# Simple expressions for the long walk distance

Pavel Chebotarev [a,*,1], R.B. Bapat [b,2], R. Balaji [c]

[a] Institute of Control Sciences of the Russian Academy of Sciences, 65 Profsoyuznaya Street, Moscow 117997, Russia
[b] Indian Statistical Institute, New Delhi 110 016, India
[c] Department of Mathematics, Indian Institute of Technology-Madras, Chennai 600 036, India

ARTICLE INFO

ABSTRACT

The walk distances in graphs are defined as the result of appropriate transformations of the $\sum_{k=0}^{\infty}(tA)^k$ proximity measures, where $A$ is the weighted adjacency matrix of a connected weighted graph and $t$ is a sufficiently small positive parameter. The walk distances are graph-geodetic, moreover, they converge to the shortest path distance and to the so-called long walk distance as the parameter $t$ approaches its limiting values. In this paper, simple expressions for the long walk distance are obtained. They involve the generalized inverse, minors, and inverses of submatrices of the symmetric irreducible singular M-matrix $\mathcal{L} = \rho I - A$, where $\rho$ is the Perron root of $A$.

© 2012 Elsevier Inc. All rights reserved.

## 1. Introduction

The walk distances for graph vertices are a parametric family of graph distances proposed in [4]. Along with their modifications they generalize [5] the logarithmic forest distances [3], resistance distance, shortest path distance, and the weighted shortest path distance. The walk distances are

---

* Corresponding author.
   E-mail addresses: chv@member.ams.org (P. Chebotarev), rbb@isid.ac.in (R.B. Bapat), balaji5@iitm.ac.in (R. Balaji).

graph-geodetic: for a distance [3] $d(i, j)$ on the vertices of a graph $G$ this means that $d(i, j) + d(j, k) = d(i, k)$ if and only if every path in $G$ connecting $i$ and $k$ visits $j$.

The long walk distance, $d^{\text{LW}}(i, j)$, is obtained by letting the parameter of the walk distances go to one of its limiting values (approaching the other limiting value leads to the shortest path distance). A number of expressions for $d^{\text{LW}}(i, j)$ are given in [5]. In this paper, we obtain simple expressions in terms of the matrix $\mathcal{L} = \rho I - A$, where $A$ is the weighted adjacency matrix of a connected weighted graph $G$ on $n$ vertices and $\rho$ is the Perron root of $A$. $\mathcal{L}$ is a symmetric irreducible singular M-matrix, so rank $\mathcal{L} = n - 1$ and $\mathcal{L}$ is positive semidefinite. In [5], $\mathcal{L}$ was called the *para-Laplacian matrix* of $G$. The expressions presented in this paper generalize some classical expressions for the resistance distance (cf. [2]). They enable one to conclude that the long walk distance can be considered as the counterpart of the resistance distance obtained by replacing the Laplacian matrix $L = \text{diag}(A\mathbf{1}) - A$ and the vector $\mathbf{1}$ (of $n$ ones) which spans Ker $L$ with the para-Laplacian matrix $\mathcal{L}$ and the eigenvector of $A$ spanning Ker $\mathcal{L}$. If $A$ has constant row sums, then $\mathcal{L} = L$ and these distances coincide.

## 2. Notation

In the graph definitions we mainly follow [7]. Let $G$ be a weighted multigraph (a weighted graph where multiple edges are allowed) with vertex set $V(G) = V$, $|V| = n > 1$, and edge set $E(G)$. Loops are allowed; throughout the paper we assume that $G$ is connected. For brevity, we will call $G$ a *graph*. For $i, j \in V$, let $n_{ij} \in \{0, 1, \ldots\}$ be the number of edges incident to both $i$ and $j$ in $G$; for every $q \in \{1, \ldots, n_{ij}\}$, $w_{ij}^q > 0$ is the weight of the $q$th edge of this type. Let

$$a_{ij} = \sum_{q=1}^{n_{ij}} w_{ij}^q$$

(if $n_{ij} = 0$, we set $a_{ij} = 0$) and $A = (a_{ij})_{n \times n}$; $A$ is the symmetric *weighted adjacency matrix* of $G$. In this paper, all matrix entries are indexed by the vertices of $G$.

The *weighted Laplacian matrix* of $G$ is $L = \text{diag}(A\mathbf{1}) - A$, where $\mathbf{1}$ is the vector of $n$ ones.

For $v_0, v_m \in V$, a $v_0 \rightarrow v_m$ path (*simple path*) in $G$ is an alternating sequence of vertices and edges $v_0, e_1, v_1, \ldots, e_m, v_m$ where all vertices are distinct and each $e_i$ is an edge incident to $v_{i-1}$ and $v_i$. The unique $v_0 \rightarrow v_0$ path is the "sequence" $v_0$ having no edges.

Similarly, a $v_0 \rightarrow v_m$ walk in $G$ is an *arbitrary* alternating sequence of vertices and edges $v_0, e_1, v_1, \ldots, e_m, v_m$ where each $e_i$ is incident to $v_{i-1}$ and $v_i$. The *length* of a walk is the number $m$ of its edges (including loops and repeated edges). The *weight* of a walk is the product of the $m$ weights of its edges. The weight of a set of walks is the sum of the weights of its elements. By definition, for any vertex $v_0$, there is one $v_0 \rightarrow v_0$ walk $v_0$ with length 0 and weight 1.

Let $r_{ij}$ be the weight of the set $\mathcal{R}^{ij}$ of all $i \rightarrow j$ walks in $G$, provided that this weight is finite. $R = R(G) = (r_{ij})_{n \times n}$ will be referred to as the *matrix of the walk weights*.

By $d^{\text{r}}(i, j)$ we denote the *resistance distance* between $i$ and $j$, i.e., the effective resistance between $i$ and $j$ in the resistive network whose line conductances equal the edge weights $w_{ij}^q$ in $G$. The resistance distance has several expressions via the generalized inverse, minors, and inverses of the submatrices of the weighted Laplacian matrix of $G$ (see, e.g., [9,1] or the papers by Subak-Sharpe and Styan published in the 60s and cited in [5]).

**Definition 1.** For a multigraph $G$ with vertex set $V$, a function $d : V \times V \rightarrow \mathbb{R}$ is called *graph-geodetic* provided that $d(i, j) + d(j, k) = d(i, k)$ holds if and only if every path in $G$ connecting $i$ and $k$ contains $j$.

Graph-geodetic functions can also be called *bottleneck additive* or *cutpoint additive*.

---

[3]  In this paper, a *distance* is assumed to satisfy the axioms of metric.

## 3. The walk distances

For any $t > 0$, consider the graph $tG$ obtained from $G$ by multiplying all edge weights by $t$. If the matrix of the walk weights of $tG$, $R_t = R(tG) = (r_{ij}(t))_{n \times n}$, exists, then[4]

$$R_t = \sum_{k=0}^{\infty} (tA)^k = (I - tA)^{-1},$$

where $I$ denotes the identity matrix of appropriate dimension.

By assumption, $G$ is connected, while its edge weights are positive, so $R_t$ is positive whenever it exists. The walk distances can be introduced as follows.

**Definition 2.** For a connected graph $G$, the *walk distances* on $V(G)$ are the functions $d_t(i, j) : V(G) \times V(G) \to \mathbb{R}$ and the functions positively proportional to them, where

$$d_t(i, j) = -\ln \left( \frac{r_{ij}(t)}{\sqrt{r_{ii}(t)\, r_{jj}(t)}} \right). \tag{1}$$

**Lemma 1** [4]**.** *For any connected $G$, if $R_t = (r_{ij}(t))$ exists, then* (1) *determines a graph-geodetic distance on $V(G)$.*

Regarding the existence (finiteness) of $R_t$, since $G$ is connected, $A$ is irreducible, so the Perron–Frobenius theory of nonnegative matrices provides the following result.

**Lemma 2.** *For any weighted adjacency matrix $A$ of a connected graph $G$, the series $R_t = \sum_{k=0}^{\infty}(tA)^k$ with $t > 0$ converges to $(I - tA)^{-1}$ if and only if $t < \rho^{-1}$, where $\rho = \rho(A)$ is the spectral radius of $A$. Moreover, $\rho$ is an eigenvalue of $A$; as such $\rho$ has multiplicity 1 and a positive eigenvector.*

Thus, the walk distance $d_t(i, j)$ with parameter $t$ exists if and only if $0 < t < \rho^{-1}$.

A topological interpretation of the walk distances in terms of the weights of routes and circuits in $G$ is given in [6].

## 4. The long walk distance

The *long walk distance* is defined [5] as the limit of the walk distance (1) as $t \to (\rho^{-1})^-$ with a scaling factor depending on $t$, $\rho$, and $n$. Namely,

$$\begin{aligned}
d^{\mathrm{LW}}(i, j) &\overset{\text{def}}{=} \lim_{t \to (\rho^{-1})^-} \frac{2d_t(i, j)}{n\rho^2(\rho^{-1} - t)} \\
&= \lim_{t \to (\rho^{-1})^-} \frac{\ln r_{ii}(t) + \ln r_{jj}(t) - 2\ln r_{ij}(t)}{n\rho^2(\rho^{-1} - t)}, \quad i, j \in V.
\end{aligned}$$

It is shown [5] that this limit always exists and defines a graph-geodetic metric.

There is a two-way connection between the long walk distance and the resistance distance. In particular, the following theorem holds.

**Theorem 1** [5]**.** *Let $G$ be a connected graph on $n$ vertices with weighted adjacency matrix $A$; let $p = (p_1, \dots, p_n)^{\mathrm{T}}$ be the Perron vector of $A$. Suppose that $p' = (p_1', \dots, p_n')^{\mathrm{T}} = \sqrt{n}p/\|p\|_2$, where*

---

[4] In the more general case of weighted *digraphs*, the $ij$-entry of the matrix $R_t - I$ is called the *Katz similarity* between vertices $i$ and $j$. Katz [8] proposed it to evaluate the social status taking into account all $i \to j$ paths.

$\|p\|_2 = (\sum_{i=1}^n p_i^2)^{1/2}$, and $P' = \operatorname{diag} p'$. Then the long walk distance in G coincides with the resistance distance in the graph $G'$ whose weighted adjacency matrix is $P'AP'$.

Using Theorem 1 and two classical expressions for the resistance distance one has:

**Corollary 1** [5]. *In the notation of Theorem 1,*

$$d^{\mathrm{LW}}(i,j) = \frac{\det (L'_{\overline{ii}})_{\bar{j}\bar{j}}}{\det L'_{\overline{vv}}}, \quad j \neq i, \quad \forall v \in V,$$

$$d^{\mathrm{LW}}(i,j) = \ell'^{-}_{ii} + \ell'^{-}_{jj} - 2\ell'^{-}_{ij},$$

*where $L' = P'\mathcal{L}P'$, $\mathcal{L} = \rho I - A$, $X_{\overline{uv}}$ is X with row u and column v removed, and $L'^{-} = (\ell'^{-}_{ij})$ is any g-inverse[5] of $L'$.*

Note that by Lemma 13 in [5], $L'$ is the weighted Laplacian matrix of the connected graph $G'$ introduced in Theorem 1. Therefore, the Matrix tree theorem implies that $\det L'_{\overline{vv}}$, the denominator of the above determinant formula for $d^{\mathrm{LW}}(i,j)$, is strictly positive.

Now we obtain simple expressions for $d^{\mathrm{LW}}(i,j)$ in terms of the matrix $\mathcal{L} = \rho I - A$.

## 5. Simple expressions for the long walk distance

**Theorem 2.** *In the notation of Theorem 1 and Corollary 1, for all $i,j \in V$ such that $j \neq i$,*

$$d^{\mathrm{LW}}(i,j) = \frac{\det (\mathcal{L}_{\overline{ii}})_{\bar{j}\bar{j}}}{p_j'^2 \det \mathcal{L}_{\overline{ii}}}, \tag{2}$$

$$d^{\mathrm{LW}}(i,j) = z^{\mathrm{T}}(i,j)\mathcal{L}^{-} z(i,j), \tag{3}$$

$$d^{\mathrm{LW}}(i,j) = z_{\overline{u}}^{\mathrm{T}}(i,j)(\mathcal{L}_{\overline{vu}})^{-1} z_{\overline{v}}(i,j), \quad \forall u,v \in V, \tag{4}$$

*where $\mathcal{L}^{-}$ is any g-inverse of $\mathcal{L}$ and $z(i,j)$ is the n-vector whose ith element is $1/p_i'$, jth element is $-1/p_j'$, and the other elements are 0.*

**Proof.** By virtue of Corollary 1, for $j \neq i$ we have

$$d^{\mathrm{LW}}(i,j) = \frac{\det ((P'\mathcal{L}P')_{\overline{ii}})_{\bar{j}\bar{j}}}{\det (P'\mathcal{L}P')_{\overline{ii}}} = \frac{(\prod_{k \notin \{i,j\}} p_k')^2 \det (\mathcal{L}_{\overline{ii}})_{\bar{j}\bar{j}}}{(\prod_{k \neq i} p_k')^2 \det \mathcal{L}_{\overline{ii}}} = \frac{\det (\mathcal{L}_{\overline{ii}})_{\bar{j}\bar{j}}}{p_j'^2 \det \mathcal{L}_{\overline{ii}}},$$

that is, (2) holds. To prove (3), consider the matrix[6] $H = (h_{ij}) \in \mathbb{R}^{n \times n}$ such that

$$h_{ij} = \begin{cases} -\dfrac{p_i' \det (\mathcal{L}_{\overline{ii}})_{\bar{j}\bar{j}}}{2p_j' \det \mathcal{L}_{\overline{ii}}}, & j \neq i, \\ 0, & j = i. \end{cases} \tag{5}$$

**Lemma 3.** *H is a g-inverse of $\mathcal{L}$: $\mathcal{L}H\mathcal{L} = \mathcal{L}$.*

**Proof.** Let $D_{G'}^{\mathrm{r}}$ be the matrix of resistance distances in the graph $G'$ introduced in Theorem 1. Then, by [12, Theorem 10.1.4],

$$D_{G'}^{\mathrm{r}} = \ell_{G'}^{-} \mathbf{1}^{\mathrm{T}} + \mathbf{1}(\ell_{G'}^{-})^{\mathrm{T}} - 2L_{G'}^{-}, \tag{6}$$

---

[5]  A matrix Z is a *g-inverse* of X whenever $X = XZX$. A simple choice of $L'^{-}$ is $(L' + \frac{1}{n}\mathbf{11}^{\mathrm{T}})^{-1}$ [12, Theorem 10.1.4].

[6]  H generalizes the *zero-axial matrix Z* studied in [13].

where $L_{G'}^-$ is any g-inverse of $L_{G'}$, $L_{G'}$ is the weighted Laplacian matrix of $G'$, and $\ell_{G'}^-$ is the column vector containing the diagonal entries of $L_{G'}^-$. As $L_{G'}\mathbf{1} = \mathbf{0}$ and $\mathbf{1}^{\mathrm{T}}L_{G'} = \mathbf{0}^{\mathrm{T}}$, (6) implies that

$$L_{G'}(-\tfrac{1}{2}D_{G'}^{\mathrm{r}})L_{G'} = L_{G'}.$$

Since by Lemma 13 in [5], $L_{G'} = P'\mathcal{L}P'$ and by Theorem 1, $D_{G'}^{\mathrm{r}} = D_G^{\mathrm{LW}}$ (where $D_G^{\mathrm{LW}}$ is the matrix of long walk distances in $G$), we have

$$P'\mathcal{L}P'(-\tfrac{1}{2}D_G^{\mathrm{LW}})P'\mathcal{L}P' = P'\mathcal{L}P'. \tag{7}$$

It follows from (2) and (5) that

$$H = P'(-\tfrac{1}{2}D_G^{\mathrm{LW}})P',$$

moreover, $P'$ is invertible. Thus, (7) yields $\mathcal{L}H\mathcal{L} = \mathcal{L}$, as desired. $\square$

By (2) and (5),

$$d^{\mathrm{LW}}(i,j) = -2\frac{h_{ij}}{p_i'p_j'} = z^{\mathrm{T}}(i,j)\,Hz(i,j), \quad j \neq i, \tag{8}$$

where $z(i,j)$ is the $n$-vector whose $i$th element is $1/p_i'$, $j$th element is $-1/p_j'$, and the other elements are 0.

Let $i$ and $j \neq i$ be fixed and let $z = z(i,j)$. Lemma 2 implies that 0 is a simple eigenvalue of $\mathcal{L}$, while $p$ spans $\mathrm{Ker}\,\mathcal{L}$. Hence, rank $\mathcal{L} = n - 1$ and the range of $\mathcal{L}$ is $p^{\perp}$, the orthogonal complement of $p$. Since $z \in p^{\perp}$, it follows that $z = \mathcal{L}x$ for some $n$-vector $x$. Consequently, if $\mathcal{L}^-$ is a g-inverse of $\mathcal{L}$, then

$$z^{\mathrm{T}}(i,j)\mathcal{L}^-z(i,j) = x^{\mathrm{T}}\mathcal{L}\mathcal{L}^-\mathcal{L}x = x^{\mathrm{T}}\mathcal{L}x = z^{\mathrm{T}}(i,j)Hz(i,j) = d^{\mathrm{LW}}(i,j),$$

where the last two transitions follow from Lemma 3 and Eq. (8). This proves (3).

Since $\mathcal{L}$ is a singular M-matrix, the principal minors of $\mathcal{L}$ are strictly positive. By [5, Lemma 7] (which follows from Theorem 3.1 in [10]) $l^{vu}/l^{vv} = p_u/p_v$, where $l^{vu}$ is the cofactor of the entry $l_{vu}$ of $\mathcal{L}$. Hence, $\mathcal{L}_{\overline{vu}}$ is non-singular. In view of (3), proving (4) reduces to showing that

$$\mathcal{L}\mathcal{L}^{\sim}_{(uv)}\mathcal{L} = \mathcal{L}, \tag{9}$$

where $\mathcal{L}^{\sim}_{(uv)} \in \mathbb{R}^{n \times n}$ is the matrix that has all zeros in row $u$ and column $v$ and satisfies $(\mathcal{L}^{\sim}_{(uv)})_{\overline{uv}} = (\mathcal{L}_{\overline{vu}})^{-1}$. Eq. (9) can be derived from the general result [12, Section 11.2]. Alternatively, the multiplication of block matrices verifies that for $(l_{km}) = \mathcal{L}$ and $\mathcal{L}' = (l'_{km}) \overset{\text{def}}{=} \mathcal{L}\mathcal{L}^{\sim}_{(uv)}\mathcal{L}$ it holds that $\mathcal{L}'_{\overline{uv}} = \mathcal{L}_{\overline{uv}}$, $l'_{u\overline{v}} = l_{u\overline{v}}$, $l'_{\overline{u}v} = l_{\overline{u}v}$, and $l'_{uv} = l_{u\overline{v}}(\mathcal{L}_{\overline{uv}})^{-1}l_{\overline{u}v}$, where $l_{u\overline{v}}$ and $l_{\overline{u}v}$ are $\mathcal{L}$'s row $u$ and column $v$ with $l_{uv}$ removed, respectively. Now to prove (9) it remains to show that

$$l_{u\overline{v}}(\mathcal{L}_{\overline{uv}})^{-1}l_{\overline{u}v} = l_{uv}, \tag{10}$$

which is true due to the following lemma.

**Lemma 4.** *If* $X = (x_{km})_{n \times n}$ *is a singular matrix, while* $X_{\overline{uv}}$ *is non-singular, then*

$$x_{uv} = x_{u\overline{v}}\,(X_{\overline{uv}})^{-1}x_{\overline{u}v},$$

*where* $x_{u\overline{v}}$ *and* $x_{\overline{u}v}$ *are X's row u and column v with* $x_{uv}$ *removed, respectively.*

Lemma 4 is a special case of [11, Eq. (6.2.1)] and can alternatively be proved by first expanding det $X$ along row $u$ and then expanding the cofactors of the entries in row $u$ (except for $x_{uv}$) along column $v$. Lemma 4 verifies (10) and thus, (9) and (4). $\square$

In conclusion, it is worth introducing a rescaled version of the long walk distance, $\tilde{d}^{\mathrm{LW}}(i, j) \overset{\text{def}}{=} n\|p\|_2^2\, d^{\mathrm{LW}}(i, j)$, where, as before, $p$ is the Perron vector of $A$ and $n$ is the number of vertices. It has the following desirable property: if for a sequence of connected graphs $(G^k)$ on $n$ vertices, we have [7] $a_{12}^k = a_{21}^k = 1,\ k = 1, 2, \ldots$, whereas all other entries $a_{ij}^k$ tend to zero as $k$ goes to infinity, then $\tilde{d}^{\mathrm{LW}}(1, 2)$ tends to 1 (while all other non-zero distances tend to infinity). If $G$ is balanced, i.e., all its vertices have the same weighted degree, then, obviously, both versions of the long walk distance coincide with the resistance distance: $\tilde{d}^{\mathrm{LW}} \equiv d^{\mathrm{LW}} \equiv d^{\mathrm{r}}$.

## References

[1] R.B. Bapat, Resistance distance in graphs, Math. Student 68 (1999) 87–98.
[2] R.B. Bapat, S. Sivasubramanian, Identities for minors of the Laplacian, resistance and distance matrices, Linear Algebra Appl. 435 (2011) 1479–1489.
[3] P. Chebotarev, A class of graph-geodetic distances generalizing the shortest-path and the resistance distances, Discrete Appl. Math. 159 (2011) 295–302.
[4] P. Chebotarev, The graph bottleneck identity, Adv. Appl. Math. 47 (2011) 403–413.
[5] P. Chebotarev, The walk distances in graphs, Discrete Appl. Math. 160 (2012) 1484–1500.
[6] P. Chebotarev, M. Deza, A topological interpretation of the walk distances, arXiv preprint math.CO/1111.0284 (2011). urlhttp://arxiv.org/abs/1111.0284, in: A. Mucherino, C. Lavor, L. Liberti, N. Maculan (Eds.), Distance Geometry: Theory, Methods, and Applications, 2013. Available from: <http://www.springer.com/mathematics/geometry/book/978-1-4614-5127-3>.
[7] F. Harary, Graph Theory, Addison-Wesley Reading, MA, 1969.
[8] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1953) 39–43.
[9] D.J. Klein, M. Randić, Resistance distance, J. Math. Chem. 12 (1993) 81–95.
[10] C.D. Meyer Jr., Limits and the index of a square matrix, SIAM J. Appl. Math. 26 (1974) 469–478.
[11] C.D. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, Philadelphia, PA, 2000.
[12] C.R. Rao, S.K. Mitra, Generalized Inverse of Matrices and Its Applications, Wiley, New York, 1971.
[13] G.P.H. Styan, G.E. Subak-Sharpe, Inequalities and equalities associated with the Campbell-Youla generalized inverse of the indefinite admittance matrix of resistive networks, Linear Algebra Appl. 250 (1997) 349–370.

---

[7] The superscript $k$ refers to the graph $G^k$.