the out-forest dimension of \(\Gamma\), and \(\bar{K} = \bigcup_{i=1}^p K_i\).

Suppose, without loss of generality, that the vertices of \(\Gamma\) 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 \(V(\Gamma) \setminus \bar{K}\). Such a numeration we call standard.
transmission was successful, then

that the transmission was successful. As a result, if one knows that the information was corrupted at root

References

Then

bases of arbitrary weighted preference relations, incomplete tournaments, irregular pairwise contests, etc.

In our opinion, the latter method can be recommended as a well-grounded approach to scoring objects on

(6)] that produces the score vectors

Corollary 1. (of Theorem 3). For any Markov chain inversely corresponding to \( \Gamma \), \( x = \frac{1}{n} \tilde{J} \cdot (1, \ldots, 1)' \) is the limiting state distribution, provided that the initial state distribution is uniform.

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

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 \( \Gamma \).

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

Then the columns of \( \tilde{J} \) can be considered. A nice interpretation of this vector is given by

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

As a reasonable ultimate score vector, the arithmetic mean \( x = \frac{1}{n} \tilde{J} \cdot (1, \ldots, 1)' \) 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 \( \Gamma \), \( x = \frac{1}{n} \tilde{J} \cdot (1, \ldots, 1)' \) 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 \( \Gamma \) 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 \( \tilde{J} \tau \), instead of \( \tilde{J} \), are free of this feature. On the other hand, both methods violate the self-consistent monotonicity axiom [29], and so do the methods that count the walks between vertices. This axiom is satisfied by the generalized Borda method [43, 44] that produces the score vectors \( J' (\tau) : (\text{od}(1) - \text{id}(1), \ldots, \text{od}(n) - \text{id}(n))' \), where \( J' (\tau) \) is the matrix \( J(\tau) \) of the undirected graph corresponding to \( \Gamma \) [45]. 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.

6. 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, \( i, j \in \text{V}(\Gamma) \), 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 [46] and comparable with certain models of [47].

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] P. Chebotarev, R. Agaev, Forest matrices around the Laplacian matrix, Linear Algebra and its Applications 356 (2002) 253–274.
[4] F. Harary, Graph Theory, Addison-Wesley, Reading, MA, 1969.
[5] A. A. Zykov, Teoriya Konechnykh Grafov (Theory of Finite Graphs), Nauka, Novosibirsk, 1969, in Russian.
[6] M. Fiedler, J. Sedláček, O W-basich orientovaných grafů, Časopis pro Pěstování Matematiky 83 (1958) 214–225.
[7] T. Schwartz, The Logic of Collective Choice, Columbia University Press, New York, 1986.
Y. L. Pavlov, Random Forests, VSP, Utrecht, 2000.

H. A. David, The Method of Paired Comparisons, 2nd Edition, Griffin, 1988.

H. A. David, D. M. Andrews, Nonparametric methods of ranking from paired comparisons, in: M. A. Fligner, J. S. Verducci (Eds.), Econometric Decision Models, Lecture Notes in Economics and Mathematical Systems, Vol. 453, Springer, Berlin, 1997, pp. 100–124.

P. Chebotarev, E. Shamis, The forest metrics for graph vertices, Electronic Notes in Discrete Mathematics 11 (2002) 98–107.

W. D. Cook, M. Kress, Ordinal Information and Preference Structures: Decision Models and Applications, Prentice-Hall, Englewood Cliffs, 1992.

A. R. Belkin, M. S. Levin, Priyatiye reshenii: kombinatornye modeli approksimatsii informatsii (Decision Making: Combinatorial Models of Information Approximation), Nauka, Moscow, 1990, in Russian.

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.

P. Y. Chebotarev, E. 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.

J.-F. Laslier, Tournament Solutions and Majority Voting, Springer, Berlin, 1997.

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/s10737-012-0171-8.

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.

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.

J. W. Moon, N. J. Pullman, On generalized tournament matrices, SIAM Review 12 (1970) 384–399.

I. A. Ushakov, The problem of choosing the preferable object, Izv. Akad. Nauk SSSR. Tekhn. Kibernet. No. 4 (1971) 3–7, in Russian.

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.

G. Slutzki, O. Volij, Ranking participants in generalized tournaments, International Journal of Game Theory 33 (2) (2005) 255–270.

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.

J. C. Maxwell, Electricity and Magnetism, 3rd Edition, Oxford Univ. Press, London/New York, 1892, vol. 1, part II, p. 409.

S. R. Caplan, D. Zeilberger, T.L. Hill’s graphical method for solving linear equations, Advances in Applied Mathematics 3 (1982) 377–383.

K. A. Berman, Y. Liu, Generalized bicycles, Discrete Applied Mathematics 78 (1997) 27–40.

P. Y. Chebotarev, Generalization of the row sum method for incomplete paired comparisons, Automation and Remote Control 50 (8, Part 2) (1989) 1103–1113.

P. Y. Chebotarev, Aggregation of preferences by the generalized row sum method, Mathematical Social Sciences 27 (1994) 293–320.

E. Shamis, Graph-theoretic interpretation of the generalized row sum method, Mathematical Social Sciences 27 (1994) 321–333.

C. Lenart, A generalized distance in graphs and centered partitions, SIAM Journal on Discrete Mathematics 11 (2) (1998) 293–304.

Y. L. Pavlov, Random Forests, VSP, Utrecht, 2000.