Discrete Applied Mathematics 159 (2011) 295–302

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

A class of graph-geodetic distances generalizing the shortest-path and the resistance distances

Pavel Chebotarev *

Institute of Control Sciences of the Russian Academy of Sciences, 65 Profsoyuznaya Street, Moscow 117997, Russia

ARTICLE INFO

Article history:
Received 2 July 2009
Received in revised form 1 October 2010
Accepted 26 November 2010
Available online 21 December 2010

Keywords:
Shortest path distance
Resistance distance
Forest distance
Matrix forest theorem
Spanning rooted forest
Transitional measure
Graph bottleneck identity
Regularized Laplacian kernel

ABSTRACT

A new class of distances for graph vertices is proposed. This class contains parametric families of distances which reduce to the shortest-path, weighted shortest-path, and the resistance distances at the limiting values of the family parameters. The main property of the class is that all distances it comprises are graph-geodetic: $d(i, j) + d(j, k) = d(i, k)$ if and only if every path from i to k passes through j. The construction of the class is based on the matrix forest theorem and the transition inequality.

1. Introduction

The classical distance for graph vertices is the shortest path distance [1]. Another distance, which is almost classical, is the resistance distance [17,18,16], which is proportional to the commute-time distance [21,14,2].

The forest distances $\tilde{d}_\alpha(i, j)$ [6] form a one-parametric family converging to the discrete distance as $\alpha \to 0$ ($\tilde{d}_0(i, j) = 1$ whenever vertices i and j are distinct) and becoming proportional to the resistance distance as $\alpha \to \infty$. The parameter α controls the relative influence of short and long paths connecting two vertices on the distance between them.

In a recent paper [26] (see also [20]), the authors construct a parametric family of graph dissimilarity measures whose extrema are the weighted shortest path distance and the resistance distance. It is noteworthy that in clustering tasks, the best performance is obtained with intermediate values of the family parameter. On the other hand, the corresponding intermediate measures break the triangle inequality, so they need not be distances (in this paper, we use the term "distance" in the sense of a metric space).

Thus, there is a demand in certain applications (these include data analysis, computer science, mathematical chemistry and some others) for a class of graph *distances* whose extreme properties are similar to those of the dissimilarity measures proposed in [26]. Such a class is introduced in this paper. It comprises logarithmically transformed forest distances, and its construction is based on the matrix forest theorem [7] and the transition inequality [3]. The logarithmic transformation not only leads to the shortest-path / weighted shortest-path distance at $\alpha \to 0^+$ and to the resistance distance at $\alpha \to \infty$, but also, for every $\alpha > 0$, it ensures the remarkable *graph-geodetic property*: $d(i, j) + d(j, k) = d(i, k)$ if and only if every path from i to k passes through j.

* Tel.: +7 495 3348869; fax: +7 495 4202016.
E-mail addresses: chv@member.ams.org, Pavel4e@gmail.com.

0166-218X/$ – see front matter
doi:10.1016/j.dam.2010.11.017

We now introduce the necessary notation. Let G be a weighted multigraph (a weighted graph, where multiple edges are allowed) with vertex set $V(G)=\{1,\ldots,n\}$, $n>1$ and edge set $E(G)$. We assume that G has no loops. For $i,j\in V(G)$, let $n_{ij}\in\{0,1,\ldots\}$ be the number of edges incident to both i and j in G; for every $p\in\{1,\ldots,n_{ij}\}$, $w_{ij}^{p}>0$ is the weight of the pth edge of this type; let $w_{ij}=\sum_{p=1}^{n_{ij}}w_{ij}^{p}$ (if $n_{ij}=0$, we set $w_{ij}=0$) and $W=(w_{ij})_{n\times n}$. W is the symmetric *matrix of total edge weights* of G.

A *rooted tree* is a connected and acyclic weighted graph in which one vertex, called the *root*, is marked. A *rooted forest* is a graph, all of whose connected components are rooted trees. The roots of those trees are, by definition, the roots of the rooted forest.

By the weight of a weighted graph H, $w(H)$, we mean the product of the weights of all its edges. If H has no edges, then $w(H)=1$. The weight of a set $\mathcal{S}$ of graphs, $w(\mathcal{S})$, is the total weight of the graphs belonging to $\mathcal{S}$; the weight of the empty set is zero. If the weights of all edges are unity, i.e. the graphs in $\mathcal{S}$ are actually unweighted, then $w(\mathcal{S})$ reduces to the cardinality of $\mathcal{S}$.

For a given weighted multigraph G, by $\mathcal{F}=\mathcal{F}(G)$, $\mathcal{F}_{ij}=\mathcal{F}_{ij}(G)$, and $\mathcal{F}_{ij}^{(p)}=\mathcal{F}_{ij}^{(p)}(G)$ we denote the set of all spanning rooted forests of G, the set of all forests in $\mathcal{F}$ that have vertex i belonging to a tree rooted at j, and the set of all forests in $\mathcal{F}_{ij}$ that have exactly p edges. Let

$$f=w(\mathcal{F}),\qquad f_{ij}=w(\mathcal{F}_{ij}),\quad\text{and}\quad f_{ij}^{(p)}=w(\mathcal{F}_{ij}^{(p)}),\quad i,j\in V(G),\ 0\le p<n; \tag{1}$$

by F we denote the matrix $(f_{ij})_{n\times n}$; F is called the *matrix of forests* of G.

Let $L=(\ell_{ij})$ be the Laplacian matrix of G, i.e.,

$$\ell_{ij}=\begin{cases}-w_{ij}, & j\ne i,\\ \sum_{k\ne i}w_{ik}, & j=i.\end{cases}$$

Consider the matrix

$$Q=(q_{ij})=(I+L)^{-1}.$$

By the matrix forest theorem[1] [8,7,5], for any weighted multigraph G, Q does exist and

$$q_{ij}=\frac{f_{ij}}{f},\quad i,j=1,\ldots,n. \tag{2}$$

Consequently, $F=fQ=f\cdot(I+L)^{-1}$ holds. Q can be considered as a matrix providing a proximity (similarity) measure for the vertices of G [7,4].

By $d^s(i,j)$ we denote the *shortest path distance*,[2] i.e., the number of edges in a shortest path between i and j in G; by $d^r(i,j)$ we denote the *resistance distance* between i and j defined as follows:

$$d^r(i,j)=\ell_{ii}^{+}+\ell_{jj}^{+}-2\ell_{ij}^{+}, \tag{3}$$

where $(\ell_{ij}^{+})_{n\times n}=L^{+}$ is the Moore–Penrose generalized inverse of the Laplacian matrix L of G. $d^r(i,j)$ is equal to the effective resistance between i and j in the resistive network whose line conductances equal the edge weights w_{ij}^{p} in G. If G is connected, then[3]

$$L^{+}=(L+\bar J)^{-1}-\bar J, \tag{4}$$

where $\bar J$ is the $n\times n$ matrix with all entries $\frac{1}{n}$. Furthermore, by [9, Theorem 3]

$$\ell_{ij}^{+}=\frac{f_{ij}^{(n-2)}-\frac{1}{n}f^{(n-2)}}{nt},\quad i,j\in V(G)$$

holds, where $f^{(n-2)}$ is the total weight of spanning rooted forests with $n-2$ edges and t is the total weight of spanning trees in G. By virtue of (3) this yields

Corollary 1 (*To Theorem 3 of [9] and (4)*). *If G is connected, then*

$$d^r(i,j)=x_{ii}+x_{jj}-2x_{ij}=\frac{f_{ii}^{(n-2)}+f_{jj}^{(n-2)}-2f_{ij}^{(n-2)}}{nt},\quad i,j\in V(G), \tag{5}$$

where $(x_{ij})=(L+\bar J)^{-1}$.

The forest representation (5) is a counterpart of the classical 2-tree expression for $d^r(i,j)$ (see, e.g., [22, Theorem 7-4] and [23]); it will be of use in Section 4.

In Section 2 we introduce a new class of intrinsic graph distances and in Sections 3–5 we study its properties.

1 Cf. Theorems 1–3 in [11].

2 The weighted shortest path distance will be considered in Section 5.

3 In fact, for a connected graph, $L^{+}=(L+\alpha\bar J)^{-1}-\alpha^{-1}\bar J$ with any $\alpha\ne 0$ (Propositions 7 and 8 in [9], where the more general case of a multicomponent graph is considered). This expression with $\alpha=n$ is presented in [18, page 88]. For other related references, see Remarks on Proposition 15 in [5].

2. Logarithmic forest distances

Suppose that G is a connected weighted multigraph. Let

$$Q_\alpha = (I + L_\alpha)^{-1}, \tag{6}$$

where α is a real parameter, I is the identity matrix, and L_α is the Laplacian matrix of the multigraph G_α resulting from G by a certain transformation of edge weights. This transformation generally depends on α; for example, if every edge weight is multiplied by $\alpha > 0$, then[4] $L_\alpha = \alpha L$, where L is the Laplacian matrix of G.

Define the matrix H_α as follows:

$$H_\alpha = \gamma\,(\alpha - 1)\,\overrightarrow{\log_\alpha Q_\alpha}, \tag{7}$$

where $\alpha > 0, \alpha \neq 1$, γ is a positive factor, and $\overrightarrow{\varphi(Q_\alpha)}$ with φ being a function stands for elementwise operations, i.e., operations applied to each entry of Q_α separately. Finally, consider

$$D_\alpha = \frac{1}{2}(h_\alpha \mathbf{1}' + \mathbf{1} h_\alpha') - H_\alpha, \tag{8}$$

where h_α is the column vector containing the diagonal entries of H_α, h_α' is the transpose of h_α, $\mathbf{1}$ and $\mathbf{1}'$ being the column of n ones and its transpose. The elementwise form of (8) is: $d_{ij}(\alpha) = \frac{1}{2}(h_{ii}(\alpha) + h_{jj}(\alpha)) - h_{ij}(\alpha),\ i, j = 1, \ldots, n$. This is a standard transformation used to obtain a metric from a symmetric similarity measure (see, e.g., the inverse covariance mapping in [13]). As Theorem 1 below states, D_α determines a metric on the vertices of G.

Since $\lim_{\alpha\to 1}((\alpha - 1)/\ln \alpha) = 1$, we extend Eq. (7) to $\alpha = 1$ as follows:

$$H_1 = \gamma\,\overrightarrow{\ln Q_1}, \tag{9}$$

which preserves continuity. This extension is assumed throughout the paper.

Theorem 1. *For any connected multigraph G and any $\alpha, \gamma > 0$, $D_\alpha = (d_{ij}(\alpha))_{n\times n}$ defined by Eqs. (6)–(9) is a matrix of distances on $V(G)$.*

Before proving Theorem 1 we represent the entries of D_α in terms of the weights of spanning forests in G_α. Let

$$f_{ij}(\alpha) = w(\mathcal{F}_{ij}(G_\alpha)), \quad i, j = 1, \ldots, n. \tag{10}$$

Proposition 1. *For any connected multigraph G and any $\alpha, \gamma > 0$, the matrix $D_\alpha = (d_{ij}(\alpha))$ defined by Eqs. (6)–(9) exists and*

$$d_{ij}(\alpha) = \begin{cases} \gamma\,(\alpha - 1)\log_\alpha \dfrac{\sqrt{f_{ii}(\alpha)\,f_{jj}(\alpha)}}{f_{ij}(\alpha)}, & \alpha \neq 1 \\[2ex] \gamma\,\ln \dfrac{\sqrt{f_{ii}(1)\,f_{jj}(1)}}{f_{ij}(1)}, & \alpha = 1 \end{cases}, \quad i, j = 1, \ldots, n.$$

Proof. Applying the matrix forest theorem (2) to G_α one obtains that Q_α exists and its entries are strictly positive, provided that G is connected. Therefore H_α and D_α also exist.

Let $q_{ij}(\alpha)$ and $h_{ij}(\alpha)$ be the notation for the entries of Q_α and H_α, respectively. For any positive $\alpha \neq 1$ and γ, Eqs. (6)–(8) and the matrix forest theorem (2) imply

$$\begin{aligned} d_{ij}(\alpha) &= \frac{1}{2}(h_{ii}(\alpha) + h_{jj}(\alpha)) - h_{ij}(\alpha) \\ &= \gamma\,(\alpha - 1)\left[\frac{1}{2}(\log_\alpha q_{ii}(\alpha) + \log_\alpha q_{jj}(\alpha)) - \log_\alpha q_{ij}(\alpha)\right] \\ &= \gamma\,(\alpha - 1)\log_\alpha \frac{\sqrt{q_{ii}(\alpha)\,q_{jj}(\alpha)}}{q_{ij}(\alpha)} = \gamma\,(\alpha - 1)\log_\alpha \frac{\sqrt{f_{ii}(\alpha)\,f_{jj}(\alpha)}}{f_{ij}(\alpha)} \end{aligned}$$

for every $i, j = 1, \ldots, n$. If $\alpha = 1$, then the desired expression follows similarly using (9). □

Proof of Theorem 1. Proving this theorem amounts to showing that for every $i, j, k \in V(G)$:

(i) $d_{ij}(\alpha) = 0$ if and only if $i = j$ and
(ii) $d_{ij}(\alpha) + d_{jk}(\alpha) - d_{ki}(\alpha) \geq 0$ (triangle inequality).

---

4 In this case, (6) is called the *regularized Laplacian kernel* of G with diffusion factor α (see [7,25,24]).

Note that the symmetry and non-negativity of D_α (which are sometimes considered as part of the definition of distance) follow from (i) and (ii) by putting $k = j$ and $k = i$ in the triangle inequality.

Let $\alpha \neq 1$. If $i = j$, then by (8), $d_{ij}(\alpha) = 0$. Conversely, if $d_{ij}(\alpha) = 0$, then by Proposition 1, $f_{ii}(\alpha) f_{jj}(\alpha) = (f_{ij}(\alpha))^2$ holds. If $i \neq j$, then $f_{ij}(\alpha) < f_{jj}(\alpha)$, since, by definition, $\mathcal{F}_{ij}(G_\alpha) \subseteq \mathcal{F}_{jj}(G_\alpha)$ and $\mathcal{F}_{jj}(G_\alpha) \smallsetminus \mathcal{F}_{ij}(G_\alpha)$ contains the trivial spanning rooted forest having no edges and weight unity. Since Q_α is symmetric, $f_{ij}(\alpha) < f_{ii}(\alpha)$. Consequently, $i \neq j$ contradicts the assumption $d_{ij}(\alpha) = 0$, hence $i = j$.

To prove (ii), observe that (7), (8), and (2) for any positive $\alpha \neq 1$ imply

$$\begin{aligned} d_{ij}(\alpha) + d_{jk}(\alpha) - d_{ki}(\alpha) &= \frac{1}{2}(h_{ii}(\alpha) + h_{jj}(\alpha) + h_{jj}(\alpha) + h_{kk}(\alpha) - h_{kk}(\alpha) - h_{ii}(\alpha)) - h_{ij}(\alpha) - h_{jk}(\alpha) + h_{ki}(\alpha) \\ &= h_{jj}(\alpha) + h_{ki}(\alpha) - h_{ij}(\alpha) - h_{jk}(\alpha) \\ &= \gamma\,(\alpha - 1)\log_\alpha \frac{f_{jj}(\alpha) f_{ki}(\alpha)}{f_{ij}(\alpha) f_{jk}(\alpha)}. \end{aligned} \tag{11}$$

Since Q_α is symmetric and the matrix $F_\alpha = (f_{ij}(\alpha))_{n\times n}$ determines a transitional measure for G_α [3, item 1 of Corollary 3], we have that[5] $f_{jj}(\alpha) f_{ki}(\alpha) \geq f_{ij}(\alpha) f_{jk}(\alpha)$ (the transition inequality) and so (11) implies that $d_{ij}(\alpha) + d_{jk}(\alpha) - d_{ki}(\alpha) \geq 0$. For $\alpha = 1$ (i) and (ii) are proved similarly. □

Theorem 1 enables us to give the following definition.

Definition 1. Suppose that G is a connected weighted multigraph and $\alpha > 0$. A *logarithmic forest distance with parameter* α *on* G is a function $d_\alpha : V(G) \times V(G) \to \mathbb{R}$ such that $d_\alpha(i, j) = d_{ij}(\alpha)$, where $D_\alpha = (d_{ij}(\alpha))$ is defined by Eqs. (6)–(9).

In Definition 1, the scaling factor γ of (7) and (9) and the transformation $G \to G_\alpha$ are regarded as internal parameters of logarithmic forest distances. In Section 3, we show that all such distances are graph-geodetic. In Sections 4 and 5, logarithmic forest distances with specific γ and $G \to G_\alpha$ transformations and desirable asymptotic properties are considered. Section 5 also contains natural requirements a $G \to G_\alpha$ transformation should satisfy.

3. The logarithmic forest distances are graph-geodetic

The key property of the logarithmic forest distances is that they are graph-geodetic.[6]

Definition 2. For a multigraph G, a function $d : V(G) \times V(G) \to \mathbb{R}$ is *graph-geodetic* whenever for all $i, j, k \in V(G)$, $d(i, j) + d(j, k) = d(i, k)$ holds if and only if every path in G from i to k contains j.

If $d(i, j)$ is a distance on the set of graph vertices, then the property of being graph-geodetic is a natural condition of strengthening the triangle inequality to equality. The shortest path distance clearly possesses the "if" (but not the "only if") part of the graph-geodetic property; the "if" part of this property for the resistance distance is established by Lemma E in [18]. The ordinary distance in a Euclidean space satisfies a similar condition resulting from substituting "line segment" for "path in G" in Definition 2.

Theorem 2. *For any connected multigraph G and any $\alpha > 0$, each logarithmic forest distance $d_\alpha(i, j)$ is graph-geodetic.*

Note that Theorem 2 is not tantamount to item 2 of Corollary 3 in [3], since the construction of logarithmic distances in the present paper differs from that in [3].

Proof. Since $F_\alpha = (f_{ij}(\alpha))_{n\times n}$ is symmetric and determines a transitional measure for G_α [3, item 1 of Corollary 3], we have that $f_{jj}(\alpha) f_{ki}(\alpha) = f_{ij}(\alpha) f_{jk}(\alpha)$ is true if and only if every path in G_α from i to k contains j (the graph bottleneck identity). Owing to (11) and the analogous expression for $\alpha = 1$, this equality is equivalent to $d_\alpha(i, j) + d_\alpha(j, k) - d_\alpha(k, i) = 0$. On the other hand, G_α is constructed in such a way that it shares the set of paths with G. Consequently, $d_\alpha(i, j) + d_\alpha(j, k) - d_\alpha(k, i) = 0$ holds if and only if every path in G from i to k contains j. □

Graph-geodetic functions have many interesting properties. One of them, as mentioned in [18], is a simple connection (such as that obtained in [15]) between the cofactors and the determinant of G's distance matrix and those of the maximal blocks of G that have no cut points. Another example is the recursive Theorem 8 in [19]. Clearly, for a tree, all the $n(n-1)/2$ values of a graph-geodetic distance are determined by the $n - 1$ values corresponding to the pairs of adjacent vertices. The logarithmic forest distances, as well as their limiting cases, i.e., the shortest-path, weighted shortest-path, and resistance distances (see Sections 4 and 5), need not be Euclidean; however, by Blumenthal's "Square-Root" theorem, the corresponding "square-rooted" distances satisfy the 3-Euclidean condition (cf. [19]).

It can be observed that the "ordinary" forest distances [6] defined without the logarithmic transformation (6) are not generally graph-geodetic.

[5] In this proof, we cannot formally apply Theorem 1 of [3] since the construction of logarithmic distances in the present paper has some difference from that in [3].

[6] This term is borrowed from [19].

4. The shortest-path and resistance distances in the framework of logarithmic forest distances

Consider the family of logarithmic forest distances determined by the $G \to G_\alpha$ edge weight transformation

$$w_{ij}^p(\alpha) = \alpha w_{ij}^p, \quad i, j = 1, \dots, n,\ p = 1, \dots, n_{ij} \tag{12}$$

(which implies $L_\alpha = \alpha L$) and the scaling factor

$$\gamma = \ln(e + \alpha^{\frac{2}{n}}), \tag{13}$$

where e is Euler's constant. It turns out that the shortest-path and the resistance distances are the limiting functions of this family.

Proposition 2. *For any connected multigraph G and every $i, j \in V(G)$, $d_\alpha(i, j)$ with $G \to G_\alpha$ transformation (12) and scaling factor (13) converges to the shortest path distance $d^s(i, j)$ as $\alpha \to 0^+$.*

Proof. Denote by m the shortest path distance $d^s(i, j)$ between i and $j \neq i$. Observe that the weight of every forest that belongs to $\mathcal{F}_{ii}(G_\alpha)$ and has at least one edge vanishes with $\alpha \to 0^+$, whereas $\mathcal{F}_{ii}(G_\alpha)$ contains one forest without edges whose weight is unity. Taking this into account and using Proposition 1 and (1) one obtains

$$\lim_{\alpha\to 0^+} d_\alpha(i, j) = \lim_{\alpha\to 0^+} \left(-\log_\alpha \frac{\sqrt{1 \cdot 1}}{\alpha^m (f_{ij}^{(m)} + o(1))} \right),$$

where $o(1) \to 0$ as $\alpha \to 0^+$. Consequently,

$$\lim_{\alpha\to 0^+} d_\alpha(i, j) = \lim_{\alpha\to 0^+} (m + \log_\alpha f_{ij}^{(m)}) = m = d^s(i, j). \quad \square$$

Proposition 3. *For any connected multigraph G and every $i, j \in V(G)$, $d_\alpha(i, j)$ with $G \to G_\alpha$ transformation (12) and scaling factor (13) converges to the resistance distance $d^r(i, j)$ as $\alpha \to \infty$.*

Proof. Observe that for every $i, j \in V(G)$, $f_{ij}^{(n-1)}$ is the total weight of all spanning trees in G. Denote this weight by t; since G is connected, $t > 0$. By Proposition 1 one has

$$\lim_{\alpha\to\infty} d_\alpha(i, j) = \lim_{\alpha\to\infty} \left(\frac{2\alpha}{n} \ln\alpha\, (\ln\alpha)^{-1} \ln \frac{\sqrt{\alpha^{n-1}\left(t + \frac{1}{\alpha} f_{ii}^{(n-2)} + o\left(\frac{1}{\alpha}\right)\right) \alpha^{n-1}\left(t + \frac{1}{\alpha} f_{jj}^{(n-2)} + o\left(\frac{1}{\alpha}\right)\right)}}{\alpha^{n-1}\left(t + \frac{1}{\alpha} f_{ij}^{(n-2)} + o\left(\frac{1}{\alpha}\right)\right)} \right),$$

where $o(\frac{1}{\alpha})$ denotes expressions such that $\alpha \cdot o(\frac{1}{\alpha}) \to 0$ as $\alpha \to \infty$. Hence

$$\lim_{\alpha\to\infty} d_\alpha(i, j) = \frac{2}{n} \lim_{\alpha\to\infty} \ln \frac{\sqrt{\left(1 + \frac{f_{ii}^{(n-2)}}{\alpha t}\right)^\alpha \left(1 + \frac{f_{jj}^{(n-2)}}{\alpha t}\right)^\alpha}}{\left(1 + \frac{f_{ij}^{(n-2)}}{\alpha t}\right)^\alpha} = \frac{2}{n} \ln \frac{\sqrt{\exp\left(\frac{f_{ii}^{(n-2)}}{t}\right) \exp\left(\frac{f_{jj}^{(n-2)}}{t}\right)}}{\exp\left(\frac{f_{ij}^{(n-2)}}{t}\right)}$$
$$= \frac{f_{ii}^{(n-2)} + f_{jj}^{(n-2)} - 2f_{ij}^{(n-2)}}{nt}.$$

Consequently, by Corollary 1 of Section 1, $\lim_{\alpha\to\infty} d_\alpha(i, j) = d^r(i, j)$. $\square$

Note that for logarithmic forest distances with arbitrary positive scaling factors γ, "converges" in Propositions 2 and 3 must be replaced by "becomes proportional".

5. The weighted shortest path distance in the present framework

In the theory of electrical networks, the edge weight w_{ij}^p is interpreted as the conductance, and the Laplacian matrix $L = (\ell_{ij})$ is termed the admittance matrix. The *weighted shortest path distance* $d^{ws}(i, j)$ is defined as follows:[7]

$$d^{ws}(i, j) = \min_\pi \sum_{e\in\pi} r_e,$$

[7] This formula corrects Eq. (6.2) in [18]; cf. the first inequality in [12, p. 261].

where the minimum is taken over all paths π from i to j and the sum is over all edges e in π; $r_{\mathrm{e}} = 1/w_{\mathrm{e}}$ is called the *resistance* of the edge e, where w_{e} is the weight of this edge.

It turns out that the weighted shortest path distance, as well as the ordinary shortest path distance, fits into the framework of logarithmic forest distances. To show this, it suffices to consider the $G \to G_\alpha$ transformation

$$w_{ij}^p(\alpha) = \psi_\alpha(r_{ij}^p), \quad \text{where } r_{ij}^p = 1/w_{ij}^p,\ i,j = 1,\ldots,n,\ p = 1,\ldots,n_{ij}, \tag{14}$$

with

$$\psi_\alpha(r) = \alpha^r. \tag{15}$$

Proposition 4. *For any connected multigraph G and every $i, j \in V(G)$, $d_\alpha(i,j)$ with $G \to G_\alpha$ transformation (14)–(15) converges to the weighted shortest path distance $d^{ws}(i,j)$ as $\alpha \to 0^+$, provided that the scaling factor γ in (7) goes to 1 as $\alpha \to 0^+$.*

Proof. Let G_α be the multigraph with edge weights $\alpha^{r_{ij}^p}$. Using the notation (10), for all vertices i and $j \neq i$, just as in the proof of Proposition 2 we derive

$$\lim_{\alpha\to 0^+} d_\alpha(i,j) = \lim_{\alpha\to 0^+}\left(-\log_\alpha \frac{\sqrt{1\cdot 1}}{f_{ij}(\alpha)}\right) = \lim_{\alpha\to 0^+} \log_\alpha f_{ij}(\alpha).$$

For every $0 < \alpha < 1$,

$$f_{ij}(\alpha) = \sum_{\mathrm{F}\in\mathcal{F}_{ij}(G_\alpha)} w(\mathrm{F}) = \sum_{\mathrm{F}\in\mathcal{F}_{ij}(G)} \prod_{\mathrm{e}\in E(\mathrm{F})} \alpha^{r_{\mathrm{e}}} = \sum_{\mathrm{F}\in\mathcal{F}_{ij}(G)} \alpha^{\sum\limits_{\mathrm{e}\in E(\mathrm{F})} r_{\mathrm{e}}} = \kappa_{ij}(\alpha)\,\alpha^{d^{ws}(i,j)}, \tag{16}$$

where $1 \le \kappa_{ij}(\alpha) \le |\mathcal{F}_{ij}(G)|$. In (16) we use the fact that for every path from i to j, $\mathcal{F}_{ij}(G)$ contains a forest sharing the set of edges with this path. Consequently,

$$\lim_{\alpha\to 0^+} d_\alpha(i,j) = \lim_{\alpha\to 0^+} \log_\alpha\left(\kappa_{ij}(\alpha)\,\alpha^{d^{ws}(i,j)}\right) = d^{ws}(i,j). \quad \square$$

Remark. By definition, G_α results from G by a certain transformation of edge weights. This means that $V(G_\alpha) = V(G)$ and for every $i, j \in V(G)$, G and G_α have the same number of edges incident to both i and j (this fact is used in the proof of Theorem 2). Since the weight of every edge is positive, $\psi_\alpha(r)$ must be positive for every $r > 0$ and every α in the definition domain. Moreover, recall that the edge weight w_{ij}^p is interpreted as the conductance of the corresponding edge, $w_{ij} = \sum_{p=1}^{n_{ij}} w_{ij}^p$, and $w_{ij} = 0 = w_{ij}(\alpha)$ holds if and only if $n_{ij} = 0$. Since the absence of direct ij-connections, i.e., the case of $n_{ij} = 0$, can also be interpreted as the zero conductance of such connections, $w_{ij}(\alpha) = \sum_{p=1}^{n_{ij}} w_{ij}^p(\alpha)$ should be *small* whenever the conductances w_{ij}^p of ij-edges are small (i.e., whenever their resistances are large). Formally, the continuity condition we have just described is stated as follows: $\lim_{r\to\infty} \psi_\alpha(r) = 0$ for every α in the definition domain. Finally, a natural requirement is that $\psi_\alpha(r)$ should be decreasing for every α (this monotonicity condition along with the above limiting condition implies the positivity of $\psi_\alpha(r)$). Note that the transformation (15) satisfies these conditions if and only if $\alpha \in]0,1[$. Furthermore, the edge weight transformations we consider in this paper are such that for each $r > 0$, $\lim_{\alpha\to 0^+} \psi_\alpha(r) = 0$ and $\psi_\alpha(r)$ strictly increases in α; except for $\psi_\alpha(r) = \alpha^r$, these transformations are increasing functions of $\frac{\alpha}{r}$.

Using Propositions 3 and 4 one can easily define a parametric family of logarithmic forest distances converging to the weighted shortest path distance as $\alpha \to 0^+$ and to the resistance distance as $\alpha \to \infty$. Such a family is not unique. Perhaps, the most interesting family with such asymptotic properties is the one described in Proposition 5.

Proposition 5. *For any connected multigraph G and every $i, j \in V(G)$, the logarithmic forest distance $d_\alpha(i,j)$ defined by:*

- *(6) and (8) with (7) replaced by*

$$H_\alpha = \gamma\,\alpha\,\overrightarrow{\ln Q_\alpha}, \tag{17}$$

- *$G \to G_\alpha$ transformation (14) with*

$$\psi_\alpha(r) = \frac{\alpha}{r}\, e^{-\frac{r}{\alpha}}, \quad \textit{and} \tag{18}$$

- *any positive scaling factor $\gamma = \gamma(\alpha)$ such that $\lim_{\alpha\to 0^+}\gamma(\alpha) = 1$ and $\lim_{\alpha\to\infty}\gamma(\alpha) = \frac{2}{n}$*

converges to the weighted shortest path distance $d^{ws}(i,j)$ as $\alpha \to 0^+$ and to the resistance distance $d^{r}(i,j)$ as $\alpha \to \infty$.

Comparing (17) with (7) shows that the family of distances introduced in Proposition 5 is contained in the class of logarithmic forest distances (6)–(9). As a scaling factor in (17) that meets the requirements of Proposition 5, one can take, for example, $\gamma(\alpha) = (\frac{2}{n}\alpha + \beta)/(\alpha + \beta)$, where $\beta > 0$ is a parameter.

Proof. Let G_α be the multigraph with edge weights assigned by (14) and (18). Since the function (18) is asymptotically equivalent to α/r as $\alpha \to \infty$, using Proposition 3 we conclude that for every $i, j \in V(G)$, $\lim_{\alpha\to\infty} d_\alpha(i,j) = d^{r}(i,j)$.

Furthermore, for all vertices i and $j \neq i$ and the distance $d_\alpha(i,j)$ under consideration, similarly to the proof of Proposition 2 we have

$$\lim_{\alpha\to 0^+} d_\alpha(i,j) = \lim_{\alpha\to 0^+}\left(\alpha \ln \frac{\sqrt{1\cdot 1}}{f_{ij}(\alpha)}\right) = \lim_{\alpha\to 0^+}(-\alpha \ln f_{ij}(\alpha)). \tag{19}$$

The definition of the graph weight and Eqs. (10), (14), and (18) yield

$$f_{ij}(\alpha) = \sum_{\mathsf{F}\in\mathcal{F}_{ij}(G_\alpha)} w(\mathsf{F}) = \sum_{\mathsf{F}\in\mathcal{F}_{ij}(G)} \prod_{\mathsf{e}\in E(\mathsf{F})} \frac{\alpha}{r_\mathsf{e}} e^{-r_\mathsf{e}/\alpha} = \sum_{\mathsf{F}\in\mathcal{F}_{ij}(G)} \alpha^{m_\mathsf{F}} w_\mathsf{F} e^{-d_\mathsf{F}/\alpha},$$

where $m_\mathsf{F} = |E(\mathsf{F})|$, $w_\mathsf{F} = \prod_{\mathsf{e}\in E(\mathsf{F})} w_\mathsf{e}$, and $d_\mathsf{F} = \sum_{\mathsf{e}\in E(\mathsf{F})} r_\mathsf{e}$.

Observe that if $\mathsf{F}, \mathsf{F}' \in \mathcal{F}_{ij}(G)$ and

(a) $d_\mathsf{F} < d_{\mathsf{F}'}$ or
(b) $d_\mathsf{F} = d_{\mathsf{F}'}$ and $m_\mathsf{F} < m_{\mathsf{F}'}$ or
(c) $d_\mathsf{F} = d_{\mathsf{F}'}$, $m_\mathsf{F} = m_{\mathsf{F}'}$, and $w_\mathsf{F} > w_{\mathsf{F}'}$,
then for each small enough $\alpha > 0$, $\alpha^{m_\mathsf{F}} w_\mathsf{F} e^{-d_\mathsf{F}/\alpha} > \alpha^{m_{\mathsf{F}'}} w_{\mathsf{F}'} e^{-d_{\mathsf{F}'}/\alpha}$ holds. Consequently, there exists $\alpha_0 > 0$ such that for all $\alpha \in]0, \alpha_0[$ and some $\kappa_{ij}(\alpha)$ satisfying $1 \le \kappa_{ij}(\alpha) \le |\mathcal{F}_{ij}(G)|$,

$$f_{ij}(\alpha) = \kappa_{ij}(\alpha)\,\alpha^{m_{\bar{\mathsf{F}}}}\, w_{\bar{\mathsf{F}}} e^{-d^{ws}(i,j)/\alpha} \tag{20}$$

is true, where $\bar{\mathsf{F}}$ is a forest $\mathsf{F} \in \mathcal{F}_{ij}(G)$ that satisfies (a) or (b) or the nonstrict version (with $w_\mathsf{F} \ge w_{\mathsf{F}'}$) of (c) w.r.t. each $\mathsf{F}' \in \mathcal{F}_{ij}(G)$. Substituting (20) in (19) results in

$$\lim_{\alpha\to 0^+} d_\alpha(i,j) = \lim_{\alpha\to 0^+}(-\alpha(\ln(\kappa_{ij}(\alpha)\, w_{\bar{\mathsf{F}}}) + m_{\bar{\mathsf{F}}}\ln\alpha - d^{ws}(i,j)/\alpha)) = d^{ws}(i,j). \quad \square$$

6. Concluding remarks

Thus, the main property of the logarithmic forest distances introduced by means of Theorem 1 and Proposition 1 is that they are graph-geodetic: $d(i,j) + d(j,k) = d(i,k)$ if and only if every path connecting i and k contains j (Theorem 2).

Three classical distances, namely, the shortest-path, the resistance, and the weighted shortest-path distances, all fit, as limiting cases, into the framework of logarithmic forest distances. The two former distances can be obtained by the use of the edge weight transformation (12), which generates the regularized Laplacian kernel, or, in other words, by putting $\psi_\alpha(r) = \alpha/r$ in (14) (Propositions 2 and 3). To obtain the latter distance, one can put[8] $\psi_\alpha(r) = \alpha^r$ (Proposition 4).

To define a parametric family of logarithmic forest distances whose limiting cases are the weighted shortest path distance and the resistance distance, it suffices to put $\psi_\alpha(r) = \frac{\alpha}{r}\, e^{-\frac{r}{\alpha}}$ in (14) (Proposition 5).

The proofs of Theorems 1 and 2 are based on the fact that the matrix $F = (f_{ij})$ of spanning rooted forests determines a transitional measure [3] on the corresponding multigraph. That is why it can be useful to study the graph-geodetic distances produced by the other transitional measures considered in [3].

We conclude with several remarks.

On intercomponent distances

Throughout the paper, we assumed that G is connected. Otherwise, if G has more than one component and i and j belong to different components, then, by the matrix forest theorem (2), $q_{ij} = f_{ij} = 0$. Consequently, if $\log_\alpha(\cdot)$ and $\ln(\cdot)$ are considered as functions mapping to the extended line $\mathbb{R}\cup\{-\infty, +\infty\}$, then (8) leads to the distance $+\infty$ between i and j, which seems quite natural.

On the parameter α and the length of paths between vertices

The parameter α of logarithmic forest distances controls the relative influence of short, medium, and long paths between vertices i and j on the distance $d_\alpha(i,j)$. As $\alpha \to 0$, only the (weighted) shortest paths matter; the long paths have the maximum effect as $\alpha \to \infty$.

On the "mixture" of the shortest-path and resistance distances

The simplest way of "generalizing" both the (weighted) shortest-path and the resistance distances is to consider the convex combination of the form $d'_\alpha(i,j) = (1-\alpha)d^s(i,j) + \alpha d^r(i,j)$, where $\alpha \in [0,1]$. However, this approach seems quite poor from both theoretical and practical points of view. First, it does not presuppose any underlying model that might provide a deeper insight by unifying the shortest-path and the resistance distances; thus, the mixture seems just "mechanical". Second, consider, for example, a path on four vertices: let $V(G) = \{1,2,3,4\}$ and $E(G) = \{(1,2),(2,3),(3,4)\}$. Then $d^s(1,2) = d^s(2,3) = d^r(1,2) = d^r(2,3) = 1$, and therefore $d'_\alpha(1,2) = d'_\alpha(2,3)$ for all

[8] It can be shown that $\psi_\alpha(r) = e^{-r/\alpha}$ is also suitable for this purpose.

$\alpha \in [0, 1]$. On the other hand, in applications, there are models and intuitive heuristics that result in either $d(1, 2) > d(2, 3)$ or $d(1, 2) < d(2, 3)$. Indeed, suppose that the distance $d(i, j)$ should depend on the whole set of routes between i and j: the shorter and more numerous are the routes, the smaller must be the distance. Then the inequality $d(1, 2) > d(2, 3)$ is suggested by the observation that there are three routes of length 3 between vertices 2 and 3 (namely, (2, 3, 2, 3), (2, 1, 2, 3), and (2, 3, 4, 3)) and only two routes of length 3 between vertices 1 and 2 ((1, 2, 1, 2) and (1, 2, 3, 2)). On the other hand, if the *relative* numbers of routes are important, then the opposite inequality $d(1, 2) < d(2, 3)$ can be justified by the observation that (1, 2) is the unique route of length 1 starting at vertex 1, whereas (2, 3) and (3, 2) are not unique routes of length 1 starting at vertices 2 and 3, respectively. It is worth mentioning that the inequality $d(1, 2) < d(2, 3)$ holds true for the *quasi-Euclidean graph distance* [19].

The above example demonstrates that distances providing $d(1, 2) = d(2, 3)$ are insufficient for the numerous applications of graph theory. As regards the forest distances, the logarithmic forest distances provide $d_\alpha(1, 2) < d_\alpha(2, 3)$, whereas with the "ordinary" forest distances [6], we have $\tilde{d}_\alpha(1, 2) > \tilde{d}_\alpha(2, 3)$.

On some physical and probabilistic interpretations of graph distances

In the view of Chen and Zhang [10], ". . . the shortest-path [distance] might be imagined to be more relevant when there is corpuscular communication (along edges) between two vertices, whereas the resistance distance might be imagined to be more relevant when the communication is wave-like". The authors do not develop this idea in depth; presumably they have in mind that a corpuscle always takes a shortest path between vertices, while a wave takes all paths simultaneously. As has been shown in this paper, the shortest-path, weighted shortest-path, and resistance distances are extreme examples of the logarithmic forest distances. The forest distance between vertices i and j is interpreted as the probability of choosing a forest partition separating i and j in the model of random forest partitions [6, Proposition 5]. As $\alpha \to 0$, transformation (7) preserves only those partitions that connect i and j by a (weighted) shortest path and separate all vertices this path does not involve; thereby the (weighted) shortest path distance results in this case, as we see in Propositions 2, 4 and 5. When $\alpha \to \infty$ and $\psi_\alpha(r) \sim \alpha/r$, this transformation preserves only the partitions determined by two disjoint trees, which leads to the resistance distance, as Propositions 3 and 5 demonstrate.

Acknowledgements

This work was partially supported by RFBR Grant 09-07-00371 and the RAS Presidium Program "Development of Network and Logical Control".

I am grateful to Marco Saerens for discussing with me the problem of generalizing the shortest-path and the resistance distances.

References

[1] F. Buckley, F. Harary, Distance in Graphs, Addison-Wesley, Redwood City, CA, 1990.
[2] A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky, P. Tiwari, The electrical resistance of a graph captures its commute and cover times, in: Proc. 21st Annual ACM Symp. on Theory of Computing, ACM Press, Seattle, 1989, pp. 574–586.
[3] P. Chebotarev, The graph bottleneck identity, Adv. Appl. Math. URL: http://dx.doi.org/10.1016/j.aam.2010.11.001 (in press).
[4] P. Chebotarev, Spanning forests and the golden ratio, Discrete Appl. Math. 156 (2008) 813–821.
[5] P. Chebotarev, R. Agaev, Forest matrices around the Laplacian matrix, Linear Algebr Appl. 356 (2002) 253–274.
[6] P. Chebotarev, E. Shamis, The forest metrics for graph vertices, Electron. Notes Discrete Math. 11 (2002) 98–107.
[7] P.Yu. Chebotarev, E.V. Shamis, The matrix-forest theorem and measuring relations in small social groups, Autom. Remote Control 58 (1997) 1505–1514.
[8] P.Yu. Chebotarev, E. Shamis, On the proximity measure for graph vertices provided by the inverse Laplacian characteristic matrix, in: Abstracts of the Conference "Linear Algebra and its Applications", 10–12 July, 1995, University of Manchester, Manchester, UK, 1995, pp. 6–7. URL: http://www.ma.man.ac.uk/~higham/laa95/abstracts.ps.
[9] P.Yu. Chebotarev, E.V. Shamis, On proximity measures for graph vertices, Autom. Remote Control 59 (1998) 1443–1459.
[10] H. Chen, F. Zhang, Resistance distance and the normalized Laplacian spectrum, Discrete Appl. Math. 155 (2007) 654–661.
[11] F. Chung, W. Zhao, PageRank and random walks on graphs, in: Proceedings of "Fete of Combinatorics and Computer Science" Conference in honor of Laci Lovász, Keszthely, Hungary, August 11–15, 2008 (in press).
[12] M.M. Deza, E. Deza, Encyclopedia of Distances, Springer, Berlin, Heidelberg, 2009.
[13] M.M. Deza, M. Laurent, Geometry of Cuts and Metrics, Springer, Berlin, 1997.
[14] F. Göbel, A.A. Jagers, Random walks on graphs, Stochastic Process. Appl. 2 (1974) 311–336.
[15] R.L. Graham, A.J. Hoffman, H. Hosoya, On the distance matrix of a directed graph, J. Graph Theory 1 (1977) 85–88.
[16] V. Gurvich, Metric and ultrametric spaces of resistances, Discrete Appl. Math. 158 (2010) 1496–1505.
[17] A.D. Gvishiani, V.A. Gurvich, Metric and ultrametric spaces of resistances, Russian Math. Surveys 42 (1987) 235–236.
[18] D.J. Klein, M. Randić, Resistance distance, J. Math. Chem. 12 (1993) 81–95.
[19] D.J. Klein, H.Y. Zhu, Distances and volumina for graphs, J. Math. Chem. 23 (1998) 179–195.
[20] A. Mantrach, L. Yen, J. Callut, K. Françoisse, M. Shimbo, M. Saerens, The Sum-over-paths covariance kernel: a novel covariance measure between nodes of a directed graph, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1112–1126.
[21] C.St.J.A. Nash-Williams, Random walk and electric currents in networks, Math. Proc. Cambridge Philos. Soc. 55 (1959) 181–194.
[22] S. Seshu, M.B. Reed, Linear Graphs and Electrical Networks, Addison-Wesley, Reading, MA, 1961.
[23] L.W. Shapiro, An electrical lemma, Math. Mag. 60 (1987) 36–38.
[24] M. Shimbo, T. Ito, D. Mochihashi, Yu. Matsumoto, On the properties of von Neumann kernels for link analysis, Mach. Learn. 75 (2009) 37–67.
[25] A.J. Smola, R.I. Kondor, Kernels and regularization of graphs, in: Proc. 16th Annual Conf. on Learning Theory, 2003, pp. 144–158.
[26] L. Yen, M. Saerens, A. Mantrach, M. Shimbo, A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances, in: 14th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining, 2008, pp. 785–793.