MINORS OF TREE DISTANCE MATRICES

HARRY RICHMAN, FARBOD SHOKRIEH, AND CHENXI WU

Abstract. We prove a combinatorial formula for the principal minors of the distance matrix of a weighted tree. This generalizes a result of Graham and Pollak. The identity involves counts of rooted spanning forests of the underlying tree.

Contents

1. Introduction

Suppose $G = (V, E)$ is a tree with n vertices. Let D denote the distance matrix of G. In [4], Graham and Pollak proved that

$$\det D = (-1)^{n-1} 2^{n-2} (n-1). \tag{1}$$

This identity is remarkable in that the result does not depend on the tree structure, beyond the number of vertices. The identity (1) was motivated by a problem in data communication, and inspired much further research on distance matrices.

The main result of this paper is to generalize (1) by replacing $\det D$ with any of its principal minors. For a subset $S \subset V(G)$, let $D[S]$ denote the submatrix consisting of the S-indexed rows and columns of D.

Theorem 1. *Suppose G is a tree with n vertices, and distance matrix D. Let $S \subset V(G)$ be a nonempty subset of vertices. Then*

$$\det D[S] = (-1)^{|S|-1} 2^{|S|-2} \left((n-1)\,\kappa(G;S) - \sum_{\mathcal{F}_2(G;S)} (2 - \deg^o(F,*))^2 \right), \tag{2}$$

where $\kappa(G;S)$ is the number of S-rooted spanning forests of G, $\mathcal{F}_2(G;S)$ is the set of $(S,*)$-rooted spanning forests of G, and $\deg^o(F,*)$ denotes the outdegree of the $*$-component of F.

For definitions of $(S,*)$-rooted spanning forests and other terminology, see Section 2. Note that the quantity $\deg^o(F,*)$ satisfies the bounds

$$1 \le \deg^o(F,*) \le |S|.$$

Date: v1, December 6, 2022 (Preliminary draft, not for circulation).

When $S = V$ is the full vertex set, the set of V-rooted spanning forests is a singleton, consisting of the subgraph with no edges, so $\kappa(G;V) = 1$; and moreover the set $\mathcal{F}_2(G;V)$ of $(V,*)$-rooted spanning forests is empty. Thus (2) recovers the Graham–Pollak identity (1) when $S = V$.

1.1. **Weighted trees.** If $\{\alpha_e : e \in E\}$ is a collection of positive edge weights, the α-distance matrix D_α is defined by setting the (u,v)-entry to the sum of the weights α_e along the unique path from u to v. Then

$$\det D_\alpha = (-1)^{n-1} 2^{n-2} \sum_{e\in E} \alpha_e \prod_{e\in E} \alpha_e. \tag{3}$$

This weighted version of (1) was proved by Bapat–Kirkland–Neumann [1]. The weighted identity (3) reduces to (1) when taking all unit weights, $\alpha_e = 1$. We also prove the following weighted version of our main theorem.

Theorem 2. *Suppose $G = (V,E)$ is a finite, weighted tree with edge weights $\{\alpha_e : e \in E\}$, and corresponding weighted distance matrix $D = D_\alpha$. For any nonempty subset $S \subset V$, we have*

$$\det D[S] = (-1)^{|S|-1} 2^{|S|-2} \left(\sum_{E(G)} \alpha_e \sum_{\mathcal{F}_1(G;S)} w(T) - \sum_{\mathcal{F}_2(G;S)} (2 - \deg^o(F,*))^2 w(F) \right). \tag{4}$$

where $\mathcal{F}_1(G;S)$ is the set of S-rooted spanning forests of G, $\mathcal{F}_2(G;S)$ is the set of $(S,*)$-rooted spanning forests of G, $w(T)$ and $w(F)$ denote the α-weights of the forests T and F, and $\deg^o(F,*)$ is the outdegree of the $*$-component of F, as above.

Theorem 2 also reduces to Theorem 1 when taking all unit weights, $\alpha_e = 1$. It is worth observing that depending on the chosen subset $S \subset V$, the distances appearing in the submatrix $D[S]$ may ignore a large part of the ambient tree G. We could instead replace G by the subtree $\mathrm{conv}(S,G)$ consisting of the union of all paths between vertices in S, which we call the *convex hull* of $S \subset G$. To apply formula (2) or (4) "efficiently," we should replace G on the right-hand side with the subtree $\mathrm{conv}(S,G)$. However, the formulas as stated are true even without this replacement due to cancellation of terms.

1.2. **Applications of the main theorem.** Given a matrix A, let $\mathrm{cof}\, A$ denote the *sum of cofactors* of A, i.e.

$$\mathrm{cof}\, A = \sum_{i=1}^{|S|} \sum_{j=1}^{|S|} (-1)^{i+j} \det A_{ij}$$

If A is invertible, then $\mathrm{cof}\, A$ is related to the sum of entries of the matrix inverse A^{-1} by a factor of $\det A$, i.e. $\mathrm{cof}\, A = (\det A)(\mathbf{1}^\intercal A^{-1} \mathbf{1})$. In [2], Bapat and Sivasubramanian showed the following identity for the sum of cofactors of a distance submatrix $D[S]$ of a tree.

$$\mathrm{cof}\, D[S] = (-2)^{|S|-1} \sum_{T \in \mathcal{F}_1(G;S)} w(T). \tag{5}$$

Using (1.2), an immediate corollary to Theorem 2 is the following identity.

Theorem 3. *Suppose $G = (V,E)$ is a finite, weighted tree with edge weights $\{\alpha_e : e \in E\}$. Let $D = D_\alpha$ denote the weighted distance matrix of G. For any nonempty subset $S \subset V$, we have*

$$\frac{\det D[S]}{\mathrm{cof}\, D[S]} = \frac{1}{2} \left(\sum_{e\in E} \alpha_e - \frac{\sum_{F\in\mathcal{F}_2(G;S)} w(F) k(F,*)^2}{\sum_{T\in\mathcal{F}_1(G;S)} w(T)} \right)$$

where $k(F,*) = 2 - \deg^o(F,*)$.

♢ add remark / theorem that det/cof is achieved as result of optimization problem ♢

We remark that the calculation of $\det D[S]$ is related to the following quadratic optimization problem: for all vectors $\mathbf{m} \in \mathbb{R}^S$,

$$\begin{aligned} \text{optimize objective function:} \quad & \mathbf{m}^\intercal D[S]\mathbf{m} \\ \text{with constraint:} \quad & \mathbf{1}^\intercal \mathbf{m} = 1. \end{aligned}$$

The gradient of the objective function is $2D[S]\mathbf{m}$, and the gradient of the constraint is $\mathbf{1}$. By the theory of Lagrange multipliers, the optimal solution $\mathbf{m}^*$ is a vector satisfying

$$D[S]\mathbf{m}^* = \lambda \mathbf{1} \qquad \text{for some } \lambda \in \mathbb{R}.$$

The constant λ is in fact the optimal objective value, since

$$(\mathbf{m}^*)^\intercal D[S]\mathbf{m}^* = (D[S]\mathbf{m}^*)^\intercal \mathbf{m}^* = \lambda(\mathbf{1}^\intercal \mathbf{m}^*) = \lambda.$$

(The above computation uses the fact that $D[S]$ is a symmetric matrix, and the given constraint $\mathbf{1}^\intercal \mathbf{m} = 1$.) On the other hand, assuming $D[S]$ is invertible we have $\mathbf{m}^* = \lambda(D[S]^{-1}\mathbf{1})$, so that

$$1 = \mathbf{1}^\intercal \mathbf{m}^* = \lambda(\mathbf{1}^\intercal D[S]^{-1}\mathbf{1}) = \lambda \frac{\operatorname{cof} D[S]}{\det D[S]}.$$

Thus the optimal objective value is $\lambda = \dfrac{\det D[S]}{\operatorname{cof} D[S]}$.

Proposition 4. *If $D[S]$ is a principal submatrix of distance matrix indexed by S, then*

$$\frac{\det D[S]}{\operatorname{cof} D[S]} = \max\{\mathbf{m}^\intercal D[S]\mathbf{m} : \mathbf{m} \in \mathbb{R}^S,\ \mathbf{1}^\intercal \mathbf{m} = 1\}$$

where $\operatorname{cof} D[S]$ *denotes the sum of cofactors,*

$$\operatorname{cof} A = \sum_{i=1}^{|S|} \sum_{j=1}^{|S|} (-1)^{i+j} \det A_{ij}$$

Theorem 5 (Monotonicity of principal minor ratios)**.** *If $A, B \subset V(G)$ are nonempty subsets with $A \subset B$, then*

$$\frac{\det D[A]}{\operatorname{cof} D[A]} \leq \frac{\det D[B]}{\operatorname{cof} D[B]}.$$

Theorem 6 (Bounds on principal minor ratios)**.** *Suppose $G = (V, E)$ is a finite, weighted tree with distance matrix D.*

(1) If $S \subset V(G)$ is nonempty,

$$0 \leq \frac{\det D[S]}{\operatorname{cof} D[S]} \leq \frac{1}{2} \sum_{E(G)} \alpha_e.$$

(2) If $\operatorname{conv}(S, G)$ denotes the subtree of G consisting of all paths between points of $S \subset V(G)$,

$$\frac{\det D[S]}{\operatorname{cof} D[S]} \leq \frac{1}{2} \sum_{E(\operatorname{conv}(S,G))} \alpha_e.$$

(3) If γ is a simple path between vertices $s_0, s_1 \in S$, then

$$\frac{1}{2} \sum_{e \in \gamma} \alpha_e \leq \frac{\det D[S]}{\operatorname{cof} D[S]}$$

Theorem 7 (Nonsingular minors)**.** *Let G be a finite tree with (weighted) distance matrix D, and let $S \subset V(G)$ be a subset of vertices. If $|S| \geq 2$ then $\det D[S] \neq 0$.*

1.3. **Previous work.** A formula for the inverse matrix D^{-1} was found by Graham and Lovász in graham-lovasz [3].

1.4. **Notation.** G a finite graph, loops and parallel edges allowed, possibly disconnected
$E(G)$ edge set of G
$V(G)$ vertex set of G
$\kappa(G)$ number of spanning trees of G
$\mathcal{F}_1(G)$ the set of spanning trees of G
$\mathcal{F}_2(G)$ the set of 2-component forests of G
$\mathcal{F}_1(G;S)$ the set of S-rooted spanning forests of G
$\mathcal{F}_2(G;S)$ the set of $(S,*)$-rooted spanning forests of G

2. Graphs and matrices

For background on enumeration problems for graphs and trees, see Moon [5].

Given a graph $G=(V,E)$ with edge weights $\{\alpha_e : e \in E\}$, for any edge subset $A \subset E$ we define the *weight* of A as

$$w(A) = \prod_{e \in A} \alpha_e.$$

We define the *co-weight* of A as

$$w(\overline{A}) = \prod_{e \notin A} \alpha_e.$$

By abuse of notation, if H is a subgraph of G, we use $w(\overline{H})$ to denote $w(\overline{E(H)})$.

2.1. **Spanning trees and forests.** A *spanning tree* of a graph G is a subgraph which is connected, has no cycles, and contains all vertices of G. A *spanning forest* of a graph G is a subgraph which has no cycles and contains all vertices of G.

Given a set of vertices $S=\{v_1,v_2,\ldots,v_k\}$, an *S-rooted spanning forest* of G is a spanning forest which has exactly one vertex v_i in each connected component. An *$(S,*)$-rooted spanning forest* of G is a spanning forest which has $|S|+1$ components, where $|S|$ components each contain one vertex of S, and the additional component is disjoint from S. We call the component disjoint from S the "floating component."

Let

$$\kappa_k(v_1|v_2|\cdots|v_k)$$

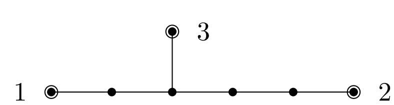
denote the number of k-component spanning trees which have a vertex v_i in each component. If $S=\{v_1,\ldots,v_k\}$, then $\kappa_k(v_1|\cdots|v_k) = \kappa(G/S)$.

If u,v,w are vertices, then let

$$\kappa_2(uv|w)$$

denote the number of two-forests which have u,v in one component and w in the other component.

Example 8. Suppose G is the tree with unit edge lengths shown below.



Let S be the set of three leaf vertices. Then $\mathcal{F}_1(G;S)$ contains 11 forests, while $\mathcal{F}_2(G;S)$ contains 19 forests. These are shown in Figures 1 and 2, respectively.

2.2. **Laplacian matrix.** Given a graph $G=(V,E)$, let $L \in \mathbb{R}^{V\times V}$ denote the *Laplacian matrix* of G. If G is a weighted graph with edge weights $\alpha_e \in \mathbb{R}_{>0}$ for $e \in E$, let L denote the weighted Laplacian matrix of G.

Given $S \subset V$, let $L[\overline{S}]$ denote the matrix obtained from L by removing the rows and columns indexed by S.

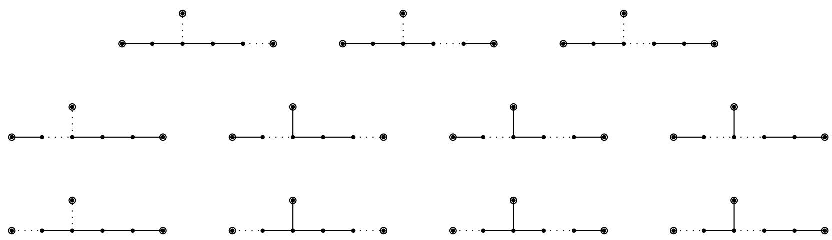


FIGURE 1. Forests in $\mathcal{F}_1(G;S)$. fig:1-forests

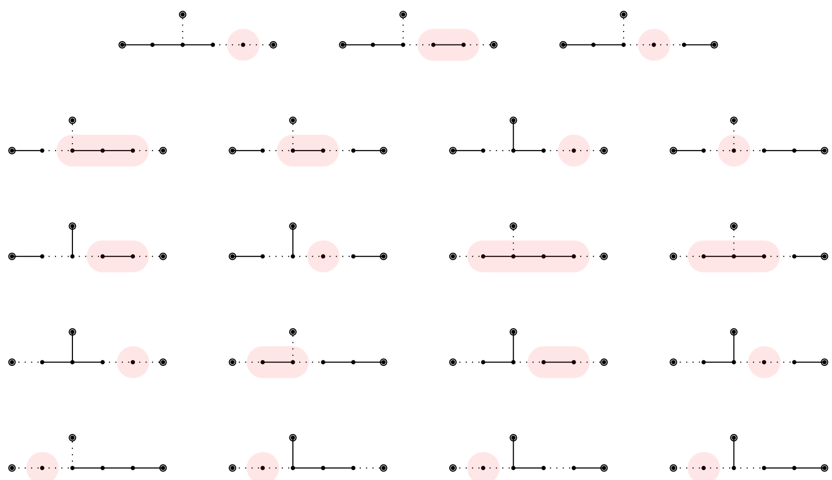


FIGURE 2. Forests in $\mathcal{F}_2(G;S)$. fig:2-forests

Definition 9 (Weighted Laplacian matrix)**.** Given a graph $G=(V,E)$ and edge weights $\{\alpha_e : e \in E\}$, the *weighted Laplacian matrix* $L_\alpha \in \mathbb{R}^{V\times V}$ is defined by

$$(L_\alpha)_{v,w} = \begin{cases} 0 & \text{if } v \neq w \text{ and } (v,w) \notin E \\ \alpha_e^{-1} & \text{if } v \neq w \text{ and } (v,w) = e \in E \\ -\displaystyle\sum_{e \in N(v)} \alpha_e^{-1} & \text{if } v = w. \end{cases}$$

For any graph G, let $\kappa(G)$ denote the number of spanning trees of G. The following theorem is due to Kirchhoff. For any graph G, let $\kappa(G)$ denote the number of spanning trees of G.

thm:matrix-tree **Theorem 10** (All-minors matrix tree theorem)**.** *Let $G=(V,E)$ be a finite graph, and let L denote the Laplacian matrix of G. Then for any nonempty vertex set $S \subset V$,*

$$\det L[\overline{S}] = \kappa(G;S).$$

Note that $\kappa(G;S)$ is also the number of spanning trees of the quotient graph G/S, which "glues together" all vertices in S as a single vertex.

The following result is due to Bapat–Siviasubramanian.

Theorem 11 (Distance matrix cofactor sums bapat-sivasubramanian [2])**.** *Given a tree G, let D be the distance matrix of G, and L the Laplacian matrix. Let $S \subset V(G)$ be a nonempty subset of vertices of G. Then*

$$\operatorname{cof} D[S] = (-2)^{|S|-1} \det L[\overline{S}].$$

ec:tree-distance

2.3. **Tree distance.** For $e \in E$ and $v, w \in V$, let

$$\delta(e; v, w) = \begin{cases} 1 & \text{if } e \text{ separates } v \text{ from } w, \\ 0 & \text{otherwise.} \end{cases}$$

(Note that $\delta(e; v, v) = 0$ for any e and v.) Given an edge $e \in E$ and vertices $v, w \in V$, let

$$\delta(e; v, w) = \begin{cases} 1 & \text{if } e \text{ lies on } v \sim w \text{ path,} \\ 0 & \text{otherwise.} \end{cases}$$

We can express the tree distance $d(v, w)$ as a sum over edges

$$d(v, w) = \sum_{e \in E(G)} \delta(e; v, w) \qquad \text{where } \delta(e; v, w) = \begin{cases} 1 & \text{if } e \text{ lies on } v \sim w \text{ path,} \\ 0 & \text{otherwise.} \end{cases}$$

We have the following perspectives on the function $\delta(e; v, w)$:

- If we fix v and w, then $\delta(-; v, w) : E(G) \to \{0, 1\}$ is the indicator function for the unique v w path in G.
- On the other hand if we fix e and v, then the deletion $G \setminus e$ has two connected components, and $\delta(e; v, -) : V(G) \to \{0, 1\}$ is the indicator function for the component of $G \setminus e$ not containing v.

rop:distance-sum

Proposition 12 (Weighted tree distance)**.** *For a tree $G = (V, E)$ with weights $\{\alpha_e : e \in E\}$, the weighted distance function satisfies*

$$d_\alpha(v, w) = \sum_{e \in E} \alpha_e \, \delta(e; v, w).$$

2.4. **Principal cuts.** Given a tree $G = (V, E)$ and an edge $e \in E$, the edge deletion $G \setminus e$ contains two connected components. Using the implicit orientation on $e = (e^+, e^-)$, we let $(G \setminus e)^+$ denote the component that contains endpoint e^+, respectively $(G \setminus e)^-$ and endpoint e^-.

For any $e \in E$ and $v \in V$, we let $(G \setminus e)^v$ denote the component of $G \setminus e$ containing v, respectively $(G \setminus e)^{\overline{v}}$ ♢ or $(G \setminus e)^{-v}$ ♢ for the component not containing v.

2.5. **Outdegree of rooted forest.** Given a rooted forest F in $\mathcal{F}(G; S)$ and $s \in S$, let $F(s)$ denote the s-component of F. We define the *outdegree* $\deg^o(F, s)$ by

eq:outdeg

$$\deg^o(F, s) = \#\{e = (a, b) \in E : a \in F(s),\, b \notin F(s)\}. \tag{6}$$

In words, $\deg^o(F, s)$ is the number of edges which connect the s-component of F to a different component.

$$\#\{e \in E : e \text{ connects the } s\text{-component of } F \text{ to a different component}\}$$

If F is a forest in $\mathcal{F}_2(G; S)$, let $\deg^o(F, *)$ denote the outdegree of the floating component.

lem:outdeg-sum

Lemma 13. *Suppose G is a tree and $H \subset G$ is a (nonempty) connected subgraph. Then*

$$\sum_{v \in V(H)} (2 - \deg(v)) = 2 - \deg^o(H).$$

Proof. This is straightforward to check by induction on $|V(H)|$, with base case $|V(H)| = 1$: if $H = \{v\}$ consists of a single vertex, then $\deg^o(H) = \deg(v)$. ☐

3. Proofs

In this section we prove Theorem thm:w-main 2.

Outline of proof: given a subset $S \subset V$ and distance submatrix $D[S]$, we will

(i) Find vector $\mathbf{m} \in \mathbb{R}^S$ such that $D[S]\mathbf{m} = \lambda \mathbf{1} \in \mathbb{R}^S$.
(ii) Compute the sum of entries of $\mathbf{m}$, i.e. $\mathbf{1}^\intercal \mathbf{m}$.
(iii) Using (i), relate the sum $\mathbf{1}^\intercal \mathbf{m}$ to the sum of entries of the inverse matrix $D[S]^{-1}$:

$$\mathbf{1}^\intercal \mathbf{m} = \lambda(\mathbf{1}^\intercal D[S]^{-1}\mathbf{1}) = \lambda \frac{\operatorname{cof} D[S]}{\det D[S]}.$$

where $\operatorname{cof} D[S]$ is the sum of cofactors of $D[S]$.
(iv) Use known expression for $\operatorname{cof} D[S]$ to compute

$$\det D[S] = \lambda (\operatorname{cof} D[S]) \left(\mathbf{1}^\intercal \mathbf{m}\right)^{-1}.$$

The interesting part of this expression will turn out to be in the constant λ.

Example 14. Suppose G is a tree consisting of three paths joined at a central vertex. Let S consist of the central vertex, and the three endpoints of the paths. The corresponding submatrix of the distance matrix is

$$D[S] = \begin{bmatrix} 0 & a & b & c \\ a & 0 & a+b & a+c \\ b & a+b & 0 & b+c \\ c & a+c & b+c & 0 \end{bmatrix}.$$

Following the steps outlined above:

(i) The vector $\mathbf{m} = \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ satisfies $D[S]\mathbf{m} = (a+b+c)\mathbf{1}$
(ii) The sum of entries of $\mathbf{m}$ is $\mathbf{1}^\intercal \mathbf{m} = 2$.
(iii) We have

$$2 = \mathbf{1}^\intercal \mathbf{m} = \lambda(\mathbf{1}^\intercal D[S]^{-1}\mathbf{1}) = \lambda \frac{\operatorname{cof} D[S]}{\det D[S]}.$$

(iv) The cofactor sum $\operatorname{cof} D[S]$ is $-8abc$, so the determinant is

$$\det D[S] = \lambda \frac{\operatorname{cof} A}{\mathbf{1}^\intercal \mathbf{m}} = (a+b+c)(-8abc)\frac{1}{2} = -4(a+b+c)abc.$$

3.1. Warmup case: $S = V$.

Proposition 15. *Let $G = (V, E)$ a tree, and consider the vector $\mathbf{m} \in \mathbb{R}^V$ defined by*

$$\mathbf{m}_v = 2 - \deg v \qquad \textit{for each } v \in V.$$

Then $\mathbf{1}^\intercal \mathbf{m} = \sum_{v \in V}(2 - \deg v) = 2$.

Proof. For any graph, $\sum_{v\in V} \deg v = 2|E|$. Since G is a tree, $|E| = |V| - 1$. □

-distance-warmup **Proposition 16.** *Let $\mathbf{m}$ be the vector defined above, and let D be the distance matrix of G. Then $D\mathbf{m} = \lambda\mathbf{1}$ for some constant λ.*

Proof. It suffices to show that for each edge e, with endpoints (e^+, e^-), we have

$$(D\mathbf{m})_{(e^+)} = (D\mathbf{m})_{(e^-)}.$$

We compute

eq:12-1

$$(D\mathbf{m})_{(e^+)} - (D\mathbf{m})_{(e^-)} = \sum_{v\in V}(d(v,e^+) - d(v,e^-))(2-\deg v)$$
$$= \sum_{v\in (G\setminus e)^-} \alpha_e(2-\deg v) - \sum_{v\in (G\setminus e)^+} \alpha_e(2-\deg v) \tag{7}$$

since

$$d(v,e^+) - d(v,e^-) = \begin{cases} \alpha_e & \text{if } v \text{ is closer to } e^- \text{ than } e^+, \\ -\alpha_e & \text{if } v \text{ is closer to } e^+ \text{ than } e^-. \end{cases}$$

♢ TO DO: define notation $(G\setminus e)^\pm$ ♢ For each sum in (6), we apply Proposition ♢ cite ♢ to obtain

$$\alpha_e \sum_{v\in (G\setminus e)^-} (2-\deg v) = \alpha_e(2-\deg^o((G\setminus e)^-) = \alpha_e.$$

The same identity applies to the sum over $(G\setminus e)^+$, so $(D\mathbf{m})_{(e^+)} = (D\mathbf{m})_{(e^-)}$ as desired. □

3.2. **General case: $S\subset V$.** Fix a tree $G=(V,E)$ and a subset $S\subset V$.

dfn:m-vector **Definition 17.** Let $\mathbf{m} = \mathbf{m}(G;S)\in\mathbb{R}^S$ be defined by

eq:m-vector

$$\mathbf{m}_v = \sum_{T\in\mathcal{F}_1(G;S)} (2-\deg^o(T,v))w(T) \qquad \text{for each } v\in S. \tag{8}$$

where $\deg^o(T,v)$ is the outdegree of the v-component of T, (5).

Let $\mathbf{1}$ denote the all-ones vector.

Proposition 18. *For* $\mathbf{m}$ *defined above,* $\mathbf{1}^\intercal\mathbf{m} = 2\sum_{T\in\mathcal{F}_1(G;S)} w(T)$.

Proof. We have

$$\mathbf{1}^\intercal\mathbf{m} = \sum_{s\in S}\mathbf{m}_s = \sum_{s\in S}\left(\sum_{T\in\mathcal{F}_1(G;S)}(2-\deg^o(T,s))w(T)\right)$$
$$= \sum_{T\in\mathcal{F}_1(G;S)} w(T)\left(\sum_{s\in S}\sum_{v\in T(s)} 2-\deg(v)\right)$$
$$= \sum_{T\in\mathcal{F}_1} w(T)\left(\sum_{v\in V} 2-\deg(v)\right) = \sum_{T\in\mathcal{F}_1} w(T)\cdot 2.$$

In the second line we apply Lemma 13 and exchange the outer summations. To obtain the third line, we observe that the vertex sets of $T(s)$ for $s\in S$ form a partition of V, since T is an S-rooted spanning forest. Finally we again apply Lemma 13 for the last equality, as $\deg^o(G)=0$. □

Corollary 19. *If G is a graph with unit edge weights $\alpha_e = 1$, then the vector* $\mathbf{m}$ *defined in (7) satisfies* $\mathbf{1}^\intercal\mathbf{m} = 2\,\kappa(G;S)$.

Theorem 20. *With* $\mathbf{m} = \mathbf{m}(G;S)$ *defined as in (7),* $D[S]\mathbf{m} = \lambda\mathbf{1}$ *for the constant*

$$\lambda = \sum_{E(G)}\alpha_e \sum_{\mathcal{F}_1(G;S)} w(T) - \sum_{\mathcal{F}_2(G;S)} (2-\deg^o(F,*))^2\, w(F)$$

where $\deg^o(F,w)$ *is the out-degree of the w-component of F (as a spanning forest).*

sec:tree-distance

Proof. For $e \in E$ and $v, w \in V$, let $\delta(e; v, w)$ denote the function defined in Section 2.3 ♢ check section ♢ For any $v \in S$, we have

eq:14-1

$$
\begin{aligned}
(D[S]\mathbf{m})_v &= \sum_{s\in S} d(v,s)\mathbf{m}_s \\
&= \sum_{s\in S}\left(\sum_{e\in E(G)} \alpha_e\, \delta(e;v,s)\right)\left(\sum_{T\in\mathcal{F}_1(G;S)} (2-\deg^o(T,s))w(T)\right) \\
&= \sum_{T\in\mathcal{F}_1} w(T)\sum_{e\in E}\alpha_e\left(\sum_{s\in S}\delta(e;v,s)(2-\deg^o(T,s))\right) \\
&= \sum_{T\in\mathcal{F}_1} w(T)\sum_{e\in E}\alpha_e\left(\sum_{s\in S}\delta(e;v,s)\sum_{u\in T(s)}(2-\deg(u))\right). \qquad (9)
\end{aligned}
$$

eq:14-1

We introduce additional notation to handle the double sum in parentheses in (8). Each S-rooted spanning tree T naturally induces a surjection $\pi_T : V \to S$, defined by

$$\pi_T(u) = s \qquad \text{if and only if} \qquad u \in T(s).$$

Using this notation,

$$(D[S]\mathbf{m})_v = \sum_{T\in\mathcal{F}_1} w(T)\sum_{e\in E}\alpha_e\left(\sum_{u\in V}(2-\deg(u))\delta(e;v,\pi_T(u))\right)$$

We will compare the above expression with the one obtained after replacing $\delta(e; v, \pi_T(u))$ with $\delta(e; v, u)$. From ♢ cite previous prop ♢ , for any $v \in V$ and $e \in E$ we have

$$\sum_{u\in V}(2-\deg(u))\delta(e;v,u) = 2-\deg^o((G\setminus e)^{\overline{v}}) = 1.$$

Thus

$$(D[S]\mathbf{m})_v - \sum_{T\in\mathcal{F}_1} w(T)\sum_{e\in E}\alpha_e = \sum_{T\in\mathcal{F}_1} w(T)\sum_{e\in E}\alpha_e\sum_{u\in V}(2-\deg(u))\left(\delta(e;v,\pi_T(u))-\delta(e;v,u)\right)$$

When $e \in E$ and $v \in V$ are fixed, $u \mapsto \delta(e; v, u)$ is the indicator function of one component of the principal cut $G \setminus e$. Recall that $\delta(e; \cdot, \cdot)$ is a $(0, 1)$-valued pseudometric on V. We have

$$\delta(e;v,\pi_T(u))-\delta(e;v,u) = \begin{cases} 0 & \text{if } \delta(e;\pi_T(u),u)=0 \\ 1 & \text{if } \delta(e;\pi_T(u),u)=1 \text{ and } \delta(e;v,\pi_T(u))=1 \\ -1 & \text{if } \delta(e;\pi_T(u),u)=1 \text{ and } \delta(e;v,\pi_T(u))=0 \end{cases}$$

Now consider varying u over all vertices, when e, T, and v are fixed. We have the following three cases:

Case 1: if $e \notin T$, then u and $\pi_T(u)$ are on the same side of the principal cut $G \setminus e$, for every vertex u. In this case $\delta(e; v, \pi_T(\cdot)) - \delta(e; v, \cdot) = 0$.

Case 2: if $e \in T(s_0)$ and s_0 is separated from v by e, then $\delta(e; v, \pi_T(\cdot)) - \delta(e; v, \cdot)$ is the indicator function for the floating component of $T \setminus e$.

Case 3: if $e \in T(s_0)$ and s_0 is on the same component as v from e, then $\delta(e; v, \pi_T(\cdot)) - \delta(e; v, \cdot)$ is the negative of the indicator function for the floating component of $T \setminus e$.

Thus when multiplying the above term by $(2 - \deg(u))$ and summing over all vertices u, we obtain

$$\sum_{u\in V}(2-\deg(u))(\delta(e;v,\pi_T(u))-\delta(e;v,u)) = \begin{cases} 0 & \text{if } e\notin T, \\ 2-\deg^o(T\setminus e,*) & \text{if } e\in T(s_0) \text{ and } \delta(e;v,s_0)=1, \\ -(2-\deg^o(T\setminus e,*)) & \text{if } e\in T(s_0) \text{ and } \delta(e;v,s_0)=0. \end{cases}$$

FIGURE 3. Edge $e \in T(s_0)$ with $\delta(e; v, s_0) = 1$ (left) and $\delta(e; v, s_0) = 0$ (right). The floating component of $T \setminus e$ is highlighted.

Thus

eq:1

$$(D[S]\mathbf{m})(v) - \sum_{T\in\mathcal{F}_1} w(T) \sum_{e\in E} \alpha_e = \sum_{T\in\mathcal{F}_1} w(T) \sum_{s_0\in S} \left(\sum_{\substack{e\in T(s_0)\\ \delta(e;v,s_0)=1}} \alpha_e (2 - \deg^o(T\setminus e, *)) - \sum_{\substack{e\in T(s_0)\\ \delta(e;v,s_0)=0}} \alpha_e (2 - \deg^o(T\setminus e, *)) \right). \tag{10}$$

We now rewrite the above expression in terms of $\mathcal{F}_2(G; S)$, observing that the deletion $T \setminus e$ is an $(S, *)$-rooted spanning forest of G, if $e \in T$, and that the corresponding weights satisfy

$$w(F) = \alpha_e \cdot w(T) \qquad \text{if} \qquad F = T \setminus e.$$

Thus

$$\text{(3.2)} = \sum_{F\in\mathcal{F}_2} w(F)(2 - \deg^o(F, *)) \sum_{T\in\mathcal{F}_1} \sum_{s_0\in S} \left(\sum_{\substack{e\in T(s_0)\\ \delta(e;v,s_0)=1}} \mathbb{1}(F = T\setminus e) - \sum_{\substack{e\in T(s_0)\\ \delta(e;v,s_0)=0}} \mathbb{1}(F = T\setminus e) \right)$$

$$= \sum_{F\in\mathcal{F}_2} w(F)(2 - \deg^o(F, *)) \Big(\#\{T \in \mathcal{F}_1 : F = T \setminus e \text{ for some } e \in T(s_0),\ \delta(e; v, s_0) = 1\}$$
$$- \#\{T \in \mathcal{F}_1 : F = T \setminus e \text{ for some } e \in T(s_0),\ \delta(e; v, s_0) = 0\} \Big)$$

Next, we note that $F = T \setminus e$ is equivalent to $T = F \cup e$, and in particular this only occurs when we choose the edge e to be in the floating boundary $\partial F(*)$:

FIGURE 4. Edge $e \in \partial F(*)$ with $\delta(e; v, F(*)) = 0$ (left) and $\delta(e; v, F(*)) = 1$ (right). The floating component $F(*)$ is highlighted.

Now for any $e \notin F$, let $\delta(e; v, F(*)) = \delta(e; v, x)$ for any $x \in F(*)$, i.e.

$$\delta(e; v, F(*)) = \begin{cases} 1 & \text{if } e \text{ lies on path from } v \text{ to } F(*), \\ 0 & \text{otherwise.} \end{cases}$$

The condition $F = T \setminus e$ for some $e \in T(s_0)$ with $\delta(e; v, s_0) = 1$ (resp. $\delta(e; v, s_0) = 0$) is equivalent to $T = F \cup e$ for some $e \in \partial F(*)$ with $\delta(e; v, F(*)) = 0$ (resp. $\delta(e; v, F(*)) = 1$). Thus

$$(3.2) = \sum_{F \in \mathcal{F}_2} w(F)(2 - \deg^o(F, *)) \Big(\#\{e \in \partial F(*) : \delta(e; v, F(*)) = 0\} - \#\{e \in \partial F(*) : \delta(e; v, F(*)) = 1\} \Big).$$

Finally, we observe that for any forest F in $\mathcal{F}_2(G; S)$, there is exactly one edge e in the boundary $\partial F(*)$ of the floating component which satisfies $\delta(e; v, F(*)) = 1$, namely the unique boundary edge on the path from the floating component $F(*)$ to v. The previous expression (3.2) simplifies as

$$\#\{e \in \partial F(*) : \delta(e; v, F(*)) = 1\} = 1 \qquad \text{and} \qquad \#\{e \in \partial(F, *) : \delta(e; v, F(*)) = 0\} = \deg^o(F, *) - 1.$$

Thus

$$\begin{aligned}(3.2) &= \sum_{F \in \mathcal{F}_2} w(F)(2 - \deg^o(F, *)) \Big((\deg^o(F, *) - 1) - (1) \Big) \\ &= -\sum_{F \in \mathcal{F}_2} w(F)(2 - \deg^o(F, *))^2.\end{aligned}$$

as desired. ☐

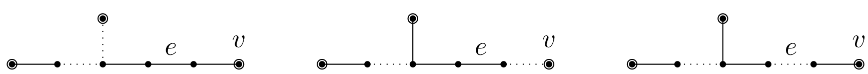


FIGURE 5. Components rooted in $S(G \setminus e)^{\overline{v}}$.

♢ MOVE TO REMARK? If $e \in \mathrm{conv}(G, S)$, then $S(G\setminus e)^{\overline{v}}$ is nonempty and ♢

Remark 21. The set $\mathcal{F}_2(G; S)$ of $(S, *)$-rooted spanning forests of G can be partitioned into two types: "active" and "inactive".

$$\mathcal{F}_2(G; S) = \mathcal{F}_2^{in}(G; S) \sqcup \mathcal{F}_2^{out}(G; S),$$

where

$$\begin{aligned}\mathcal{F}_2^{in}(G; S) &= \{F \in \mathcal{F}_2(G; S) \text{ such that } \deg^o(*, F) \geq 2\}, \\ \mathcal{F}_2^{out}(G; S) &= \{F \in \mathcal{F}_2(G; S) \text{ such that } \deg^o(*, F) = 1\}.\end{aligned}$$

Remark 22. A key step in the above proof is the use of the map

$$E(G) \times \mathcal{F}_1(G; S) \mapsto S \sqcup \{\text{error}\}$$

defined by

$$(e, T) \mapsto \begin{cases} s & \text{if } e \in T(s), \\ \text{error} & \text{if } e \notin T. \end{cases}$$

Remark 23. For a given spanning forest $F \in \mathcal{F}_2(G; S)$, there are exactly $\deg^o(F, *)$-many choices of pairs $(T, e) \in \mathcal{F}_1(G; S) \times E(G)$ such that $F = T \setminus e$. Consider the "deletion" map

$$E(G) \times \mathcal{F}_1(G; S) \to \mathcal{F}_2(G; S) \sqcup \mathcal{F}_1(G; S)$$

defined by ...

$$(e, T) \mapsto \begin{cases} T \setminus e & \text{if } e \in T, \\ T & \text{if } e \notin T. \end{cases}$$

For a forest F in $\mathcal{F}_2(G; S)$, the preimage under this map has $\deg^o(F, *)$ elements.

There is an associated "union" map

$$E(G) \times \mathcal{F}_2(G;S) \longrightarrow \mathcal{F}_1(G;S) \sqcup \mathcal{F}_2(G;S)$$

defined by

$$(e,F) \mapsto \begin{cases} F \cup e & \text{if } e \notin F, \\ F & \text{if } e \in F \end{cases}$$

4. Optimization: quadratic programming

5. Physical interpretation

If we consider G as a network of wires with each edge containing a unit resistor, which is grounded at all nodes in S, then $\mathbf{m}_S$ records the currents flowing to S when current is added on $V \setminus S$ in the amount $2 - \deg v$ for each $v \notin S$.

5.1. Alternate proof.

Let $\mathbf{1}$ denote the all-ones vector. When we choose a subset $S \subset V(G)$, we no longer have a single "obvious" replacement for $\mathbf{m}$ inside $\mathbb{R}^S$. Instead, we can take an average over S-rooted spanning forests.

In the outline above, our first goal is to find a "special" vector $\mathbf{m} \in \mathbb{R}^S$ satisfying $D[S]\mathbf{m} = \lambda\mathbf{1}$. We can approach this first goal as follows: consider $\mathbb{R}^S$ inside the larger vector space $\mathbb{R}^V = \mathbb{R}^S \oplus \mathbb{R}^{V\setminus S}$, and we wish to find vectors $\mathbf{n}_i \in \mathbb{R}^V$ satisfying $\pi_S(D\mathbf{n}_i) = \lambda_i\mathbf{1}$. By finding sufficiently many such vectors $\mathbf{n}_i$, we can hope to find a linear combination that lies inside $\mathbb{R}^S \oplus \{0\}$.

prop:n-vector

Proposition 24. *Suppose $v \in V \setminus S$. For each $s_j \in S$, let $\mu(v,s_j)$ = current flowing to s_j when G is grounded at S and one unit of current enters G at v. Explicitly,*

$$\begin{aligned} \mu(v,s) &= \frac{\#\text{ of } S\text{-rooted spanning forests of } G \text{ whose } s_j\text{-component contains } v}{\#\text{ of } S\text{-rooted spanning forests of } G} \\ &= \frac{\sum_{\mathcal{F}_1(G/S)} \mathbb{1}(v \in T(s))}{\kappa(G/S)} \\ &= \frac{\kappa_r(s_1|\cdots|s_j v|\cdots|s_r)}{\kappa_r(s_1|\cdots|s_r)} \end{aligned}$$

Consider the vector $\mathbf{n} = \mathbf{n}(G;S,v) \in \mathbb{R}^V$ defined by

$$\mathbf{n}_v = 1, \qquad \mathbf{n}_s = -\mu(v,s) \text{ if } s \in S, \qquad \mathbf{n}_w = 0 \text{ if } w \notin S \cup v$$

Then $D\mathbf{n}$ is constant on S, i.e. $\pi_S(D\mathbf{n}) = \lambda\mathbf{1}$ for some λ.

Proof sketch. For any $s, s' \in S$, consider tracking the value of $D\mathbf{n}$ along path from s to s'. The value of $D\mathbf{n}$ changes according to current flow in the corresponding network, i.e. $D\mathbf{n}$ records electrical potential. By assumption S is grounded, so $D\mathbf{n}$ takes the same value at s and s'. □

Theorem 25. *Let G be a tree, S a nonempty subset of vertices, and $D[S]$ the corresponding submatrix of the distance matrix. Suppose $\mathbf{m} = \mathbf{m}(G;S) \in \mathbb{R}^S$ is defined by* eq:m-vector *(7);*

$$\mathbf{m}(G;S)_v = \sum_{T \in \mathcal{F}_1(G;S)} \sum_{w \in T_v} (2 - \deg w) = \sum_{T \in \mathcal{F}_1(G;S)} 2 - \deg^o(T,v).$$

Then $D[S]\mathbf{m} = \lambda\mathbf{1}$ for some constant λ.

Proof. The vector $\mathbf{m} = \mathbf{m}(G;S)$ can be expressed as a linear combination

$$\begin{aligned}\mathbf{m}(G;S) &= \kappa(G;S)\left(\sum_{v\in V}(2-\deg v)\boldsymbol{\delta}(v) - \sum_{v\in V\setminus S}(2-\deg v)\mathbf{n}(G;S,v)\right)\\ &= \kappa(G;S)\left(\mathbf{m}(G;V) - \sum_{v\in V\setminus S}(2-\deg v)\mathbf{n}(G;S,v)\right)\end{aligned}$$

♢ TODO: elaborate on this equation ♢ From Proposition prop:m-distance-warmup 16 we know that $D\mathbf{m}(G;V)$ is constant on V, and from Proposition prop:n-vector 26 we know that $D\mathbf{n}(G;S,v)$ is constant on S. Hence by linearity, $D\mathbf{m}(G;S)$ is constant on S. ☐

Proposition 26. *Let $G = (V,E)$ be a tree, and $S \subset V$. Suppose we label $S = \{s_1,\dots,s_r\}$ and $V\setminus S = \{t_1,\dots,t_{n-r}\}$. For each $t_i \in V\setminus S$, consider $\mathbf{f}_i \in \mathbb{R}^V$ defined by*

Example 27. If

$$D[S\cup t] = \begin{bmatrix} 0 & a & b & c\\ a & 0 & a+b & a+c\\ b & a+b & 0 & b+c\\ c & a+c & b+c & 0\end{bmatrix}$$

then

$$\begin{bmatrix} 0 & a & b & c\\ a & 0 & a+b & a+c\\ b & a+b & 0 & b+c\\ c & a+c & b+c & 0\end{bmatrix}\begin{bmatrix} ab+ac+bc\\ -bc\\ -ac\\ -ab\end{bmatrix} = \begin{bmatrix} -3abc\\ -abc\\ -abc\\ -abc\end{bmatrix}$$

6. Examples

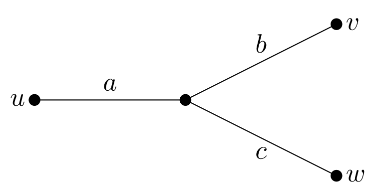
Example 28. Suppose G is a tree consisting of three paths joined at a central vertex. Let S consist of the central vertex, and the three endpoints of the paths. The corresponding minor of the distance matrix is

$$D[S] = \begin{bmatrix} 0 & a & b & c\\ a & 0 & a+b & a+c\\ b & a+b & 0 & b+c\\ c & a+c & b+c & 0\end{bmatrix} \sim \begin{bmatrix} 0 & a & b & c\\ a & -a & a & a\\ b & b & -b & b\\ c & c & c & -c\end{bmatrix} \sim \begin{bmatrix} 0 & a & b & c\\ a & -2a & 0 & 0\\ b & 0 & -2b & 0\\ c & 0 & 0 & -2c\end{bmatrix}.$$

The determinant is

$$\det D[S] = -4(a+b+c)abc.$$

Example 29. Suppose Γ is a tripod with lengths a, b, c and corresponding leaf vertices u, v, w.



Let $S = \{u, v, w\}$. Then

$$D[S] = \begin{bmatrix} 0 & a+b & a+c\\ a+b & 0 & b+c\\ a+c & b+c & 0\end{bmatrix}.$$

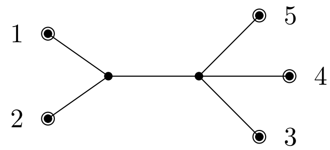
and

$$\det D[S] = 2(a+b)(a+c)(b+c) = 2\left((a+b+c)(ab+ac+bc) - abc\right).$$

The "special vector" that satisfies $D[S]\mathbf{m} = \lambda\mathbf{1}$ in this example is

$$\mathbf{m} = \begin{bmatrix} a(b+c) & b(a+c) & c(a+b) \end{bmatrix}^{\intercal}.$$

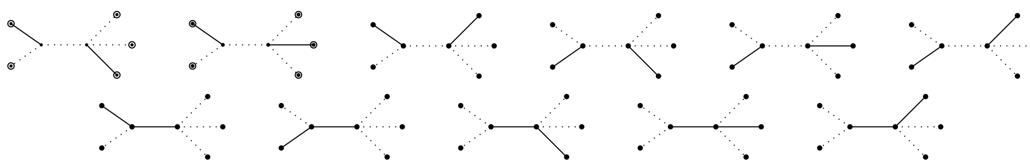
Example 30. Suppose Γ is the tree with unit edge lengths shown below, with five leaf vertices.



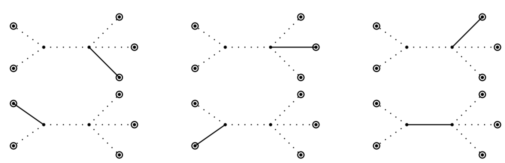
Let S denote the set of five leaf vertices. Then

$$D[S] = \begin{bmatrix} 0 & 2 & 3 & 3 & 3 \\ 2 & 0 & 3 & 3 & 3 \\ 3 & 3 & 0 & 2 & 2 \\ 3 & 3 & 2 & 0 & 2 \\ 3 & 3 & 2 & 2 & 0 \end{bmatrix}.$$

Forests in $\mathcal{F}_1(G;S)$:



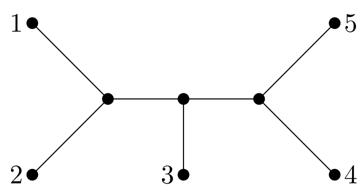
Forests in $\mathcal{F}_2(G;S)$:



and

$$\det D[S] = 368 = (-1)^4 2^3 \left(6 \cdot 11 - (3 \cdot 1^2 + 2 \cdot 2^2 + 1 \cdot 3^2)\right)$$

Example 31. Suppose Γ is the tree with unit edge lengths shown below, with five leaf vertices and three internal vertices.



Let S denote the set of five leaf vertices. Then

$$D[S] = \begin{bmatrix} 0 & 2 & 3 & 4 & 4 \\ 2 & 0 & 3 & 4 & 4 \\ 3 & 3 & 0 & 3 & 3 \\ 4 & 4 & 3 & 0 & 2 \\ 4 & 4 & 3 & 2 & 0 \end{bmatrix}$$

and

$$\det D[S] = 864 = (-1)^4 2^3 \left(7 \cdot 21 - (14 \cdot 1^2 + 4 \cdot 2^2 + 1 \cdot 3^2)\right)$$

7. Further work

richman-shokrieh-wu

See [6].

7.1. **Symanzik polynomials.**

Acknowledgements

The authors would like to thank Ravindra Bapat for helpful discussion.

References

kirkland-neumann [1] R. Bapat, S. J. Kirkland, and M. Neumann. On distance matrices and Laplacians. *Linear Algebra Appl.*, 401:193–209, 2005.

-sivasubramanian [2] R. B. Bapat and S. Sivasubramanian. Identities for minors of the Laplacian, resistance and distance matrices. *Linear Algebra Appl.*, 435(6):1479–1489, 2011.

graham-lovasz [3] R. L. Graham and L. Lovász. Distance matrix polynomials of trees. *Adv. in Math.*, 29(1):60–88, 1978.

graham-pollak [4] R. L. Graham and H. O. Pollak. On the addressing problem for loop switching. *Bell System Tech. J.*, 50:2495–2519, 1971.

moon [5] J. W. Moon. *Counting labelled trees*. Canadian Mathematical Monographs, No. 1. Canadian Mathematical Congress, Montreal, Que., 1970. From lectures delivered to the Twelfth Biennial Seminar of the Canadian Mathematical Congress (Vancouver, 1969).

hman-shokrieh-wu [6] D. H. Richman, F. Shokrieh, and C. Wu. Capacity on metric graphs, 2022. in preparation.