PRINCIPAL 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 trees with edge lengths. We use arguments motivated by potential theory on graphs. Our formulas can be expressed in terms of evaluations of Symanzik polynomials.

Contents

1. Introduction

Suppose $G = (V, E)$ is a tree with n vertices. Let D denote the *distance matrix* of G, defined by setting the (u, v)-entry to the length of the unique path from u to v. In [GP71], Graham and Pollak prove

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

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 $S \subseteq V$, let $D[S]$ denote the submatrix consisting of the S-indexed rows and columns of D.

Date: November 19, 2024.
2020 *Mathematics Subject Classification.* 05C50, 05C05, 05C12, 05C30, 31C15 .

Theorem A. *Suppose G is a tree with n vertices whose distance matrix is D. Let $S \subseteq V$ be a nonempty subset of vertices. Then*

$$\det D[S] = (-1)^{|S|-1}2^{|S|-2}\Big((n-1)\,\kappa(G;S) - \sum_{F\in\mathcal{F}_2(G;S)} (\mathrm{outdeg}(F,*)-2)^2\Big),$$

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 $\mathrm{outdeg}(F,*)$ *denotes the outdegree of the floating component of F.*

For definitions of rooted spanning forests as well as other terminology, see §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 $\kappa(G;V) = 1$. Moreover, the set $\mathcal{F}_2(G;V)$ of $(V,*)$-rooted spanning forests is empty. Thus Theorem A recovers the Graham–Pollak identity (1).

1.1. **Trees with edge lengths.** Assume the tree G is endowed with a collection of positive real edge lengths $\alpha = \{\alpha_e \colon e \in E\}$. The *distance matrix* (with respect to α), which we will again denote by D, is defined by setting the (u,v)-entry to the sum of the edge lengths α_e along the unique path from u to v. The relation (1) has a generalization for the distance matrix of a tree with edge lengths, proved by Bapat, Kirkland, and Neumann ([BKN05, Corollary 2.5]):

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

We, in fact, prove the following more general result. For an edge subset $A \subseteq E$ let $w(A) = \prod_{e\notin A}\alpha_e$. (The quantity $w(A)$ is sometimes known as the "weight" of A, while other sources call it the "coweight.")

Theorem B. *Suppose $G = (V,E)$ is a finite tree with edge lengths $\{\alpha_e \colon e \in E\}$, and corresponding distance matrix D. For any nonempty subset $S \subseteq V$, we have*

$$\det D[S] = (-1)^{|S|-1}2^{|S|-2}\left(\sum_{e\in E}\alpha_e \sum_{T\in\mathcal{F}_1(G;S)} w(T) - \sum_{F\in\mathcal{F}_2(G;S)} w(F)\,(\mathrm{outdeg}(F,*)-2)^2\right), \tag{3}$$

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 weights of the forests T and F, and $\mathrm{outdeg}(F,*)$ *is the outdegree of the floating component of F, as above.*

See §2 for explanations of the other terminology used here.

Example. Suppose G is the tree with unit edge lengths shown in Figure 1, with five leaf vertices and three internal vertices. Let $S = \{1,2,3,4,5\}$ be the set of leaf vertices. The corresponding distance submatrix is

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

whose determinant is 864.

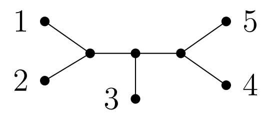


FIGURE 1. Tree with five leaves.

The tree G has 7 edges and 21 S-rooted spanning forests. There are 19 $(S,*)$-rooted spanning forests. Of the floating components in these forests, 14 have outdegree three, 4 have outdegree four, and 1 has outdegree five. By Theorem A,

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

1.2. **Normalized principal minors.** One may ask how the expressions $\det D[S]$ vary, as we fix a tree distance matrix and vary the vertex subset S. To our knowledge there is no nice behavior among these determinants but, as S varies there is nice behavior of a "normalized" version 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 the sum of entries of the matrix inverse A^{-1} multiplied by a factor of $\det A$, i.e. $\operatorname{cof} A = (\det A)(\mathbf{1}^\intercal A^{-1} \mathbf{1})$.

In [BS11], Bapat and Sivasubramanian show the following identity for the sum of cofactors of a distance submatrix $D[S]$ of a tree with edge lengths $\{\alpha_e \colon e \in E\}$,

$$\operatorname{cof} D[S] = (-2)^{|S|-1} \sum_{T \in \mathcal{F}_1(G;S)} w(T), \tag{4}$$

where $\mathcal{F}_1(G;S)$ denotes the set of S-rooted spanning forests of G (see §2 and Theorem 4.4). An immediate consequence of Theorem B, together with (4), is the identity

$$\frac{\det D[S]}{\operatorname{cof} D[S]} = \frac{1}{2} \left(\sum_{e \in E} \alpha_e - \frac{\sum_{F \in \mathcal{F}_2(G;S)} w(F)\, (\operatorname{outdeg}(F,*) - 2)^2}{\sum_{T \in \mathcal{F}_1(G;S)} w(T)} \right). \tag{5}$$

We will refer to $\det D[S]/\mathrm{cof}\, D[S]$ as the *normalized principal minor* corresponding to the vertex subset S. It turns out that normalized principal minors satisfy a monotonicity condition.

Theorem C (Monotonicity of normalized principal minors)**.** *If $A, B \subseteq V$ are nonempty vertex subsets with $A \subseteq B$, then*

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

The essential observation behind this result is an intriguing connection to the theory of quadratic optimization; we will show that the normalized principal minor $\det D[S]/\,\mathrm{cof}\, D[S]$ is the solution of the following optimization problem: for all vectors $\mathbf{u} \in \mathbb{R}^S$,

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

This result is obtained using the method of Lagrange multipliers, and relies on knowledge of the signature of $D[S]$. The proof is given in §4.

For a nonempty subset S, equation (5) implies the upper bound in

$$0 \leq \frac{\det D[S]}{\mathrm{cof}\, D[S]} \leq \frac{1}{2} \sum_{e \in E(G)} \alpha_e,$$

while the lower bound is implied by the quadratic optimization formulation. We get refined bounds by making use of the monotonicity property, Theorem C. Let $\mathrm{conv}(S, G)$ denote the convex hull of S, i.e. the subtree of G consisting of all paths between points of S.

Theorem D (Bounds on principal minor ratios)**.** *Suppose $G = (V, E)$ is a finite tree with edge lengths $\{\alpha_e \colon e \in E\}$ and with distance matrix D.*

(a) For a nonempty subset $S \subseteq V$,

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

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

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

The monotonicity result, Theorem C, was found independently by Devriendt [Dev22, Property 3.38], in the more general context of effective resistance matrices of graphs.

Remark. It is worth observing that the terms appearing in the submatrix $D[S]$ only depend on the lengths of edges which lie on some path between vertices in S. Depending on the chosen subset $S \subseteq V$, these distances in $D[S]$ may ignore a large part of the ambient tree G. However, the expression in Theorem B involes all edge lengths.

To calculate $\det D[S]$ in this case, using Theorem B "efficiently," 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. However, the formulas as stated are true even without this replacement due to cancellation of terms.

1.3. **Connection with Symanzik polynomials.** Recall, 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.

A "momentum function" is any $p \colon V \to \mathbb{R}$ satisfying the constraint $\sum_{v \in V} p(v) = 0$. The *second Symanzik polynomial* is

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

where $\mathcal{F}_2(G)$ is the set of two-component spanning forests of G and F_1 denotes one of the components of $F = F_1 \cup F_2$. Note that it does not matter which component we label as F_1, since the momentum constraint implies that $\sum_{v \in F_1} p(v) = -\sum_{v \in F_2} p(v)$.

We note that the expressions in Theorem B and (5) are closely related to Symanzik polynomials. Let $\psi_{(G/S)}$ and $\varphi_{(G/S)}$ denote the first and second Symanzik polynomials of the quotient graph G/S. Let p_{can} be the momentum function defined on G/S by $p_{\text{can}}(v) = \deg(v) - 2$ for $v \notin S$ and $p_{\text{can}}(v_S) = -\sum_{v \notin S} p_{\text{can}}(v)$, where v_S denotes the vertex image of S in the quotient.

The identity in Theorem B can be written as follows

$$\det D[S] = (-1)^{|S|-1} 2^{|S|-2} \left(\Big(\sum_{e \in E} \alpha_e \Big) \psi_{(G/S)}(\underline{\alpha}) - \phi_{(G/S)}(p_{\text{can}}; \underline{\alpha}) \right).$$

The "normalized" ratio (5) can be written using Symanzik polynomials as

$$\frac{\det D[S]}{\operatorname{cof} D[S]} = \frac{1}{2} \left(\sum_{e \in E} \alpha_e - \frac{\varphi_{(G/S)}(p_{\text{can}}; \underline{\alpha})}{\psi_{(G/S)}(\underline{\alpha})} \right). \tag{6}$$

1.4. **Related work.** The idea to study the ratio $(\det D / \operatorname{cof} D)$ appeared early on in the literature on distance matrices. In [GHH77], it is shown that this ratio is additive on wedge sums of graphs. This powerful observation is enough to essentially obtain the Graham–Pollak identity (1) and its weighted version (2) as corollaries.

Choudhury and Khare [CK24] prove a generalization of the Graham–Pollak identity (1) in which D can be replaced with certain principal minors, but crucially their generalization excludes all cases where $\det D[S]$ depends on the underlying tree structure. Namely, they consider precisely those minors $\det D[S]$ which can be expressed in terms of data assigned to edges (weights, edge counts, etc.), without any reference to what edges are incident to others. In contrast, Theorems A and B make evident use of the tree structure, via the spanning forests in $\mathcal{F}_1(G;S)$ and $\mathcal{F}_2(G;S)$.

It is natural to ask how our results for trees may be generalized to arbitrary finite graphs. We address this in an upcoming paper [RSW24], which involves more technical machinery. Certain ideas in this paper overlap with those developed concurrently and independently in [Dev22]. Some generalizations of our results, namely Theorem C, appear already in [Dev22, Property 3.38], which is concerned with effective resistances on graphs. On a tree, effective resistance is the same as (path-)distance. If D is the distance matrix of a tree, then any principal minor $D[S]$ is the effective resistance matrix of some (edge-weighted) graph on the vertex set S. The quantity $(\det D[S]/\operatorname{cof} D[S])$ is referred to as the *resistance radius* of S (up to a factor of 2), and the vector $\mathbf{m}$ used in §5 is called the *resistance curvature* (up to scaling) in Devriendt's terminology. The idea to consider an $\mathbf{m}$-like vector as a "curvature" also appears in [Ste23], which motivates the terminology by proving graph-theoretic analogues of some curvature-related theorems in differential geometry. The results in §3 regarding the signature of the matrices $D[S]$ can be considered special discrete cases of the *Dirichlet pairing* defined by Baker and Rumely [BR07].

Amini studies ratios of Symanzik polynomials in [Ami19], similar to the ratio appearing in (6), motivated by quantum field theory and the calculation of Feynman amplitudes [ABBGF16]. Amini's results apply for Symanzik polynomials on an arbitrary graph, in contrast to (6) where the revelant graph is a vertex-quotient of a tree. The first Symanzik polynomial can be computed using Kirchhoff's celebrated matrix-tree theorem. The second Symanzik polynomial can similarly be calculated via determinant of the generalized Laplacian matrix; see [Ami19, Section 1.1] and [Bro24, Theorem 7.1].

In [GL24], Gutiérrez and Lillo observe that Theorem A can be expressed as

$$\det D[S] = (-1)^{|S|-1}2^{|S|-2}\Big((|S|-1)\,\kappa(G;S) \\ - \sum_{F\in\mathcal{F}_2(G;S)} (\operatorname{outdeg}(F,*)-1)\,(\operatorname{outdeg}(F,*)-4) \Big)$$

through some straightforward algebraic manipulation, while also providing a new proof for this result using a nice combinatorial argument involving sign-reversing involutions on collections of paths in G.

The results in this paper may be of interest to those studying phylogenetics. In phylogenetics, one aims to find the tree that best represents the evolutionary history among a collection of organisms using biological data (e.g. DNA sequences). In this tree, leaf vertices represent modern-day species, while internal vertices represent ancestral species. There are standard methods for estimating pairwise distances between species along their evolutionary tree. This means we can often predict the distance submatrix $D[S]$ of the target tree, in which S is the set of leaf vertices, and we would like to use this information in reverse to decide what underlying tree best fits this distance data. Results such as Theorem C may lead to new tests for phylogenetic inference, or for evaluating tree instability [CBSMI24].

Structure of the paper. In §2 we review terminology and notation concerning graphs, spanning forests, Laplacian matrices, and degree-related identities. In §3 we recall some results on the spectral properties of the distance matrix and its principal submatrices. In §4 we show that the normalized principal minor ($\det D[S]/\operatorname{cof} D[S]$) is the solution to some quadratic optimization problem involving the bilinear form $D[S]$. This observaton is used to prove Theorem C (monotonicity) and Theorem D. In §5 we prove Theorems A and B, which give a combinatorial expression for $\det D[S]$ in terms of rooted spanning forests. We end the paper with some examples in §6 demonstrating Theorems A and B.

Acknowledgements. The authors would like to thank Ravindra Bapat for helpful correspondence, in particular related to the results in § 3.1. We also thank Matt Baker, Karel Devriendt, Apoorva Khare, and Sebastian Prillo for helpful discussions. We thank Álvaro Gutiérrez for helpful discussions regarding his follow-up to our work.

HR was supported by the Howard Hughes Medical Institute and the Matsen Group at the Fred Hutchinson Cancer Center. FS was partially supported by NSF CAREER DMS-2044564 grant. CW was partially supported by Simons Collaboration Grant 850685.

2. Graphs and spanning forests

Throughout, let $G = (V, E)$ be a finite graph with (positive, real) edge lengths $\{\alpha_e : e \in E\}$. A graph without edge lengths will be considered as a special case, with $\alpha_e = 1$ for all $e \in E$.

We assume all graphs in the paper are equipped with an implicit (arbitrary) orientation. This means we fix a pair of functions $\operatorname{head}: E \to V$ and $\operatorname{tail}: E \to V$, such that $\operatorname{head}(e)$ and $\operatorname{tail}(e)$ are the endpoints of e. We abbreviate $\operatorname{head}(e)$ as e^+, and $\operatorname{tail}(e)$ as e^-.

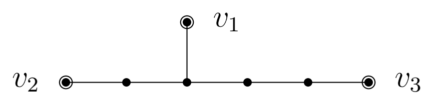
2.1. **Spanning trees and forests.** A *spanning tree* of a graph G is a subgraph which is connected, has no cycles, and contains all vertices of G. A *spanning forest* of a graph G is a subgraph which has no cycles and contains all vertices of G.

Given a set of vertices $S = \{v_1, v_2, \ldots, v_r\}$, an *$S$-rooted spanning forest* of G is a spanning forest which has exactly one vertex of S in each connected component. Given $s \in S$ and a forest F, we let $F(s)$ denote the *s-component* of F, that is, the component of F containing s.

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 [KKW15]. For an $(S,*)$-rooted spanning forest F, we let $F(*)$ denote the floating component. We sometimes refer to the floating component as the $*$-component of F. Again, for $s \in S$, we let $F(s)$ denote the s-component of F.

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.

Example 2.1. 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. Some of these are shown in Figure 2 and Figure 3, respectively.

FIGURE 2. Some forests in $\mathcal{F}_1(G;S)$.

FIGURE 3. Some forests in $\mathcal{F}_2(G;S)$, with floating component highlighted.

Remark 2.2. Let q be a fixed vertex. Then $\mathcal{F}_1(G;\{q\})$ is the set of spanning trees of G, and $\mathcal{F}_2(G;\{q\})$ is the set of two-component spanning forests of G. Note that both these sets are independent of the choice of the vertex q, and were denoted by $\mathcal{F}_1(G)$ and $\mathcal{F}_2(G)$, respectively, in §1.3.

2.2. **Laplacian matrix.** The *incidence matrix* of a graph $G=(V,E)$ (endowed with an arbitrary orientation) is the matrix $B \in \mathbb{R}^{V\times E}$ defined by

$$B_{v,e} = \mathbb{1}(v=e^+) - \mathbb{1}(v=e^-).$$

Here $\mathbb{1}(\cdot)$ denotes the indicator function.

Let $\{\alpha_e \colon e \in E\}$ denote the edge lengths as before. The Laplacian matrix is defined by

$$L = B \begin{pmatrix} \alpha_1^{-1} & & \\ & \ddots & \\ & & \alpha_m^{-1} \end{pmatrix} B^\intercal. \tag{7}$$

It is clear that L is positive semidefinite, since edge lengths are positive.

Note that the incidence matrix depends on the choice of the orientation but the Laplacian matrix does not.

2.3. **Principal minors matrix-tree theorem.** For a subset $A \subseteq E$ we define $w(A)$ as

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

If H is a subgraph of G, we use H to also denote its subset of edges $E(H)$. So, in symbols, $w(H) = w\,(E(H))$.

Let

$$\kappa(G;S) = \sum_{T \in \mathcal{F}_1(G;S)} w(T).$$

When all edge lengths are one, $\kappa(G;S)$ is the number of S-rooted spanning forests.

Remark 2.3. If G is a graph *without* edge lengths (so $\alpha_e = 1$ for all $e \in E$), then $\kappa(G;S)$ is simply the number of S-rooted spanning forests of G. In this case, $\kappa(G;S)$ is also the number of spanning trees of the quotient graph G/S, which "glues together" all vertices in S as a single vertex v_S.

The *(principal minors) matrix-tree theorem* gives a determinantal formula for computing $\kappa(G;S)$, which we now explain. Given $S \subseteq V$, let $L[\overline{S}]$ denote the matrix obtained from L by removing the rows and columns indexed by S.

Theorem 2.4. *Let $G = (V,E)$ be a finite graph with edge lengths $\{\alpha_e \colon e \in E\}$ For any nonempty vertex set $S \subseteq V$,*

$$\kappa(G;S) = \left(\prod_{e \in E} \alpha_e\right) \det L[\overline{S}].$$

Proof. See Tutte [Tut01, Section VI.6, Equation (VI.6.7)]. □

Note that the classical (Kirchhoff's) matrix-tree theorem is the special case, where S is a singleton.

2.4. **Tree splits and tree distance.** Fix a tree $G = (V,E)$ with edge lengths $\{\alpha_e \colon e \in E\}$ (and endowed with an arbitrary orientation). Given an edge $e \in E$, the edge deletion $G \setminus e$ contains two connected components. This phenomenon is referred to as a *tree split*.

Using the orientation on $e = (e^+, e^-)$, we let $(G \setminus e)^+$ denote the component that contains endpoint e^+, and let $(G \setminus e)^-$ denote the component that contains endpoint e^-. For any $e \in E$ and $v \in V$, we denote by $(G \setminus e)^v$ the component of $G \setminus e$ containing v, and $(G \setminus e)^{\overline{v}}$ the component not containing v.

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

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

Remark 2.5.

(i) If we fix an edge e, then $\delta(e; -, -) : V \times V \to \{0, 1\}$ is the indicator function for vertex pairs (v, w) that are on opposite sides of the tree split $G \setminus e$.

(ii) In particular, if we fix e and a vertex v, then $\delta(e; v, -)\colon V(G) \to \{0, 1\}$ is the indicator function for the component $(G \setminus e)^{\overline{v}}$ of the tree split $G \setminus e$ not containing v.

(iii) If we fix vertices v, w, then $\delta(-; v, w)\colon E(G) \to \{0, 1\}$ is the indicator function for the unique path between v and w.

Recall the distance between vertices v, w (with respect to edge lengths $\{\alpha_e\colon e \in E\}$), denoted by $d(v, w)$, is defined as the sum of the edge lengths α_e along the unique path from u to v.

Proposition 2.6. *For a tree $G = (V, E)$ with edge lengths $\{\alpha_e\colon e \in E\}$, the corresponding distance function satisfies*

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

Proof. This follows from Remark 2.5 (iii). ☐

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

$$\sum_{v \in V(G)} (2 - \deg(v)) = 2. \tag{8}$$

We next state a slight generalization (8), which will be useful later.

Given a connected subgraph $H \subseteq G$, we define the *edge boundary* ∂H as the set of edges which join H to its complement; i.e.

$$\partial H = \{e = \{a, b\} \in E\colon a \in V(H),\, b \notin V(H)\}.$$

We define the *outdegree* of H in G as the number of edges in its edge boundary, $\operatorname{outdeg}(H) = |\partial H|$. Note that the edge boundary and outdegree do not depend on the implicit orientation on E.

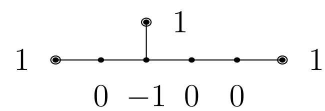
We are especially interested in the following special cases. Let G be a tree and $\emptyset \neq S \subseteq V$. For an S-rooted spanning forest F of G, and $s \in S$, we define the *outdegree* $\operatorname{outdeg}(F, s)$ as the number of edges which join $F(s)$ (the s-component of F) to a different component; i.e. outdegree of $F(s)$ in the tree G. Similarly, if

F is a $(S,*)$-rooted spanning forest of G, we let $\operatorname{outdeg}(F,*)$ denote the outdegree of the floating component $F(*)$.

Example 2.7. Consider the tree G from Example 2.1, where S is the set of three leaf vertices. In Figure 4, we show examples of rooted spanning forests of G, where each component is labelled by $\operatorname{outdeg}(F,s)$, for a leaf vertex s, or $\operatorname{outdeg}(F,*)$ for the floating component.

FIGURE 4. An S-rooted spanning forest (left) and an $(S,*)$-rooted spanning forest (right). Each forest component is labelled with its outdegree.

In consideration of the next lemma, we also show G labelled with $2-\deg(v)$ on each vertex, below.



Lemma 2.8. *Suppose G is a tree.*

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

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

(b) For any fixed edge e and fixed vertex u of G, we have

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

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 $\operatorname{outdeg}(H)=\deg(v)$.

(b) Recall that $(G\setminus e)^{\overline{u}}$ denotes the component of the tree split $G\setminus e$ that does not contain u. Its vertices are precisely those v that satisfy $\delta(e;u,v)=1$ (see Remark 2.5 (iii)), so

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

Since the subgraph $(G\setminus e)^{\overline{u}}$ has a single edge separating it from its complement, we have $\operatorname{outdeg}((G\setminus e)^{\overline{u}})=1$. Apply part (a) to obtain the result. $\square$

2.6. **Transitions between $\mathcal{F}_1(G;S)$ and $\mathcal{F}_2(G;S)$.** A key step in the proof of Theorem B uses the following "transition structure" which relates the S-rooted spanning forests $\mathcal{F}_1(G;S)$ with $(S,*)$-rooted spanning forests $\mathcal{F}_2(G;S)$, via the operations of edge-deletion and edge-union.

- Consider the "deletion map"
$$E(G) \times \mathcal{F}_1(G;S) \xrightarrow{\text{del}} \mathcal{F}_1(G;S) \sqcup \mathcal{F}_2(G;S)$$
defined by
$$(e,T) \mapsto \begin{cases} T & \text{if } e \notin T, \\ T \setminus e & \text{if } e \in T. \end{cases}$$
Given $F \in \mathcal{F}_2(G;S)$, there are exactly $\operatorname{outdeg}(F,*)$-many choices of pairs $(e,T) \in E(G) \times \mathcal{F}_1(G;S)$ such that $F = T \setminus e$.
- Consider the "union map"
$$E(G) \times \mathcal{F}_2(G;S) \longrightarrow \mathcal{F}_1(G;S) \sqcup \mathcal{F}_2(G;S)$$
defined by
$$(e,F) \mapsto \begin{cases} F \cup e & \text{if } e \in \partial F(*), \\ F & \text{if } e \notin \partial F(*) \end{cases}$$
Given $T \in \mathcal{F}_1(G;S)$, there are exactly $(|V| - |S|)$-many choices of pairs $(e,F) \in E(G) \times \mathcal{F}_2(G;S)$ such that $T = F \cup e$.

3. Principal submatrices of D as bilinear forms

Let $G = (V,E)$ be a tree with edge lengths $\alpha = \{\alpha_e : e \in E\}$. The distance matrix (with respect to α) is symmetric, so it defines a symmetric bilinear forms on the vector space $\mathbb{R}^V$.

3.1. **Signature and invertibility.** For a subset of vertices S with $|S| \geq 2$, the submatrix $D[S]$ has nonzero determinant. We give a proof in here, based on finding the signature of $D[S]$ as a bilinear form.

We first recall a celebrated result of Cauchy. Let $M[\bar{i}]$ denote the matrix obtained from M by deleting the i-th row and column.

Proposition 3.1 (Cauchy interlacing)**.** *Suppose M is a symmetric real matrix with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$. Then the submatrix $M[\bar{i}]$ has eigenvalues $\mu_1 \leq \cdots \leq \mu_{n-1}$ satisfying*
$$\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \cdots \leq \mu_{n-1} \leq \lambda_n.$$

Proof. See, for example, [HJ13, Theorem 4.3.17]. □

Lemma 3.2 ([Bap10, Lemma 8.15])**.** *Suppose D is the distance matrix of a tree with n vertices and edge lengths $\{\alpha_e : e \in E\}$. Then D has one positive eigenvalue and $n-1$ negative eigenvalues.*

Remark 3.3. The proof in [Bap10] is by induction on the number of vertices, and uses Cauchy interlacing (Proposition 3.1). Lemma 8.15 in [Bap10] is actually stated for trees without edge lengths. However, the same argument applies to trees with edge lengths if one applies Bapat, Kirkland, Neumann's result [BKN05, Corollary 2.5] (see (2)).

The the following extension of Lemma 3.2 and its proof was communicated to the authors by Bapat (personal communication).

Lemma 3.4. *Suppose D is the distance matrix of a tree $G = (V, E)$ with edge lengths $\{\alpha_e : e \in E\}$. Let $S \subseteq V$ be a subset of size $|S| \geq 2$. Then $D[S]$ has one positive eigenvalue and $|S| - 1$ negative eigenvalues. In particular,* $\det D[S] \neq 0$.

Proof. We apply decreasing induction on the size of S. For $S = V$, we have Lemma 3.2. Now suppose $|S| = k$ where $2 \leq k < n$, and assume, by induction hypothesis, that the claim holds for all vertex subsets of size greater than k. Let $S^+ \subseteq V$ be a set of $k + 1$ vertices containing S. The inductive hypothesis states that $D[S^+]$ has k negative eigenvalues and one positive eigenvalue, so Cauchy interlacing (Proposition 3.1) from $D[S^+]$ implies that $D[S]$ has at least $k - 1$ negative eigenvalues. Since all diagonal entries of $D[S]$ are zero, $D[S]$ has zero trace. Thus the remaining eigenvalue of $D[S]$ must be positive. □

3.2. **Negative definite hyperplane.** We next prove that any principal submatrix of D induces a negative semidefinite quadratic form on the hyperplane of zero-sum vectors.

Bapat, Kirkland, and Neumann ([BKN05, Theorem 2.1]) prove that

$$(D)^{-1} \; = \; -\frac{1}{2}L + \frac{1}{2}\Big(\sum_{e\in E}\alpha_e\Big)^{-1}\mathbf{m}\,\mathbf{m}^{\intercal} \tag{9}$$

where $\mathbf{m}$ is the vector with components $\mathbf{m}_v = 2 - \deg v$. The special case of (9) for trees without edge lengths had appeared in an earlier work by Graham and Lovász ([GL78, Lemma 1]).

Proposition 3.5. *Let D denote the distance matrix of a tree with edge lengths $\{\alpha_e : e \in E\}$. Let L be the Laplacian matrix. Then*

$$D \; = \; -\frac{1}{2}DLD + \frac{1}{2}\left(\sum_{e\in E}\alpha_e\right)\mathbf{1}\mathbf{1}^{\intercal}.$$

Proof. Multiply (9) by the all-ones vector $\mathbf{1}$; since $L\mathbf{1} = 0$ and $\mathbf{m}^{\intercal}\mathbf{1} = 2$ (see (8)), we obtain

$$(D)^{-1}\mathbf{1} \; = \; \left(\sum_{e\in E}\alpha_e\right)^{-1}\mathbf{m}.$$

Therefore,

$$D\mathbf{m} = \left(\sum_{e\in E}\alpha_e\right)\mathbf{1}. \tag{10}$$

The result now follows from multiplying (9) by D on both sides, and using (10). $\square$

Proposition 3.6. *Suppose D is the distance matrix of a tree with edge lengths $\{\alpha_e \colon e \in E\}$. Let $S \subset V$ be a nonempty vertex subset.*

(a) Let $H_{0,S} \subset \mathbb{R}^S$ be the hyperplane of vectors whose coordinates sum to zero. Then the function $\mathbf{u} \mapsto \mathbf{u}^\intercal D[S]\mathbf{u}$ is concave on $H_{0,S}$.
(b) Let $H_{c,S}$ be the affine hyperplane of vectors whose coordinates sum to c. Then the function $\mathbf{u} \mapsto \mathbf{u}^\intercal D[S]\mathbf{u}$ is concave on $H_{c,S}$.

Proof. (a) Suppose $\mathbf{u}$ is in $H_{0,S}$, so that $\mathbf{1}^\intercal\mathbf{u} = 0$. Using Proposition 3.5, we obtain

$$\mathbf{u}^\intercal D\mathbf{u} = -\frac{1}{2}\mathbf{u}^\intercal DLD\mathbf{u} + 0.$$

It is well-known that the Laplacian matrix L is positive semidefinite (this readily follows from (7)). Thus $\mathbf{u} \mapsto (D\mathbf{u})^\intercal L(D\mathbf{u})$ is convex on $H_{0,S}$, so the claim follows.

(b) For $\mathbf{u}$ in $H_{c,S}$, we have

$$\mathbf{u}^\intercal D\mathbf{u} = -\frac{1}{2}\mathbf{u}^\intercal DLD\mathbf{u} + \frac{1}{2}\left(\sum \alpha_e\right) c^2.$$

As $\mathbf{u}$ varies in $H_{c,S}$, the term $\frac{1}{2}\left(\sum \alpha_e\right) c^2$ remains constant. Thus the concavity depends only on the first term, which agrees with part (a). $\square$

4. Quadratic optimization

Here, we explain how the quantity $(\det D[S]/\operatorname{cof} D[S])$, introduced in §1.2 arises as the solution of a quadratic optimization problem.

Proposition 4.1. *Suppose D is the distance matrix of a tree with edge lengths $\{\alpha_e \colon e \in E\}$. If $D[S]$ is a principal submatrix of a distance matrix indexed by a nonempty subset S of vertices, then*

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

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

Assume $|S| \geq 2$. Proposition 3.6 (b) implies that the objective function $\mathbf{u} \mapsto \mathbf{u}^\intercal D[S]\mathbf{u}$ is concave on the domain $\mathbf{1}^\intercal\mathbf{u} = 1$, so any critical point is a global maximum. To find the critical points using the method of Lagrange multipliers, we must compute the gradients of the objective function and the constraint. The gradient of the objective function (i.e. $\mathbf{u}^\intercal D[S]\mathbf{u}$) is $2D[S]\mathbf{u}$, and the gradient of the constraint (i.e. $\mathbf{1}^\intercal\mathbf{u}$) is $\mathbf{1}$. By the theory of Lagrange multipliers, the optimal solution $\mathbf{u}^*$ is a vector satisfying

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

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

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

The last equality uses the given constraint $\mathbf{1}^\intercal\mathbf{u} = 1$.

On the other hand, $D[S]$ is invertible by Lemma 3.4. Therefore, we have $\mathbf{u}^* = \lambda(D[S]^{-1}\mathbf{1})$, so

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

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

Remark 4.2. If we consider G as a resistive electrical network, with each edge e representing a resistor of resistance α_e, then the optimal vector $\mathbf{u}^*$ has a physical interpretation as a current flow: it records the currents in the network when external current enters the network in the amount $(\deg(v)-2)/2$ for each $v \in V\backslash S$, and exits from the nodes in S, and the network is grounded at all nodes in S. Indeed, we give an explicit combinatorial expression for $\mathbf{u}^*$, up to a normalizing constant, in (11). It is a classical result in network theory that this gives the current flow as described; see, for example, Tutte [Tut01, Section VI.6].

We note the following restatement of Proposition 4.1, viewing $\mathbb{R}^S$ as a subspace of $\mathbb{R}^V$ where coordinates indexed by $V \setminus S$ are set to zero.

Corollary 4.3. *If $D[S]$ is a principal submatrix of a distance matrix indexed by S, then*

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

4.1. **Cofactor sums.** 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 essentially due to Bapat and Sivasubramanian ([BS11]).

Recall from §1.2 that cof M denotes the *sum of cofactors* of M, i.e. $\operatorname{cof} M = \displaystyle\sum_{i=1}^n\sum_{j=1}^n (-1)^{i+j} \det M[\bar{i},\bar{j}]$ where $M[\bar{i},\bar{j}]$ denotes the matrix with the i-th row and j-th column deleted. Recall that $\kappa(G;S) := \sum_{T\in\mathcal{F}_1(G;S)} w(T)$.

Theorem 4.4 (Distance submatrix cofactor sums)**.** *Let $G = (V,E)$ be a tree with edge lengths $\{\alpha_e \colon e \in E\}$, and let D be the distance matrix of G. Let $S \subseteq V$ be a nonempty subset of vertices. Then*

$$\operatorname{cof} D[S] = (-2)^{|S|-1}\kappa(G;S).$$

Proof. Bapat and Sivasubramanian ([BS11, Theorem 11]) show that

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

where L is the Laplacian matrix. Combine this identity with Theorem 2.4. □

The following result is a direct consequence of theorems of Bapat, Kirkland, Neumann ([BKN05]) and Bapat, Sivasubramanian ([BS11]).

Proposition 4.5. *Suppose D is the distance matrix of a tree with edge lengths $\{\alpha_e \colon e \in E\}$. Then*

$$\frac{\det D}{\operatorname{cof} D} = \frac{1}{2} \sum_{e \in E} \alpha_e.$$

Proof. Consider Theorem 4.4 with $S = V$. In this case $\mathcal{F}_1(G; V)$ is a singleton consisting of the forest T with no edges, so $w(T)$ is the product of all edge lengths. Thus

$$\operatorname{cof} D = (-2)^{n-1} \prod_{e \in E} \alpha_e.$$

Combining this with (2) yields the result. □

4.2. **Monotonicity.** Using Proposition 4.1, we can now prove Theorem C and Theorem D.

Proof of Theorem C. By Corollary 4.3, 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 on an affine subspace of $\mathbb{R}^V$, but the subspace for A is contained in the subspace for B. □

Proof of Theorem D. (a) To see

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

take B as the set of all vertices in $\operatorname{conv}(S, G)$. Then $S \subseteq B$, and apply Theorem C. By Proposition 4.5 we have

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

(b) Recall that γ is a simple path between vertices $s_0, s_1 \in S$. To see

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

let $A = \{s_0, s_1\}$. Then $A \subseteq S$ by assumption. Now apply Theorem C. By Proposition 4.5 we have

$$\frac{\det D[A]}{\operatorname{cof} D[A]} = \frac{1}{2} d(s_0, s_1) = \frac{1}{2} \sum_{e \in \gamma} \alpha_e.$$ □

5. Proof of Theorems A and B

In this section, we prove our main result, Theorem B. Theorem A is an immediate corollary. The arguments do not make use of results in §4.

5.1. **Outline of the proof.** We compute $\det D[S]$ via the following steps.

(I) Find a vector $\mathbf{m} \in \mathbb{R}^S$ such that $D[S]\mathbf{m} = \lambda \mathbf{1}$. Find λ.
(II) Compute the sum of entries of $\mathbf{m}$, i.e. $\mathbf{1}^\intercal \mathbf{m}$.
(III) Since $D[S]$ is nonsingular (Lemma 3.2), $\mathbf{m} = \lambda(D[S]^{-1}\mathbf{1})$ and
$$\mathbf{1}^\intercal \mathbf{m} = \lambda(\mathbf{1}^\intercal D[S]^{-1}\mathbf{1}) = \lambda \frac{\operatorname{cof} D[S]}{\det D[S]}.$$
Calculate $\dfrac{\det D[S]}{\operatorname{cof} D[S]} = \dfrac{\lambda}{\mathbf{1}^\intercal \mathbf{m}}$.
(IV) Multiply previous expression by $\operatorname{cof} D[S]$, using expression in Theorem 4.4, to compute $\det D[S]$.

It turns out that the entries of $\mathbf{m}$ are combinatorially meaningful (see (11)), which also gives combinatorial meaning to the constant λ.

5.2. **Equilibrium vector identities.** Fix a tree $G = (V, E)$ with edge lengths $\{\alpha_e \colon e \in E\}$ and a nonempty subset $S \subseteq V$. We first define a vector $\mathbf{m}$ which satisfies the relation $D[S]\mathbf{m} = \lambda\mathbf{1}$ for some λ.

Definition 5.1. Let $\mathbf{m} = \mathbf{m}(G; S)$ denote the vector in $\mathbb{R}^S$ be defined by

$$\mathbf{m}_v = \sum_{T \in \mathcal{F}_1(G;S)} w(T)(2 - \operatorname{outdeg}(T, v)) \qquad \text{for each } v \in S. \tag{11}$$

where $\operatorname{outdeg}(T, v)$ is the outdegree of the v-component of T (see §2.5). We call $\mathbf{m}$ the *equilibrium vector* of $(G; S)$.

The choice of terminology "equilibrium" here is in reference to potential theory, c.f. [Tsu75][Ste23].

Proposition 5.2. *Suppose S is nonempty. For the vector $\mathbf{m} = \mathbf{m}(G; S)$ defined above, we have*

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

Proof. (a) By Lemma 2.8 we can express $\operatorname{outdeg}(T, s)$ as a sum over vertices in $T(s)$,

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

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

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

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

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

where we apply equation (8) for the last equality.

(b) All edge lengths are positive, so $w(T) > 0$ for all T, and $\mathcal{F}_1(G;S)$ is nonempty as long as S is nonempty. Therefore, part (a) implies that $\mathbf{1}^\intercal \mathbf{m} > 0$. ☐

The following computation is the technical heart of our main result.

Theorem 5.3. *With* $\mathbf{m} = \mathbf{m}(G;S)$ *defined as in* (11), $D[S]\mathbf{m} = \lambda \mathbf{1}$ *for the constant*

$$\lambda = \sum_{e \in E(G)} \alpha_e \sum_{T \in \mathcal{F}_1(G;S)} w(T) - \sum_{F \in \mathcal{F}_2(G;S)} w(F)\, (2 - \operatorname{outdeg}(F,*))^2.$$

Proof. For $e \in E$ and $v, w \in V$, let $\delta(e; v, w)$ denote the function defined in §2.4. For any $v \in S$, we have

$$\begin{aligned}
(D[S]\mathbf{m})_v &= \sum_{s \in S} d(v,s) \mathbf{m}_s \\
&= \sum_{s \in S} \left(\sum_{e \in E(G)} \alpha_e\, \delta(e; v, s) \right) \left(\sum_{T \in \mathcal{F}_1(G;S)} (2 - \operatorname{outdeg}(T,s)) w(T) \right) \\
&= \sum_{T \in \mathcal{F}_1(G;S)} w(T) \sum_{e \in E} \alpha_e \left(\sum_{s \in S} \delta(e; v, s) (2 - \operatorname{outdeg}(T,s)) \right) \\
&= \sum_{T \in \mathcal{F}_1(G;S)} w(T) \sum_{e \in E} \alpha_e \left(\sum_{s \in S} \delta(e; v, s) \sum_{u \in T(s)} (2 - \deg(u)) \right). \qquad (12)
\end{aligned}$$

The second line follows from Proposition 2.6 and (11). The last line follows from using Lemma 2.8 for the subgraph $H = T(s)$.

We introduce additional notation to handle the double sum in parentheses in (12). Each S-rooted spanning forest T naturally induces a surjection $\pi_T : V \to S$, defined by

$$\pi_T(u) = s \qquad \text{if and only if} \qquad u \in T(s), \text{ the } s\text{-component of } T.$$

Using this notation,

$$(D[S]\mathbf{m})_v = \sum_{T \in \mathcal{F}_1(G;S)} w(T) \sum_{e \in E} \alpha_e \left(\sum_{u \in V} (2 - \deg(u)) \delta(e; v, \pi_T(u)) \right) \qquad (13)$$

We will compare how the above expression changes after replacing $\delta(e; v, \pi_T(u))$ with $\delta(e; v, u)$.

From Lemma 2.8 (b), we have $\sum_{u\in V}(2-\deg(u))\delta(e;v,u)=1$. Thus

$$\sum_{T\in\mathcal{F}_1(G;S)} w(T)\sum_{e\in E}\alpha_e=\sum_{T\in\mathcal{F}_1(G;S)} w(T)\sum_{e\in E}\alpha_e\left(\sum_{u\in V}(2-\deg(u))\delta(e;v,u)\right) \tag{14}$$

By subtracting equation (14) from (13), we obtain

$$\begin{aligned}(D[S]\mathbf{m})_v-&\sum_{T\in\mathcal{F}_1(G;S)} w(T)\sum_{e\in E}\alpha_e\\&=\sum_{T\in\mathcal{F}_1(G;S)} w(T)\sum_{e\in E}\alpha_e\sum_{u\in V}(2-\deg(u))\Big(\delta(e;v,\pi_T(u))-\delta(e;v,u)\Big).\end{aligned} \tag{15}$$

For the rest of the argument, let $(\star)=(D[S]\mathbf{m})_v-\sum_{T\in\mathcal{F}_1(G;S)} w(T)\sum_{e\in E}\alpha_e$.

We focus on the inner expression $\delta(e;v,\pi_T(u))-\delta(e;v,u)$, where we suppose e, v, and T are fixed. Recall that

$$\delta(e;v,\pi_T(u))-\delta(e;v,u)=\mathbb{1}(\pi_T(u)\in(G\setminus e)^{\overline{v}})-\mathbb{1}(u\in(G\setminus e)^{\overline{v}}), \tag{16}$$

c.f. Remark 2.5 (ii). 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 tree split $G\setminus e$, for every vertex u. In this case $\delta(e;v,\pi_T(\cdot))-\delta(e;v,\cdot)=0$.
- Case 2: if $e\in T$ and $\pi_T(e)$ 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 5, left.
- Case 3: if $e\in T$ and $\pi_T(e)$ 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 5, right.

FIGURE 5. Edge $e\in T$ with $\delta(e;v,\pi_T(e))=1$ (left) and $\delta(e;v,\pi_T(e))=0$ (right). The floating component of $T\setminus e$ is highlighted.

Thus, when multiplying the term (16) by $(2-\deg(u))$ and summing over all vertices u, we obtain

$$\begin{aligned}&\sum_{u\in V}(2-\deg(u))\left(\delta(e;v,\pi_T(u))-\delta(e;v,u)\right)\\&=\begin{cases}0 & \text{if } e\notin T,\\ 2-\operatorname{outdeg}(T\setminus e,*) & \text{if } e\in T(s_0) \text{ and } \delta(e;v,s_0)=1,\\ -(2-\operatorname{outdeg}(T\setminus e,*)) & \text{if } e\in T(s_0) \text{ and } \delta(e;v,s_0)=0.\end{cases}\end{aligned} \tag{17}$$

Therefore substituting (17) into (15) yields

$$(\star) = \sum_{T \in \mathcal{F}_1(G;S)} w(T) \sum_{e \in T} \alpha_e \times \\ (2 - \operatorname{outdeg}(T \setminus e, *)) \Big(\mathbb{1}(\delta(e; v, \pi_T(e)) = 1) - \mathbb{1}(\delta(e; v, \pi_T(e)) = 0) \Big), \tag{18}$$

where we use the shorthand $(\star) = (D[S]\mathbf{m})_v - \sum_{T \in \mathcal{F}_1(G;S)} w(T) \sum_{e \in E} \alpha_e$.

We now rewrite the above expression in terms of $\mathcal{F}_2(G;S)$. Observe in (18) that the deletion $T \setminus e$ is an $(S,*)$-rooted spanning forest of G, and that the corresponding weights satisfy

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

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 $\partial F(*)$.

Thus

$$\begin{aligned}(\star) &= \sum_{F \in \mathcal{F}_2} w(F)(2 - \operatorname{outdeg}(F, *)) \times \\ &\qquad \sum_{e \in \partial F(*)} \Big(\mathbb{1}(\delta(e; v, \pi_{(F \cup e)}(e)) = 1) - \mathbb{1}(\delta(e; v, \pi_{(F \cup e)}(e)) = 0) \Big) \\ &= \sum_{F \in \mathcal{F}_2} w(F)(2 - \operatorname{outdeg}(F, *)) \times \\ &\qquad \Bigg(\#\{e \in \partial F(*) \colon \delta(e; v, \pi_{(F \cup e)}(e)) = 1\} - \#\{e \in \partial F(*) \colon \delta(e; v, \pi_{(F \cup e)}(e)) = 0\} \Bigg).\end{aligned}$$

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 that $\delta(e; v, \pi_{(F \cup e)}(e)) = 1$ (respectively, $\delta(e; v, \pi_{(F \cup e)}(e)) = 0$) is equivalent to $\delta(e; v, F(*)) = 0$ (respectively, $\delta(e; v, F(*)) = 1$). For an illustration, compare Figures 6 and 7. Therefore

$$(\star) = \sum_{F \in \mathcal{F}_2} w(F)(2 - \operatorname{outdeg}(F, *)) \Bigg(\#\{e \in \partial F(*) \colon \delta(e; v, F(*)) = 0\} \\ - \#\{e \in \partial F(*) \colon \delta(e; v, F(*)) = 1\} \Bigg). \tag{19}$$

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

$$\#\{e \in \partial F(*) \colon \delta(e; v, F(*)) = 1\} = 1, \qquad \text{and}$$
$$\#\{e \in \partial F(*) \colon \delta(e; v, F(*)) = 0\} = \operatorname{outdeg}(F, *) - 1.$$

Thus, the previous expression (19) simplifies as

$$(\star) = -\sum_{F \in \mathcal{F}_2} w(F)(2 - \operatorname{outdeg}(F, *))^2$$

as desired. ☐

Figure 6. A forest $F \in \mathcal{F}_2(G; S)$ and edge $e \in \partial F(*)$, with $\delta(e; v, F(*)) = 0$ (left). After adding e to F to obtain $T = F \cup e \in \mathcal{F}_1(G; S)$, we have $\delta(e; v, \pi_T(e)) = 1$ (right).

Figure 7. A forest $F \in \mathcal{F}_2(G; S)$ and edge $e \in \partial F(*)$, with $\delta(e; v, F(*)) = 1$ (left). After adding e to F to obtain $T = F \cup e \in \mathcal{F}_1(G; S)$, we have $\delta(e; v, \pi_T(e)) = 0$ (right).

5.3. **Principal minors of D.** Finally we can prove our main theorems, giving combinatorial expressions for $\det D[S]$.

Proof of Theorem B. 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 weight $w(G) = 1$. Moreover, the subgraphs in $\mathcal{F}_2(G; \{v\})$ are precisely $G \setminus e$ for $e \in E$, and for each forest $F = G \setminus e$ we have $w(F) = \alpha_e$ and $(\operatorname{outdeg}(F, *) - 2)^2 = 1$. This shows that the right-hand side of (3) is zero.

Next, suppose $|S| \geq 2$. Lemma 3.4 states that $D[S]$ is nonsingular, so we may use the inverse matrix identity

$$\mathbf{1}^\intercal D[S]^{-1}\mathbf{1} = \frac{\operatorname{cof} D[S]}{\det D[S]}. \tag{20}$$

Let $\mathbf{m} = \mathbf{m}(G;S)$ denote the equilibrium vector (11). By Proposition 5.2 (a) and Theorem 4.4,

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

Theorem 5.3 states that $D[S]\mathbf{m} = \lambda \mathbf{1}$ for some constant λ, which is nonzero since $D[S]$ is invertible (see Lemma 3.4) and $\mathbf{m}$ is nonzero (see Proposition 5.2 (b)). Hence

$$\mathbf{1}^\intercal D[S]^{-1} \mathbf{1} = \lambda^{-1} \mathbf{1}^\intercal \mathbf{m} = \frac{\operatorname{cof} D[S]}{(-1)^{|S|-1} 2^{|S|-1} \lambda}. \tag{21}$$

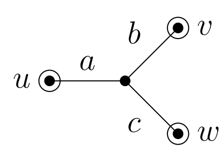
Comparing (20) with (21) gives the desired result, $\det D[S] = (-1)^{|S|-1} 2^{|S|-1} \lambda$. $\square$

Proof of Theorem A. Set all lengths α_e to 1 in Theorem B. In this case, the weights $w(T) = 1$ and $w(F) = 1$ for all forests T and F, and

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

6. Examples

Example 6.1. Suppose G is a tree consisting of three edges joined at a central vertex.



Suppose S consists of the leaf vertices $\{u, v, w\}$. Then

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

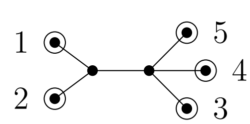
which has determinant

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

The equilibrium vector (which satisfies $D[S]\mathbf{m} = \lambda\mathbf{1}$) in this example is

$$\mathbf{m} = \begin{pmatrix} ab+ac & ab+bc & ac+bc \end{pmatrix}^\intercal.$$

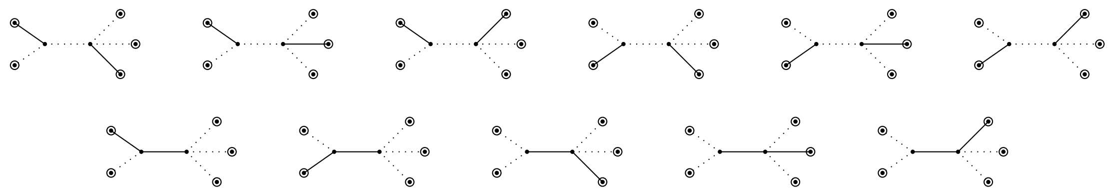
Example 6.2. Suppose G is the tree with unit edge lengths shown below, with five leaf vertices.



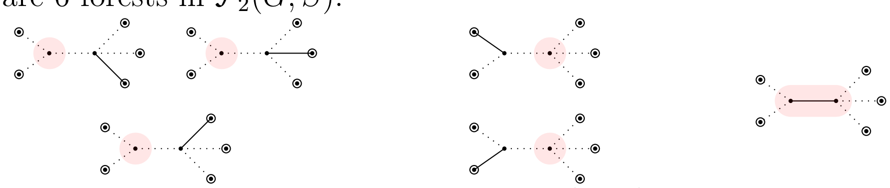
Let S denote the set of five leaf vertices. Then

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

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



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



Out of the floating components of forests in $\mathcal{F}_2(G; S)$, 3 have outdegree three, 2 have outdegree four, and 1 has outdegree five.

The determinant of the distance submatrix is

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

and the equilibrium vector is $\mathbf{m} = \begin{pmatrix} 5 & 5 & 4 & 4 & 4 \end{pmatrix}^\intercal$.

Example 6.3. Suppose G is the tree with edge lengths shown in Figure 8, with four leaf vertices and two internal vertices. Let S denote the set of four leaf vertices.

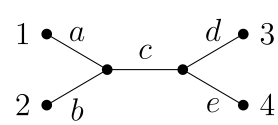


FIGURE 8. Tree with four leaves, and varying edge lengths.

Then

$$D[S] = \begin{pmatrix} 0 & a+b & a+c+d & a+c+e \\ a+b & 0 & b+c+d & b+c+e \\ a+c+d & b+c+d & 0 & d+e \\ a+c+e & b+c+e & d+e & 0 \end{pmatrix}$$

and the equilibrium vector is

$$\mathbf{m} = abd\begin{pmatrix}1\\1\\1\\-1\end{pmatrix} + abe\begin{pmatrix}1\\1\\-1\\1\end{pmatrix} + acd\begin{pmatrix}1\\0\\1\\0\end{pmatrix} + ace\begin{pmatrix}1\\0\\0\\1\end{pmatrix} + ade\begin{pmatrix}1\\-1\\1\\1\end{pmatrix}$$
$$+ bcd\begin{pmatrix}0\\1\\1\\0\end{pmatrix} + bce\begin{pmatrix}0\\1\\0\\1\end{pmatrix} + bde\begin{pmatrix}-1\\1\\1\\1\end{pmatrix}.$$

The determinant of $D[S]$ normalized by its sum of cofactors, is

$$\frac{\det D[S]}{\operatorname{cof} D[S]} = \frac{1}{2}\Big((a+b+c+d+e)$$
$$- \frac{(abcd + abce + acde + bcde) + 4(abde)}{(abd + abe + acd + ace + ade + bcd + bce + bde)}\Big).$$

References

[ABBGF16] O. Amini, S. Bloch, J. I. Burgos Gil, and J. Fresán, *Feynman amplitudes and limits of heights*, Izv. Math. **80** (2016), no. 5, 5–40. MR3588803

[Ami19] Omid Amini, *The exchange graph and variations of the ratio of the two Symanzik polynomials*, Ann. Inst. Henri Poincaré D **6** (2019), no. 2, 155–197. MR3950652

[Bap10] R. B. Bapat, *Graphs and matrices*, Universitext, Springer, London; Hindustan Book Agency, New Delhi, 2010. MR2797201

[BKN05] R. Bapat, S. J. Kirkland, and M. Neumann, *On distance matrices and Laplacians*, Linear Algebra Appl. **401** (2005), 193–209. MR2133282

[BR07] Matt Baker and Robert Rumely, *Harmonic analysis on metrized graphs*, Canad. J. Math. **59** (2007), no. 2, 225–275. MR2310616

[Bro24] Francis Brown, *Generalised graph Laplacians and canonical Feynman integrals with kinematics*, Comm. Math. Phys. **405** (2024), no. 2, Paper No. 41, 46. MR4703455

[BS11] R. B. Bapat and Sivaramakrishnan Sivasubramanian, *Identities for minors of the Laplacian, resistance and distance matrices*, Linear Algebra Appl. **435** (2011), no. 6, 1479–1489. MR2807165

[CBSMI24] Lena Collienne, Mary Barker, Marc A Suchard, and Frederick A Matsen IV, *Phylogenetic tree instability after taxon addition: empirical frequency, predictability, and consequences for online inference*, Systematic Biology (2024).

[CK24] Projesh Nath Choudhury and Apoorva Khare, *Distance matrices of a tree: two more invariants, and in a unified framework*, European J. Combin. **115** (2024), Paper No. 103787, 30. MR4623916

[Dev22] Karel Devriendt, *Graph geometry from effective resistances*, Ph.D. Thesis, 2022. Available at https://ora.ox.ac.uk/objects/uuid:bb174997-ef29-4bdd-a8cf-3db7b5ed5429.

[GHH77] R. L. Graham, A. J. Hoffman, and H. Hosoya, *On the distance matrix of a directed graph*, J. Graph Theory **1** (1977), no. 1, 85–88. MR505769

[GL24] Álvaro Gutiérrez and Adrián Lillo, *Principal minors of the distance matrix of a tree*, 2024. Preprint available at `arXiv:2407.01966`.

[GL78] R. L. Graham and L. Lovász, *Distance matrix polynomials of trees*, Adv. in Math. **29** (1978), no. 1, 60–88. MR480119

[GP71] R. L. Graham and H. O. Pollak, *On the addressing problem for loop switching*, Bell System Tech. J. **50** (1971), 2495–2519. MR289210

[HJ13] Roger A. Horn and Charles R. Johnson, *Matrix analysis*, Second, Cambridge University Press, Cambridge, 2013. MR2978290

[KKW15] Adrien Kassel, Richard Kenyon, and Wei Wu, *Random two-component spanning forests*, Ann. Inst. Henri Poincaré Probab. Stat. **51** (2015), no. 4, 1457–1464. MR3414453

[RSW24] Harry Richman, Farbod Shokrieh, and Chenxi Wu, *On capacity of metric graphs*, 2024. In preparation.

[Ste23] Stefan Steinerberger, *Curvature on graphs via equilibrium measures*, J. Graph Theory **103** (2023), no. 3, 415–436. MR4596504

[Tsu75] M. Tsuji, *Potential theory in modern function theory*, Chelsea Publishing Co., New York, 1975. Reprinting of the 1959 original. MR0414898

[Tut01] W. T. Tutte, *Graph theory*, Encyclopedia of Mathematics and its Applications, vol. 21, Cambridge University Press, Cambridge, 2001. With a foreword by Crispin St. J. A. Nash-Williams, Reprint of the 1984 original. MR1813436

Email address: hrichman@uw.edu

Email address: farbod@uw.edu

Email address: cwu367@math.wisc.edu