MINORS OF TREE DISTANCE MATRICES

HARRY RICHMAN, FARBOD SHOKRIEH, AND CHENXI WU

Abstract. We prove that the principal minors of the distance matrix of a tree satisfy a combinatorial expression involving counts of rooted spanning forests of the underlying tree. This generalizes a result of Graham and Pollak. We also give such an expression for the case of edge-weighted trees.

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*

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

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.

Date: v1, February 9, 2023 (Preliminary draft, not for circulation).

For definitions of $(S,*)$-rooted spanning forests and other terminology, see Section 2. 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. The relation (1) has an analogue for the weighted distance matrix,

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

which 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*

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

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, $w(\overline{T})$ and $w(\overline{F})$ 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 reduces to Theorem 1 when taking all unit weights, $\alpha_e = 1$.

1.2. **Applications.** Suppose we fix a tree distance matrix D. It is natural to ask, how do the expressions $\det D[S]$ vary as we vary the vertex subset S? To our knowledge there is no nice behavior among the determinants, but as S varies there is nice behavior of the "normalized" ratios $\det D[S]/\operatorname{cof} D[S]$ which we describe here.

Given a matrix A, let $\operatorname{cof} A$ denote the *sum of cofactors* of A, i.e.

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

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

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

Using the Bapat–Sivasubramanian identity (5), an immediate corollary to Theorem 2 is the following result:

$$\frac{\det D[S]}{\operatorname{cof} D[S]} = \frac{1}{2} \left(\sum_{e\in E} \alpha_e - \frac{\sum_{F\in\mathcal{F}_2(G;S)} w(\overline{F}) (\deg^o(F,*) - 2)^2}{\sum_{T\in\mathcal{F}_1(G;S)} w(\overline{T})} \right). \tag{6}$$

The expression (6) satisfies a monotonicity condition as we vary the vertex set $S \subset V(G)$.

Theorem 3 (Monotonicity of normalized principal minors)**.** *If $A, B \subset V(G)$ are nonempty subsets with $A \subset B$, then*

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

The essential observation behind this result is that $\det D[S]/\operatorname{cof} D[S]$ is calculated via the following quadratic optimization problem: for all vectors $\mathbf{m} \in \mathbb{R}^S$,

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

This result can be shown using Lagrange multipliers, and relies of knowledge of the signature of $D[S]$. For details, see Section 3.

If $S \subset V(G)$ is nonempty, the expression (6) immediately implies the bound

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

We get refined bounds by making use of the monotonicity property, Theorem 3.

Theorem 4 (Bounds on principal minor ratios)**.** *Suppose $G = (V, E)$ is a finite, weighted tree with distance matrix D.*

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

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

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

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

1.3. **Further questions.** A formula for the inverse matrix D^{-1} was found by Graham and Lovász in [3]. Does there exist a nice expression for the inverse of the matrix $D[S]$?

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
$\mathcal{F}_1(G; S)$ 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 Tutte [**?**, Chapter VI].

Let $G = (V, E)$ be a graph 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 H to also denote its subset of edges $E(H)$, so e.g. $w(\overline{H}) = w(\overline{E(H)})$.

Let M be an $n \times n$ matrix. For a subset $S \subset \{1, \dots, n\}$, let $M[S]$ denote the submatrix obtained by keeping the S-indexed rows and columns of M. Let $M[\overline{S}]$ denote the submatrix obtained by deleting the S-indexed rows and columns.

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. Let $\kappa(G)$ denote the number of spanning forests of G, and let $\kappa_r(G)$ denote the number of r-component spanning forests.

Given a set of vertices $S = \{v_1, v_2, \dots, v_k\}$, 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*, following terminology in [**?**]. Given $s \in S$ and a forest F in $\mathcal{F}_1(G;S)$ or $\mathcal{F}_2(G;S)$, we let $F(s)$ denote the s-component of F, and let $F(*)$ denote the floating component (if $F \in \mathcal{F}_2$).

Let $\kappa(G;S)$ denote the number of S-rooted spanning forests of G, and let $\kappa_2(G;S)$ denote the number of $(S,*)$-rooted spanning forests. Let $\mathcal{F}_1(G;S)$ denote the set of S-rooted spanning forests of G, and let $\mathcal{F}_2(G;S)$ denote the set of $(S,*)$-rooted spanning forests of G.

Let

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

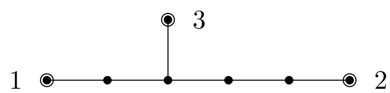
denote the number of k-component spanning trees which have a vertex v_i in each component. If $S = \{v_1, \ldots, v_k\}$, then $\kappa_k(v_1|\cdots|v_k) = \kappa(G;S) = \kappa(G/S)$.

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 5. 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.

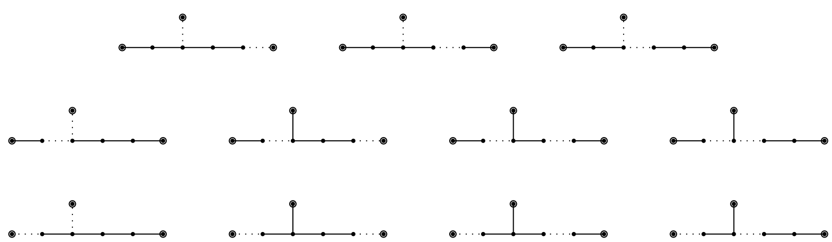


FIGURE 1. Forests in $\mathcal{F}_1(G;S)$.

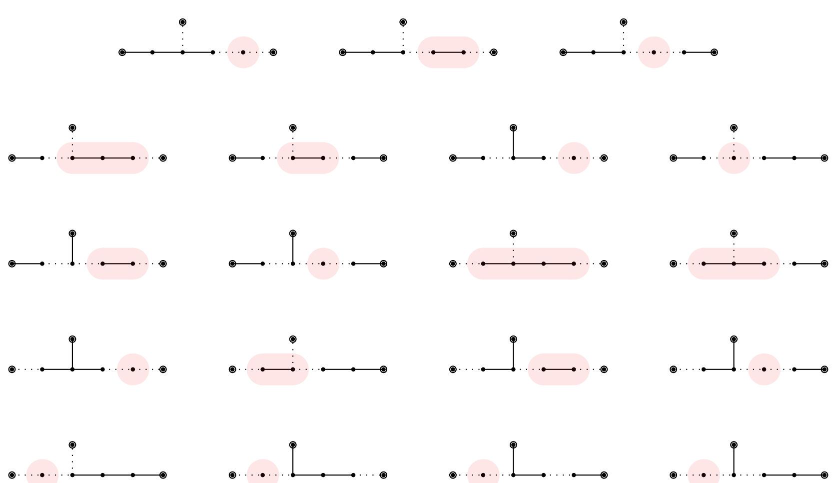


FIGURE 2. Forests in $\mathcal{F}_2(G;S)$, with floating component highlighted.

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.

Definition 6 (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

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

Given $S \subset V$, let $L[\overline{S}]$ denote the matrix obtained from L by removing the rows and columns indexed by S. For any graph G, let $\kappa(G)$ denote the number of spanning trees of G. The following theorem related minors of the (weighted) Laplacian to (weighted) counts of rooted spanning forests.

Theorem 7 (Principal-minors matrix tree theorem)**.** *Let $G = (V, E)$ be a finite graph.*

(a) Let L denote the Laplacian matrix of G. Then for any nonempty vertex set $S \subset V$,

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

(b) Let L_α denote the weighted Laplacian matrix of G, with edge weights $\{\alpha_e\}$. Then for any nonempty vertex set $S \subset V$,

$$\det L_\alpha[\overline{S}] = \sum_{T\in\mathcal{F}_1(G;S)} w(T)^{-1} = \sum_{T\in\mathcal{F}_1(G;S)} w(\overline{T}) \prod_{e\in E} \alpha_e^{-1}.$$

Proof. See Tutte [**?**, Section VI.6, Equation (VI.6.7)] or Chaiken [**?**] or Bapat [**?**, Theorem 4.7]. □

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.

Next we recall a connection between minors of the Laplacian matrix and cofactor sums of the distance matrix, when G is a tree. The result is due to Bapat–Sivasubramanian [2]. Recall that $\operatorname{cof} M$ denotes the *sum of cofactors* of M, i.e. $\operatorname{cof} M = \sum_{i=1}^n \sum_{j=1}^n (-1)^{i+j} \det M[\overline{i}, \overline{j}]$.

Theorem 8 (Distance submatrix cofactor sums)**.** *Given a tree G with edge weights, let D be the weighted distance matrix of G, and L the weighted Laplacian matrix of G. Let $S \subset V(G)$ be a nonempty subset of vertices of G. Then*

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

Proof. Bapat and Sivasubramanian [2, Theorem 11] show that

$$\operatorname{cof} D[S] = (-2)^{|S|-1} \left(\prod_{e\in E} \alpha_e \right) \det L[\overline{S}].$$

Then combine this equation with the matrix tree theorem, Theorem 8 (b). □

2.3. **Tree splits and tree distance.** In this section we describe the tree splits associated to a tree, and use their associated indicator functions to give an expression for the tree distance.

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^+, and let $(G \setminus e)^-$ denote the other component. 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 $(G \setminus e)^{\overline{v}}$ for the component not containing v.

Tree splits can be used to express the path distance between vertices in a tree. Given an edge $e \in E$ and vertices $v, w \in V$, let

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

In other words, $\delta(e; v, w) = 1$ if the vertices are in different components of the split $G \setminus e$, and $\delta(e; v, w) = 0$ if they are in the same component. Note that $\delta(e; v, v) = 0$ for any e and v.

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

(i) If we fix e and v, then $\delta(e; v, -) : V(G) \to \{0, 1\}$ is the indicator function for the component $(G \setminus e)^{\overline{v}}$ of the tree split $G \setminus e$ not containing v.

(ii) On the other hand if we fix v and w, then $\delta(-; v, w) : E(G) \to \{0, 1\}$ is the indicator function for the unique $v \sim w$ path in G.

Proposition 9 (Weighted tree distance)**.** *For a tree $G = (V, E)$ with weights $\{\alpha_e : e \in E\}$, the weighted distance function satisfies*

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

For an unweighted tree, we can express the tree distance $d(v, w)$ as the unweighted sum

$$d(v, w) = \sum_{e \in E(G)} \delta(e; v, w).$$

2.4. **Outdegree of rooted forest.** Given a vertex v in a graph, the *degree* $\deg(v)$ is the number of edges incident to v. A consequence of the "handshake lemma" of graph theory is that for any tree G, we have

$$\sum_{v \in V(G)} (2 - \deg(v)) = 2.$$

In this section we discuss some generalizations, which will be used later.

Given a connected subgraph $H \subset G$, we define the *outdegree* $\deg^o(H)$ as the number of edges which join H to its complement; i.e.

$$\deg^o(H) = \#\{e = (a, b) \in E : a \in V(H),\, b \notin V(H)\}.$$

Lemma 10. *Suppose G is a tree.*

(a) If $H \subset G$ is a (nonempty) connected subgraph, then

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

(b) If e is a fixed edge and u is a fixed vertex of G, then

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

Proof. (a) 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 $\deg^o(H) = \deg(v)$.

(b) Using part (a), we have

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

Since the component $(G \setminus e)^{\overline{u}}$ of the tree split $G \setminus e$ has a single edge separating it from its complement, $\deg^o((G \setminus e)^{\overline{u}}) = 1$ □

OLD: Given a rooted forest F in $\mathcal{F}_1(G;S)$ and $s \in S$, let $F(s)$ denote the s-component of F. We define the *outdegree* $\deg^o(F,s)$ as the number of edges which join $F(s)$ to a different component; i.e.

$$\deg^o(F,s) = \#\{e = (a,b) \in E : a \in F(s),\, b \notin F(s)\}. \tag{7}$$

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

2.5. **Miscellaneous.** Later we will make use of the fact that a submatrix $D[S]$ of a distance matrix has nonzero determinant, as long as $|S| \geq 2$.

We first recall a result of Cauchy, which states that the eigenvalues of $M[\bar{i}]$ "interlace" the eigenvalues of M. Recall that $M[\bar{i}]$ denotes the matrix obtained from M by deleting the i-th row and column.

Proposition 11 (Cauchy interlacing)**.** *Suppose M is a symmetric real matrix with ordered eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$, and the submatrix $M[\bar{i}]$ has ordered eigenvalues $\mu_1 \leq \cdots \leq \mu_{n-1}$. Then*

$$\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \cdots \leq \mu_{n-1} \leq \lambda_n.$$

Proof. See Horn–Johnson [**?**, Theorem 4.3.17]. □

Lemma 12 (Bapat [**?**, Lemma 8.15])**.** *Suppose D is the (weighted) distance matrix of a tree with n vertices. Then D has one positive eigenvalue and $n-1$ negative eigenvalues.*

Proof. Lemma 8.15 of [**?**] is stated for a non-weighted distance matrix; however, the same argument applies to a weighted distance matrix by applying Bapat–Kirkland–Neumann's result (3) on the weighted distance matrix determinant [1, Corollary 2.5]. □

The following argument was communicated to the author by Bapat, via personal communication.

Proposition 13. *Suppose D is the distance matrix of a tree $G = (V,E)$ and $S \subset V$ is a subset of size $|S| \geq 2$. Then*
(a) $D[S]$ has one positive eigenvalue and $|S|-1$ negative eigenvalues;
(b) $\det D[S] \neq 0$.

Proof. (a) Let $n = |V|$; assume $n \geq 3$. We apply (decreasing) induction on the size of S. If $|S| = n-1$, then Lemma **??** and Cauchy interlacing imply that $D[S]$ has at most one positive eigenvalue, and at least one negative eigenvalue. Since all diagonal entries of $D[S]$ are zero, $D[S]$ has zero trace. Thus $D[S]$ has exactly one positive eigenvalue as claimed. The same argument applies for smaller S, as long as $|S| \geq 2$.

(b) This follows from (a), since the determinant is the product of eigenvalues. □

Remark 14. A key step in the main 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 15. 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

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

For a forest T in $\mathcal{F}_1(G; S)$, the preimage under this map has $|E(T)|$-many elements.

3. Optimization: quadratic programming

In this section, we explain how the quantity $\dfrac{\det D[S]}{\operatorname{cof} D[S]}$ arises as the solution of a quadratic optimization problem.

Proposition 16. *If $D[S]$ is a principal submatrix of a 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 of* $D[S]$*.*

Proof. If $|S| = 1$ then $D[S]$ is the zero matrix and the statement is true trivially.

♢ TODO show that objective function is concave on domain $\mathbf{1}^\intercal\mathbf{m} = 1$ ♢

Now assume $|S| \geq 2$. 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, since $D[S]$ is invertible (Proposition **??**) 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 $\lambda = \dfrac{\det D[S]}{\operatorname{cof} D[S]}$. □

Corollary 17. *If $D[S]$ is a principal submatrix of a distance matrix indexed by S, then*

$$\frac{\det D[S]}{\operatorname{cof} D[S]} = \max\{\mathbf{m}^\intercal D\mathbf{m} : \mathbf{m} \in \mathbb{R}^V,\ \mathbf{1}^\intercal \mathbf{m} = 1,\ \mathbf{m}_v = 0 \textit{ if } v \notin S\}$$

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

Proof of Theorem 3. We are to show that for vertex subsets $A \subset B$, we have $\dfrac{\det D[A]}{\operatorname{cof} D[A]} \leq \dfrac{\det D[B]}{\operatorname{cof} D[B]}$. By Corollary **??**, both values $\dfrac{\det D[A]}{\operatorname{cof} D[A]}$ and $\dfrac{\det D[B]}{\operatorname{cof} D[B]}$ arise from optimizing the same objective function, but the constraint for A is more strict. □

Proof of **??***.* (a) To see that

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

take B as the set of vertices in $\operatorname{conv}(S, G)$. Then $S \subset B$, and

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

(b) To see that

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

take A as the set of endpoints of γ. (where If γ is a simple path between vertices $s_0, s_1 \in S$) Then $A \subset S$ by assumption, and

$$\frac{\det D[A]}{\operatorname{cof} D[A]} = \frac{1}{2} \sum_{e \in \gamma} \alpha_e.$$

□

4. Distance submatrices: Proofs

In this section we prove our main result, Theorem 2. Theorem 1 follows as an immediate corollary.

4.1. **Outline of proof.** Given a subset $S \subset V$ and distance submatrix $D[S]$, we will do the following steps.

(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. $\mathbf{1}^\intercal \mathbf{m}$, and normalize

$$\mathbf{m}^* = \frac{\mathbf{m}}{\mathbf{1}^\intercal \mathbf{m}}.$$

This solves the optimization problem of Section 3.

(iii) The optimal objective value $(\mathbf{m}^*)^\intercal D[S]\mathbf{m}^* = \lambda^*$ is

$$\lambda^* = \frac{\lambda}{\mathbf{1}^\intercal \mathbf{m}}.$$

(iv) Using Theorem **??**,

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

where $\operatorname{cof} D[S]$ is the sum of cofactors of $D[S]$. Use expression for $\operatorname{cof} D[S]$ in Theorem **??** to compute $\det D[S]$.

It turns out that the entries of $\mathbf{m}$ are combinatorially meaningful, which also gives combinatorial meaning to the constant λ.

Example 18. 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 $\mathbf{1}^\intercal \mathbf{m} = 2$, so $\mathbf{m}^* = \frac{1}{2}\mathbf{m}$.

(iii) We have

$$\lambda^* = \frac{\lambda}{\mathbf{1}^\intercal \mathbf{m}} = \frac{a+b+c}{2}.$$

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

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

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

Definition 19. Let $\mathbf{m} = \mathbf{m}(G; S)$ denote the vector in $\mathbb{R}^S$ be defined by

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

where $\deg^o(T, v)$ is the outdegree of the v-component of T, (7).

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

Proposition 20. *For the vector* $\mathbf{m}$ *defined above,*

(a) $\mathbf{1}^\intercal \mathbf{m} = 2 \sum_{T \in \mathcal{F}_1(G;S)} w(\overline{T})$*;*

(b) if all edge weights α_e *are positive,* $\mathbf{m}$ *is nonzero.*

Proof. (a) By Lemma 11 we have

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

Then exchange the order of summation in $\mathbf{1}^\intercal \mathbf{m}$,

$$\begin{aligned} \mathbf{1}^\intercal \mathbf{m} = \sum_{s \in S} \mathbf{m}_s &= \sum_{s \in S} \left(\sum_{T \in \mathcal{F}_1(G;S)} w(\overline{T}) \sum_{v \in T(s)} (2 - \deg(v)) \right) \\ &= \sum_{T \in \mathcal{F}_1(G;S)} w(\overline{T}) \left(\sum_{s \in S} \sum_{v \in T(s)} (2 - \deg(v)) \right). \end{aligned}$$

Next observe that the inner double sum is simply a sum over $v \in V$, since the vertex sets of $T(s)$ for $s \in S$ form a partition of V by definition of S-rooted spanning forest. Thus

$$\mathbf{1}^\intercal \mathbf{m} = \sum_{T \in \mathcal{F}_1} w(\overline{T}) \left(\sum_{v \in V} (2 - \deg(v)) \right) = \sum_{T \in \mathcal{F}_1} w(\overline{T}) \cdot 2$$

where we again apply Lemma 11 for the last equality, as $\deg^o(G) = 0$.

(b) If all edge weights are positive, then $w(\overline{T}) > 0$ for all T, and $\mathcal{F}_1(G; S)$ is nonempty as long as S is nonempty. Thus part (a) implies that $\mathbf{1}^\intercal \mathbf{m} > 0$. □

Corollary 21. *If* G *is a graph with unit edge weights* $\alpha_e = 1$*, then the vector* $\mathbf{m}$ *defined in* (9) *satisfies* $\mathbf{1}^\intercal \mathbf{m} = 2\,\kappa(G; S)$*.*

Theorem 22. *With* $\mathbf{m} = \mathbf{m}(G; S)$ *defined as in* (9)*,* $D[S]\mathbf{m} = \lambda \mathbf{1}$ *for the constant*

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

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

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

where in the last equality, we apply Lemma 11 to the subgraph $H = T(s)$.

We introduce additional notation to handle the double sum in parentheses in (10). 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,

$$(D[S]\mathbf{m})_v = \sum_{T\in\mathcal{F}_1} w(\overline{T}) \sum_{e\in E}\alpha_e \left(\sum_{u\in V}(2-\deg(u))\delta(e;v,\pi_T(u))\right) \qquad (10)$$

We will compare the above expression with the one obtained after replacing $\delta(e; v, \pi_T(u))$ with $\delta(e; v, u)$. From Lemma 11 (b), we have $\sum_{u\in V}(2-\deg(u))\delta(e;v,u) = 1$. Thus

$$\sum_{T\in\mathcal{F}_1} w(\overline{T}) \sum_{e\in E}\alpha_e = \sum_{T\in\mathcal{F}_1} w(\overline{T}) \sum_{e\in E}\alpha_e \left(\sum_{u\in V}(2-\deg(u))\delta(e;v,u)\right) \qquad (11)$$

By subtracting equation (12) from (11),

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

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 $G \setminus e$. We have

$$\delta(e;v,\pi_T(u)) - \delta(e;v,u) = \begin{cases} 0 & \text{if } \delta(e;\pi_T(u),u) = 0 \\ 1 & \text{if } \delta(e;\pi_T(u),u) = 1 \text{ and } \delta(e;v,\pi_T(u)) = 1 \\ -1 & \text{if } \delta(e;\pi_T(u),u) = 1 \text{ and } \delta(e;v,\pi_T(u)) = 0 \end{cases} \qquad (12)$$

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 $T \setminus e$. See Figure 3, left.

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 $T \setminus e$. See Figure 3, right.

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 when multiplying the term (13) by $(2 - \deg(u))$ and summing over all vertices u, we obtain

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

Thus

$$(13) \quad (D[S]\mathbf{m})(v) - \sum_{T \in \mathcal{F}_1} w(\overline{T}) \sum_{e \in E} \alpha_e$$
$$= \sum_{T \in \mathcal{F}_1} w(\overline{T}) \sum_{s_0 \in S} \left(\sum_{\substack{e \in T(s_0) \\ \delta(e; v, s_0) = 1}} \alpha_e (2 - \deg^o(T \setminus e, *)) - \sum_{\substack{e \in T(s_0) \\ \delta(e; v, s_0) = 0}} \alpha_e (2 - \deg^o(T \setminus e, *)) \right)$$
$$= \sum_{T \in \mathcal{F}_1} w(\overline{T}) \sum_{e \in T} \alpha_e (2 - \deg^o(T \setminus e, *)) \left(\sum_{\substack{e \in T(s_0) \\ \delta(e; v, s_0) = 1}} 1 - \sum_{\substack{e \in T(s_0) \\ \delta(e; v, s_0) = 0}} 1 \right).$$

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, and that the corresponding weights satisfy

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

Thus

$$(14) = \sum_{F \in \mathcal{F}_2} w(\overline{F}) (2 - \deg^o(F, *)) \sum_{T \in \mathcal{F}_1} \sum_{s_0 \in S} \left(\sum_{\substack{e \in T(s_0) \\ \delta(e; v, s_0) = 1}} \mathbb{1}(F = T \setminus e) - \sum_{\substack{e \in T(s_0) \\ \delta(e; v, s_0) = 0}} \mathbb{1}(F = T \setminus e) \right)$$
$$= \sum_{F \in \mathcal{F}_2} w(\overline{F}) (2 - \deg^o(F, *)) \Big(\#\{T \in \mathcal{F}_1 : F = T \setminus e \text{ for some } e \in T(s_0),\ \delta(e; v, s_0) = 1\}$$
$$- \#\{T \in \mathcal{F}_1 : F = T \setminus e \text{ for some } e \in T(s_0),\ \delta(e; v, s_0) = 0\} \Big)$$

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

$$(14) = \sum_{F \in \mathcal{F}_2} w(\overline{F})(2 - \deg^o(F, *)) \Bigg(\#\{e \in \partial F(*) : \delta(e; v, F(*)) = 0\} \\ - \#\{e \in \partial F(*) : \delta(e; v, F(*)) = 1\} \Bigg).$$

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. Hence

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

Thus the previous expression (14) simplifies as

$$\begin{aligned} (14) &= \sum_{F \in \mathcal{F}_2} w(\overline{F})(2 - \deg^o(F, *)) \Big((\deg^o(F, *) - 1) - (1) \Big) \\ &= - \sum_{F \in \mathcal{F}_2} w(\overline{F})(2 - \deg^o(F, *))^2. \end{aligned}$$

as desired. □

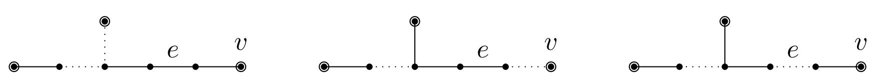


FIGURE 5. Components rooted in $S(G \setminus e)^{\overline{v}}$.

Finally we can prove our main theorem: for any nonempty subset $S \subset V(G)$,

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

Proof of Theorem 2. First, suppose $|S| = 1$. Then $D[S]$ is the zero matrix, and we must show that the right-hand side is zero. Since G is a tree, $\mathcal{F}_1(G; \{v\})$ consists of the tree G itself, with co-weight $w(\overline{G}) = 1$. Moreover, the subgraphs in $\mathcal{F}_2(G; \{v\})$ are precisely the tree splits $G \setminus e$, and for each $F = G \setminus e$ we have $w(\overline{F}) = \alpha_e$ and $\deg^o(F, *) - 2 = -1$. This shows that the right-hand size of (**??**) is zero.

Next, suppose $|S| \geq 2$. Proposition **??** states that $D[S]$ is nonsingular, so we may use the inverse matrix identity

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

Let $\mathbf{m} = \mathbf{m}(G; S)$ denote the vector (9). By Proposition **??** (a) and Theorem **??**,

$$\mathbf{1}^\intercal \mathbf{m} = 2 \sum_{T \in \mathcal{F}_1(G;S)} w(\overline{T}) = \frac{\operatorname{cof} D[S]}{(-1)^{1-|S|} 2^{2-|S|}}.$$

Theorem **??** states that $D[S]\mathbf{m} = \lambda\mathbf{1}$ for some constant λ, which is nonzero since $D[S]$ is invertible and $\mathbf{m}$ is nonzero, c.f. Proposition **??** (b). Hence

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

Comparing (**??**) with (**??**) gives the desired result, $\det D[S] = (-1)^{|S|-1}2^{|S|-1}\lambda$. □

Proof of Theorem 1. Set all weights α_e to 1 in Theorem 2. In this case, the weights $w(\overline{T}) = 1$ and $w(\overline{F}) = 2$ for all forests T and F, and

$$\sum_{e\in E} \alpha_e = n-1, \qquad \sum_{T\in\mathcal{F}_1(G;S)} w(\overline{T}) = \kappa_1(G;S).$$ □

Remark 23. 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 $\operatorname{conv}(S, G)$ 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 $\operatorname{conv}(S, G)$. However, the formulas as stated are true even without this replacement due to cancellation of terms.

5. Physical interpretation

If we consider G as a network of wires with edge e containing a resistor of resistance α_e, which is grounded at all nodes in S, then $\mathbf{m}_S$ has an interpretation as current flow: it records the currents flowing to S when current enters the vertices in the amount $\deg(v) - 2$ 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 let π_S denote the projection from $\mathbb{R}^V$ to $\mathbb{R}^S$. We wish to find vectors $\mathbf{n}_i \in \mathbb{R}^V$ satisfying $\pi_S(D\mathbf{n}_i) = \lambda_i\mathbf{1}$. OR, $D\mathbf{n}_i = \lambda_i\mathbf{1}^S \oplus (-)$. By finding sufficiently many such vectors $\mathbf{n}_i$, we can hope to find a linear combination that lies inside $\mathbb{R}^S \oplus \{0\}$.

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

$$\mathbf{n}_u = \begin{cases} 1 & \text{if } u = v, \\ -\mu(v,u) & \text{if } u \in S, \\ 0 & \text{otherwise.} \end{cases}$$

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

Explicitly,

$$\mu(v,s) = \frac{\#\text{ of } S\text{-rooted spanning forests of } G \text{ whose } s_j\text{-component contains } v}{\#\text{ of } S\text{-rooted spanning forests of } G}$$

$$= \frac{\sum_{\mathcal{F}_1(G/S)} \mathbb{1}(v \in T(s))}{\kappa(G/S)} = \frac{\sum_{\mathcal{F}_1(G/S)} \mathbb{1}(\pi_T(v) = s)}{\kappa(G/S)}$$

If $s = s_j$, then

$$\mu(v, s_j) = \frac{\kappa_r(s_1|\cdots|s_j v|\cdots|s_r)}{\kappa_r(s_1|\cdots|s_r)}$$

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'. $\square$

Theorem 25. *Let G be a tree, S a nonempty subset of vertices, and $D[S]$ the corresponding submatrix of the distance matrix of G. Suppose $\mathbf{m} = \mathbf{m}(G;S) \in \mathbb{R}^S$ is defined by* (9)*;*

$$\mathbf{m}(G;S)_s = \sum_{T\in\mathcal{F}_1(G;S)} w(\overline{T}) \sum_{u\in T(s)} (2-\deg u) = \sum_{T\in\mathcal{F}_1(G;S)} w(\overline{T})(2-\deg^o(T,s)).$$

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

$$\mathbf{m}(G;S) = \sum_{T\in\mathcal{F}_1(G;S)} w(\overline{T}) \sum_{u\in V}(2-\deg(u))\boldsymbol{\delta}(\pi_T(u)).$$

Therefore

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

$\diamondsuit$ TODO: elaborate on this equation $\diamondsuit$ From Proposition 21 we know that $D\mathbf{m}(G;V)$ is constant on V, and from Proposition 27 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. $\square$

Example 26. 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

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

5.2. **Symanzik polynomials.** We note that the expression in the main theorem, Theorem 2, is related to Symanzik polynomials, which we recall here.

Given a graph $G = (V,E)$, the *first Symanzik polynomial* is the homogeneous polynomial in edge-indexed variables $\underline{x} = \{x_e : e \in E\}$ defined by

$$\psi_G(\underline{x}) = \sum_{T\in\mathcal{F}_1(G)} \prod_{e\notin T} x_e,$$

where $\mathcal{F}_1(G)$ denotes the set of spanning trees of G.

Consider a "momentum" function $p : V \to \mathbb{R}$ which satisfies $\sum_{v\in V} p(v) = 0$. Then the second Symanzik polynomial is

$$\varphi_G(p;\underline{x}) = \sum_{F\in\mathcal{F}_2(G)} \left(\sum_{v\in F_1} p(v)\right)^2 \prod_{e\notin F} x_e,$$

where $\mathcal{F}_2(G)$ is the set of two-component spanning forests of G, and F_1 denotes one of the components of F. It doesn't matter which component we label as F_1, due to the momemtum constraint $\sum_{v\in V} p(v) = 0$.

Theorem statement:

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

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

In terms of Symanzik polynomials, let ψ and φ denote the first and second Symanzik polynomials of the quotient graph G/S. We have

$$\det D[S] = (-1)^{|S|-1} 2^{|S|-2} \left(\sum_{E(G)} \alpha_e \, \psi_{(G/S)}(\underline{\alpha}) - \phi_{(G/S)}(p; \underline{\alpha}) \right).$$

and

$$\frac{\det D[S]}{\operatorname{cof} D[S]} = \frac{1}{2} \left(\sum_{e \in E} \alpha_e - \frac{\varphi_{(G/S)}(p; \underline{\alpha})}{\psi_{(G/S)}(\underline{\alpha})} \right)$$

with momentum function $p(v) = \deg(v) - 2$ for $v \notin S$.

5.3. Warmup case: $S = V$.

Proposition 27. *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, $|E| = |V| - 1$. □

Proposition 28. *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

$$\begin{aligned}
(D\mathbf{m})_{(e^+)} - (D\mathbf{m})_{(e^-)} &= \sum_{v \in V} (d(v, e^+) - d(v, e^-))(2 - \deg v) \\
&= \sum_{v \in (G \setminus e)^-} \alpha_e (2 - \deg v) - \sum_{v \in (G \setminus e)^+} \alpha_e (2 - \deg v) \\
&= \alpha_e \left(\sum_{v \in (G \setminus e)^-} (2 - \deg v) - \sum_{v \in (G \setminus e)^+} (2 - \deg v) \right) \qquad (14)
\end{aligned}$$

since

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

For each sum in (8), we apply Lemma 11 to obtain

$$\sum_{v \in (G \setminus e)^-} (2 - \deg v) = (2 - \deg^o((G \setminus e)^-) = 1,$$

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

6. Examples

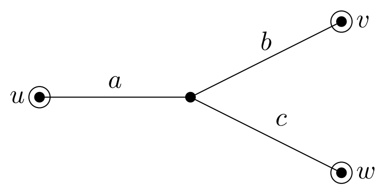
Example 29. Suppose G is a tree consisting of three paths joined at a central vertex. Let S consist of the central vertex, and the three endpoints of the paths. The corresponding minor of the distance matrix is

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

The determinant is

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

Example 30. Suppose Γ is a tripod with lengths a, b, c and corresponding leaf vertices u, v, w.



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

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

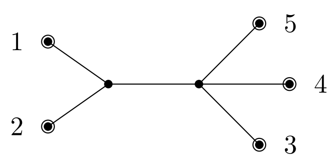
and

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

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

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

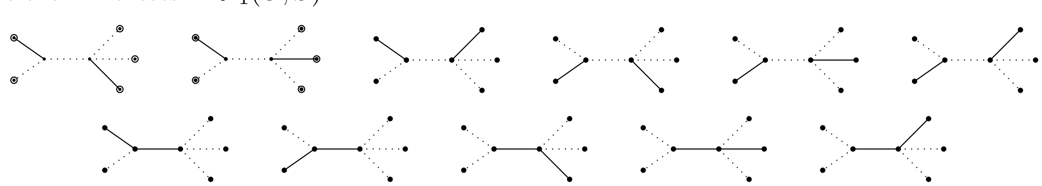
Example 31. Suppose G is the tree with unit edge lengths shown below, with five leaf vertices.



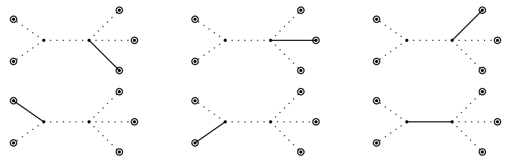
Let S denote the set of five leaf vertices. Then

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

There are 11 forests in $\mathcal{F}_1(G; S)$:



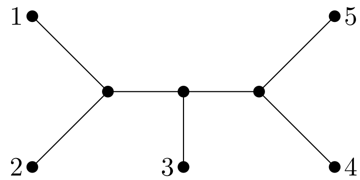
There are 6 forests in $\mathcal{F}_2(G;S)$:



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 32. Suppose G 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

It is natural to ask whether these results for trees may be generalized to arbitrary finite graphs. We address this in [6], which involve more technical machinery.

See [**?**].

Acknowledgements

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

References

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

[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.

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

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

[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).

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