FPSAC 2011, Reykjavík, Iceland DMTCS proc. **AO**, 2011, 375–386

Dissimilarity Vectors of Trees and Their Tropical Linear Spaces (Extended Abstract)

Benjamin Iriarte Giraldo[1]†

[1]*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA*

Abstract. We study the combinatorics of weighted trees from the point of view of tropical algebraic geometry and tropical linear spaces. The set of dissimilarity vectors of weighted trees is contained in the tropical Grassmannian, so we describe here the tropical linear space of a dissimilarity vector and its associated family of matroids. This gives a family of complete flags of tropical linear spaces, where each flag is described by a weighted tree.

Résumé. Nous étudions les propriétés combinatoires des arbres pondérés avec le formalisme de la géométrie tropicale et des espaces linéaires tropicaux. L'ensemble de vecteurs de dissimilarité des arbres pondérés est contenu dans la grassmannienne tropicale, donc nous décrivons ici l'espace linéaire tropical d'un vecteur de dissimilarité et sa famille de matroïdes associée. Cela permet d'obtenir une famille de drapeaux complets d'espaces linéaires tropicaux, où chaque drapeau est décrit par un arbre pondéré.

Resumen. Estudiamos la combinatoria de los árboles valuados desde el punto de vista de la geometría algebraica tropical y de los espacios lineales tropicales. El conjunto de los vectores de disimilaridad de un árbol valuado está contenido en el grassmanniano tropical y aquí describimos el espacio lineal tropical de un vector de disimilaridad y su familia asociada de matroides. Se obtiene entonces una familia de banderas completas de espacios lineales tropicales, donde cada bandera se describe mediante un árbol.

Keywords: dissimilarity vector, tropical linear space, tight span, weighted tree, matroid polytope, T-theory

1 Introduction

1.1 Basic Definitions

For every finite set E, define $\binom{E}{m}$ to be the collection of all subsets of E of size m. We adopt the convention that $[n] = \{1, 2, \ldots, n\}$, and we further define $[0] = \emptyset$. We will use the letters $\mathcal{E}$, $\mathcal{V}$ and $\mathcal{L}$ to denote the set of edges, vertices and leaves of a tree, respectively.

Throughout this article, we will consider a tree T with $n \geq 1$ leaves labeled by the set $[n]$, and such that T does not contain internal vertices of degree 2. Moreover, we will assume the existence of a weight function $\mathcal{E}(T) \xrightarrow{\omega} \mathbb{R}$ such that $\omega(e) > 0$ if e is an internal edge of T. We say that S is a **subtree** of T and write $S \subseteq_{\mathrm{st}} T$ whenever $\mathcal{V}(S) \subseteq \mathcal{V}(T)$, S is the restriction of T to $\mathcal{V}(S)$ and S is itself a tree. We will

†Supported by the San Francisco State University-Colombia (SFSU-Colombia) initiative.

also write $S \subsetneq_{\text{st}} T$ if $\mathcal{V}(S) \subsetneq \mathcal{V}(T)$. The function ω will extend to a function from the subtrees of T to the reals in the natural way so that we may speak of total weights of subtrees of T. As a convention, if S is a subtree of T and $\mathcal{E}(S) = \emptyset$, we will let $\omega(S) = 0$.

At all points, we will identify $\mathcal{L}(T)$ with $[n]$.

Now, for $m \in [n]$, we define a vector $d \in \mathbb{R}^{\binom{[n]}{m}}$ associated with the tree T and called the m-**dissimilarity vector** of T. For each $A \in \binom{[n]}{m}$, let d_A be the total weight of the subtree of T spanned by the set of leaves A. The letter d will always denote a dissimilarity vector.

At some points, we will consider instead a tree U with n leaves labeled by the set $[n]$ with $n \geq 1$, with a weight function $\mathcal{E}(U) \xrightarrow{\omega} \mathbb{R}_{\geq 0}$. The function ω will again extend to a function from the subtrees of U to the nonnegative reals in the natural way, so that we may speak of total weights of subtrees of U and define the m-dissimilarity vector of U in a completely analogous way. Furthermore, we will assume that U is trivalent, rooted, ℓ-equidistant (*i.e.* the total weight of the minimal path from every leaf to the root is a constant ℓ), and that ω induces a metric on $\mathcal{L}(U) = [n]$. Such a tree U will be referred to as an **ultrametric tree**. Along with U we consider the poset $(\mathcal{V}(U), \leq_{\text{TO}})$, called the **tree-order** of U, by which $u \leq_{\text{TO}} v$ if the minimal path from the root of U to v contains u.

1.2 Tropical Algebraic Geometry

Consider the field $\mathbb{K} = \mathbb{C}\{\{t\}\}$ of **dual Puiseux series**. Let $\mathbb{K}^* := \mathbb{K}\backslash\{0\}$. Recall that $\mathbb{K}^*$ consists of all formal expressions $\mathfrak{t} = \sum_{-\infty}^{k=p} c_k t^{k/q}$, where $p \in \mathbb{Z}$, $c_p \neq 0$, $q \in \mathbb{Z}^+$ and $c_k \in \mathbb{C}$ for all $k \leq p$. The field comes equipped with a standard valuation $\text{val} : \mathbb{K}^* \mapsto \mathbb{Q}$ by which $\text{val}(\mathfrak{t}) = p/q$. We extend this valuation to a function $\text{val} : (\mathbb{K}^*)^n \mapsto \mathbb{Q}^n$ by taking pointwise valuation.

Let $X = (X_1, \ldots, X_n)$ be a vector of variables. Consider a polynomial $f \in \mathbb{K}[X]$ and write it explicitly as $f(X) = \sum_{a \in A \subseteq \mathbb{Z}^n_{\geq 0}} \mathfrak{t}_a X^a$, where $\mathfrak{t}_a \in \mathbb{K}^*$ for all $a \in A$.

The **tropicalization** $\text{trop } f : \mathbb{R}^n \mapsto \mathbb{R}$ of f is $\text{trop } f(x) = \max\limits_{a \in A}\{\text{val}(\mathfrak{t}_a) + a \cdot x\}$, where $x = (x_1, \ldots, x_n)$ is a vector of variables and the product of vectors is the dot product.

Let $\mathcal{T}(\text{trop } f) \subseteq \mathbb{R}^n$ be the set of points where $\text{trop } f$ attains its maximum twice. In this definition, $\text{trop } f$ can be replaced by any convex piecewise linear form $L : \mathbb{R}^n \mapsto \mathbb{R}$. Call $\mathcal{T}(\text{trop } f)$ the **tropical hypersurface** of f. More generally, consider an arbitrary set $S \subseteq \mathbb{K}[X]$. Let $\text{trop } S := \{\text{trop } f \mid f \in S\}$. Define the **tropical variety** of S to be the set $\mathcal{T}(\text{trop } S) = \bigcap_{f \in S} \mathcal{T}(\text{trop } f)$. Again, $\text{trop } S$ can be replaced by a set of convex piecewise linear forms mapping $\mathbb{R}^n$ into $\mathbb{R}$.

Let $V(S) := \{\mathfrak{t} \in (\mathbb{K}^*)^n \mid f(\mathfrak{t}) = 0 \text{ for all } f \in S\}$.

We have:

Theorem 1.1 (Theorem 2.1, Speyer and Sturmfels [SS04]) *Let $I \subseteq \mathbb{K}[X]$ be an ideal. Then, the following subsets of $\mathbb{R}^n$ coincide:*

a) *The closure of the set $\{val(\mathfrak{t}) \mid \mathfrak{t} \in V(I)\}$.*

b) *The tropical variety $\mathcal{T}(trop\ I)$.*

□

Let $Z = (Z_{ij})$ be an $m \times n$ matrix of indeterminates. Consider the polynomial ring: $\mathbb{K}[Y] = \mathbb{K}[Y_A : A \in \binom{[n]}{m}]$. Let $\phi_{m,n} : \mathbb{K}[Y] \mapsto \mathbb{K}[Z]$ be the homomorphism of rings taking Y_A to the maximal minor of Z obtained from the set of columns A.

The **Plücker ideal** or ideal of **Plücker relations** is the homogeneous prime ideal $I_{m,n} = \ker(\phi_{m,n}) \subseteq \mathbb{K}[Y]$. In particular, if $P \in (\mathbb{K}^*)^{\binom{[n]}{m}}$ is a vector of **Plücker coordinates** of an m-dimensional linear subspace V of $\mathbb{K}^n$, then $P \in V(I_{m,n})$.

The tropical variety $\mathcal{T}(\text{trop } I_{m,n})$ receives the special name of **tropical Grassmannian** and is denoted by $\mathcal{G}_{m,n}$. Theorem 1.1 then shows that if P is a vector of Plücker coordinates of an m-dimensional linear subspace V of $\mathbb{K}^n$ and P has non-zero entries, then $\text{val}(P) \in \mathcal{G}_{m,n}$.

1.3 Tropical Linear Spaces

Consider an m-dimensional linear subspace V of $\mathbb{K}^n$, and let P be a vector of Plücker coordinates of V. Consider the family $\mathcal{S}_V$ of linear polynomials in $\mathbb{K}[X]$:

$$\sum_{r=0}^{m+1} (-1)^r P_{i_1 i_2 \ldots \widehat{i_r} \ldots i_m i_{m+1}} X_{i_r} \text{ for all } 1 \leq i_1 < i_2 < \cdots < i_m < i_{m+1} \leq n. \tag{1.1}$$

Then $\mathfrak{t} \in V$ if and only if $f(\mathfrak{t}) = 0$ for all $f \in \mathcal{S}_V$, defining V as a Zariski closed set.

Let $p \in \mathbb{R}^{\binom{[n]}{m}}$. Consider the set of three-term Plücker relations $S_{m,n} := \{Y_{Aij}Y_{Akl} - Y_{Aik}Y_{Ajl} + Y_{Ail}Y_{Ajk} \mid A \in \binom{[n]}{m-2} \text{ and } \{i,j,k,l\} \subseteq \binom{[n]\setminus A}{4}\} \subseteq I_{m,n} \subseteq \mathbb{K}[Y]$, so that $\mathcal{G}_{m,n} \subseteq \mathcal{T}(\text{trop } S_{m,n})$. Say that p satisfies the **tropical Plücker relations** if $p \in \mathcal{T}(\text{trop } S_{m,n})$.

In particular, if $p = \text{val}(P)$ and P is a vector of Plücker coordinates, then $p \in \mathcal{G}_{m,n}$ implies that $p \in \mathcal{T}(\text{trop } S_{m,n})$, but if $p \in \mathcal{T}(\text{trop } S_{m,n})$, we may not necessarily find such P.

Exploiting the definition of linear spaces as closed sets, let us define the tropical analogue of a linear space.

Let $p \in \mathcal{T}(\text{trop } S_{m,n})$. Considering the tropical analogues of the equations 1.1, let

$$L(p) := \bigcap_{A \in \binom{[n]}{m+1}} \mathcal{T}\left(\max_{i \in A}\{p_{A\setminus i} + x_i\}\right).$$

The set $L(p)$ is the m-dimensional **tropical linear space** associated to p.

For all $A \in \binom{[n]}{m}$, define $e^A = \sum_{i \in A} e_i$, where the sum takes place in $\mathbb{R}^n$ with its canonical basis $\{e_1, e_2, \ldots, e_n\}$. In general, for $x \in \mathbb{R}^n$, let $x^A = \sum_{i \in A} x_i$. Let $\mathscr{H}_m$ be the m-**hypersimplex** of $\mathbb{R}^n$, *i.e.* the convex hull of all vectors e^A with $A \in \binom{[n]}{m}$.

Suppose that $p \in \mathbb{R}^{\binom{[n]}{m}}$. Let $\mathscr{H}_m^p \subseteq \mathbb{R}^{n+1}$ be the convex hull of all points (e^A, p_A) with $A \in \binom{[n]}{m}$. Then p induces a regular subdivision $\mathcal{D}_p$ of $\mathscr{H}_m$ by projecting the upper faces of $\mathscr{H}_m^p$ down to $\mathscr{H}_m$, where by upper face we mean that its outer normal vector has a positive last coordinate.

Let $\mathcal{P}$ be a subpolytope of $\mathscr{H}_m$. We say that $\mathcal{P}$ is **matroidal** or that $\mathcal{P}$ is a **matroid polytope** if the collection of sets $A \in \binom{[n]}{m}$ for which $e^A \in \mathcal{P}$ is the set of bases of a rank-m matroid over $[n]$.

The following theorem motivates considering the set $S_{m,n}$.

Theorem 1.2 (Speyer [Spe08]) *The following assertions are equivalent:*

a) *The vector* $p \in \mathbb{R}^{\binom{[n]}{m}}$ *satisfies the tropical Plücker relations;*

b) *Every face of* $\mathcal{D}_p$ *is matroidal.*

□

If p satisfies the tropical Plücker relations, let $\mathfrak{M}_p$ be the family of rank-m matroids given by the faces of $\mathcal{D}_p$. Recall that a matroid over a finite set E is said to be **loopless** if every element of E is contained in at least one basis. Let $\mathfrak{M}_p^{loop}$ be the subfamily of loopless matroids of $\mathfrak{M}_p$, and let $\mathcal{D}_p^{loop}$ be the subcomplex of faces of $\mathcal{D}_p$ described by $\mathfrak{M}_p^{loop}$. Similarly, let $\mathcal{D}_p^{int}$ be the subcomplex of internal faces of $\mathcal{D}_p$.

For any fixed $x \in \mathbb{R}^n$, the projection of the face of $\mathscr{H}_m^p$ maximizing the dot product with $(-x, 1)$ gives a matroid M_x with set of bases B_x, and we have:

Theorem 1.3 (Speyer [Spe08]) *$x \in L(p)$ if and only if M_x is a loopless matroid.*

□

Tropical linear spaces can be defined differently. Let $p \in \mathcal{T}(\text{trop } S_{m,n})$. Consider the non-empty unbounded n-dimensional polyhedron $P_p := \{x \in \mathbb{R}^n \mid x^A \geq p_A \text{ for all } A \in \binom{[n]}{m}\}$. Let $\preceq$ denote componentwise inequality for vectors in $\mathbb{R}^n$, and define the **reduced tropical linear space** of p to be the set $P_p' := \{x \in P_p \mid y \preceq x \text{ with } y \in P_p \text{ implies } y = x\}$. The polyhedral complex P_p' is pure $(m-1)$-dimensional [Spe08] and coincides with the dual complex of $\mathcal{D}_p^{loop}$, see Herrmann and Joswig [HJ08, Proposition 2.3].

Take $x \in \partial(P_p)$ and let B_x' be the collection of sets $A \in \binom{[n]}{m}$ for which $x^A = p_A$. Then, $B_x' = \{A \in \binom{[n]}{m} \mid -x \cdot e^A + p_A \geq -x \cdot e^B + p_B \text{ for all } B \in \binom{[n]}{m}\}$, so $B_x' = B_x$ as above. Using the definition of P_p', we can check that $x \in P_p'$ if and only if M_x is a loopless matroid. In particular, (closed) bounded faces of P_p are faces of P_p'. Hence, we have:

- For $x \in \mathbb{R}^n$, $x \in L(p)$ if and only if M_x is loopless.
- For $x \in \partial(P_p)$, $x \in P_p'$ if and only if M_x is loopless.

For every $x \in \mathbb{R}^n$, there exists a unique t for which $x + te^{[n]} \in \partial(P_p)$. Also, $L(p)$ and the matroid M_x are invariant under translation by $e^{[n]}$, so $P_p' = L(p)/\mathbb{R}(1, \ldots, 1)$, under a natural choice of representative for each class.

Let the **tight span** $\mathscr{T}_p$ of p be the complex of bounded faces of P_p (or of P_p'). The tight span coincides with the dual complex of $\mathcal{D}_p^{int}$ [HJ08].

We now present a lemma on the topology of $L(p)$, P_p' and $\mathscr{T}_p$: all three spaces deformation retract onto any point of $\mathscr{T}_p$.

Lemma 1.4 *Let $P \subseteq \mathbb{R}^n$ be an n-dimensional unbounded polyhedron with no lines. Let $P_\emptyset$ be the complex of bounded faces of P and let $\mathcal{F}_\infty$ be the set of closed unbounded faces of P. For $A \subseteq \mathcal{F}_\infty$, define*

$$P_A = P_\emptyset \cup \left(\bigcup_{F \in A} F \right).$$

Then, P_A deformation retracts onto any point of $P_\emptyset$.

1.4 Results About Weighted Trees

The motivation of this subsection is the following result, known as the **four point condition theorem**. It is a theorem of Buneman [Bun74].

Theorem 1.5 (Pachter and Sturmfels [PS05]) *The set of* 2*-dissimilarity vectors of trees is equal to the tropical Grassmannian* $\mathcal{G}_{2,n}$*.*

□

From here, Pachter and Speyer [PS03] asked whether the set of m-dissimilarity vectors is contained in $\mathcal{G}_{m,n}$ for all $m \geq 3$. In Section 2, we answer this question affirmatively:

Theorem 1.6 (Iriarte [Iri10]) *Let* T *be a tree with* m*-dissimilarity vector* d *with* $m \geq 2$*. Then,* $d \in \mathcal{G}_{m,n}$*.*

□

This result is in the direction of solving the more general problem:

Problem 1.7 *Characterize the set of* m*-dissimilarity vectors of trees,* $m \geq 2$*.*

For the case $m = 2$, this problem is solved in a useful and concrete way by the four point condition theorem and by a result of Dress [Dre84] which relates tropical linear spaces and 2-dissimilarity vectors. The latter is also nicely obtained in Hirai [Hir06, Appendix A], with the additional verification that the face the author finds to prove it is indeed a bounded face of the reduced tropical linear space.

Theorem 1.8 (Dress [Dre84]) *A vector* $d \in \mathbb{R}^{\binom{[n]}{2}}$ *is a* 2*-dissimilarity vector if and only if* $\mathscr{T}_d$ *is a tree.*

□

In the same line of thought, using Part **b)** of Theorem 1.2 we can directly say:

Theorem 1.9 *A vector* $d \in \mathbb{R}^{\binom{[n]}{2}}$ *is a* 2*-dissimilarity vector if and only if every face of* $\mathcal{D}_d$ *is matroidal.*

□

Currently, no analogous results are known for $m \geq 3$.

2 Dissimilarity Vectors are Contained in the Tropical Grassmannian

Consider the tree T. We begin with a key lemma.

Lemma 2.1 (Cools [Coo09]) *Fix a set* $A \in \binom{[n]}{m}$*. Denote by* $\mathfrak{S}_A$ *the group of permutations of* A*. Let* $\mathcal{C}_A \subseteq \mathfrak{S}_A$ *be the set of cycles of length* m*. For any* $j \in [m] \cup \{0\}$*,* $a \in A$ *and* $\sigma \in \mathcal{C}_A$ *consider the following sum, which is independent of the choice of* a*:*

$$s_{A,\sigma} = \sum_{j=1}^{m} d_{\sigma^{j-1}(a)\sigma^{j}(a)}.$$

Then, there exists $\varsigma \in \mathcal{C}_A$ *(depending on* A*) such that* $d_A = \frac{1}{2} s_{A,\varsigma}$ *and* $s_{A,\varsigma} = \min_{\sigma \in \mathcal{C}_A} s_{A,\sigma}$*.*

□

In order to prove Theorem 1.6, it suffices only to consider the case of a trivalent tree T with rational weights. Moreover, Lemma 2.1 implies that if such T corresponds to an m-linear subspace V of $\mathbb{K}^n$, so that $d = \mathrm{val}(P)$ where P is a vector of Plücker coordinates of V, and T' is obtained from T by assigning new rational weights to the external edges, then the m-dissimilarity vector d' of T' corresponds analogously to a linear subspace V' obtained from V via a torus action on $\mathbb{K}^n$, so there exists a fixed $(t^{q_1}, t^{q_2}, \ldots, t^{q_n}) \in (\mathbb{K}^*)^n$ such that $V' = \{(t^{q_1}v_1, t^{q_2}v_2, \ldots, t^{q_n}v_n) \mid (v_1, v_2, \ldots, v_n) \in V\}$. Therefore, the following theorem implies Theorem 1.6:

Theorem 2.2 *Let U be an ultrametric tree with rational weights and m-dissimilarity vector d, $m \geq 2$. Then, $d \in \mathcal{G}_{m,n}$.*

□

Now, Theorem 2.2 follows from the following technical result proposed by Cools [Coo09]:

Proposition 2.3 (Iriarte [Iri10]) *Suppose that $3 \leq m \leq n$. Let U be an ultrametric tree with m-dissimilarity vector d, all of whose edges have rational weight. For each $e = (u, v) \in \mathcal{E}(U)$ with $u \leq_{\mathrm{TO}} v$, denote by $h(e)$ the total weight of the minimal path from u to any $j \in \mathcal{L}(U) = [n]$ such that $v \leq_{\mathrm{TO}} j$.*

For all $i \in [n-2]$ and $e \in \mathcal{E}(U)$, let $a_{i,e}$ be a generic complex number. Now, for all $i \in [n-2]$ and $j \in \mathcal{L}(U) = [n]$ let

$$\mathfrak{f}_{ij} := \sum_{\substack{e=(u,v)\in\mathcal{E}(U)\\ u\leq_{\mathrm{TO}} v\leq_{\mathrm{TO}} j}} a_{i,e} t^{h(e)},$$

and then define the $n \times n$ matrix:

$$M := \begin{pmatrix} 1 & 1 & \dots & 1 \\ \mathfrak{f}_{11} & \mathfrak{f}_{12} & \dots & \mathfrak{f}_{1n} \\ (\mathfrak{f}_{11})^2 & (\mathfrak{f}_{12})^2 & \dots & (\mathfrak{f}_{1n})^2 \\ \mathfrak{f}_{21} & \mathfrak{f}_{22} & \dots & \mathfrak{f}_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathfrak{f}_{(n-2)1} & \mathfrak{f}_{(n-2)2} & \dots & \mathfrak{f}_{(n-2)n} \end{pmatrix}.$$

For all $A \in \binom{[n]}{m}$, let $\mathfrak{m}_A$ be the $m \times m$ upper minor of M coming from the columns A. Then,

$$val(\mathfrak{m}_A) = d_A.$$

□

3 The Tropical Linear Space of a Dissimilarity Vector

In this section, we will further assume that our tree T is trivalent and that $1 < m < n$. Per Theorem 1.6, we can construct the tropical linear space P'_d of its m-dissimilarity vector. The purpose of the section is to present a geometric and combinatorial description of P'_d.

Let $S \subseteq_{\mathrm{st}} T$. Let $\mathtt{in}(S)$ be the subtree of S that consists of the internal edges and vertices of S. If S does not contain internal edges or vertices, as a technical convenience we let $\mathtt{in}(S) = \Lambda$, where Λ is the abstract tree consisting of an empty set of vertices and edges, and which we consider to be a proper subtree of every other tree. We call $\mathtt{in}(S)$ the **internal tree** of S. Let also $\overline{S}$ be the subtree of T consisting of all edges and vertices of T that are in S or adjacent to S. We call $\overline{S}$ the **extended tree** of S. For every leaf $i \in \mathcal{L}(T) = [n]$, there is a unique minimal path in T from i to S, denoted by $i \rightharpoonup S$.

Considering $i \in [m-1] \cup \{0\}$ and $S \subseteq_{\mathrm{st}} T$, we say that $(S, \mathcal{A})$ is an (m, i)-**good pair** of T if furthermore **a)** $i \leq |\mathcal{L}(S)| \leq m - 1$, **b)** $m - 1 + i \leq |\mathcal{V}(S)|$, **c)** $\mathcal{A} \subseteq \mathcal{L}(S)$ with $|\mathcal{A}| = i$, and **d)** $\mathcal{L}(S) \cap \mathcal{L}(T) \subseteq \mathcal{A}$.

3.1 Faces of P'_d and Their Matroids

We describe explicitly the faces of P'_d.

Let $S \subseteq_{\mathrm{st}} T$. For all $l \in \mathcal{L}(S)$, let H_l be the set of leaves of T whose minimal path to S meets l. Analogously, for all $l \in \mathcal{L}(\overline{S})$ let R_l be the set of leaves of T whose minimal path to $\overline{S}$ meets l. Let $\mathcal{R}_S = \{R_l\}_{l\in\mathcal{L}(\overline{S})}$. Then, $\mathcal{R}_S$ is a partition of $[n]$. For any $\mathcal{A} \subseteq \mathcal{L}(S)$, let $\mathcal{H}_{(S,\mathcal{A})} = \{H_l\}_{l\in\mathcal{A}}$. It is worth noting that $\mathcal{H}_{(S,\mathcal{A})}$ is also a collection of pairwise disjoint subsets of $[n]$, but not necessarily a partition of $[n]$. In fact, we will let $H^c_S := \{i \in [n] \mid i \notin H \text{ for all } H \in \mathcal{H}_{(S,\mathcal{L}(S))}\}$. Note also that each set in the collection $\mathcal{H}_{(S,\mathcal{A})}$ can be written as a disjoint union of at least two sets from the collection $\mathcal{R}_S$. If $\mathcal{H}_{(S,\mathcal{L}(S))}$ is a partition of $[n]$, we say that S is a **full subtree** of T.

For all $i \in [m-1] \cup \{0\}$, define

$$\begin{aligned}
\mathfrak{T}_i &:= \{(S,\mathcal{A}) \mid (S,\mathcal{A}) \text{ is an } (m,i)\text{-good pair of } T\},\\
\mathfrak{T}^{\mathtt{in}}_i &:= \{(S,\mathcal{A}) \in \mathfrak{T}_i \mid S \subseteq_{\mathrm{st}} \mathtt{in}(T)\},\\
\mathfrak{T}_* &:= \{S \mid S \text{ is a full subtree of } T \text{ with } m \text{ leaves}\},\\
\mathfrak{T}^{\mathtt{in}}_* &:= \{S \in \mathfrak{T}_* \mid S \subseteq_{\mathrm{st}} \mathtt{in}(T)\}.
\end{aligned}$$

We are now ready to present the main results of this section.

Theorem 3.1 *Let $i \in [m-2] \cup \{0\}$. Then, the set of i-dimensional faces of P'_d is in bijection with the set $\mathfrak{T}_i$, and the set of i-dimensional faces of $\mathscr{T}_d$ is in bijection with the set $\mathfrak{T}^{\mathtt{in}}_i$.*

□

Theorem 3.2 *The set of $(m-1)$-dimensional faces of P'_d is in bijection with the set $\mathfrak{T}_{m-1} \sqcup \mathfrak{T}_*$, and the set of $(m-1)$-dimensional faces of $\mathscr{T}_d$ is in bijection with the set $\mathfrak{T}^{\mathtt{in}}_{m-1} \sqcup \mathfrak{T}^{\mathtt{in}}_*$.*

□

For a tree $S \subseteq_{\mathrm{st}} T$ and sets $\mathcal{A} \subseteq \mathcal{L}(S)$, $\mathcal{B} \subseteq \mathcal{V}(T)$, let $S_{\mathcal{A}}$ be the subtree of S spanned by the set of vertices $\mathcal{V}(S)\backslash\mathcal{A}$, and let $S^{\mathcal{B}}$ be the subtree of T spanned by the set of vertices $\mathcal{V}(S) \cup \mathcal{B}$.

Proposition 3.3 *Let F and F' be faces of $\mathscr{T}_d$. Suppose that F corresponds to an $(m, |\mathcal{A}|)$-good pair $(S,\mathcal{A})$ of T and that F' corresponds to an $(m, |\mathcal{A}'|)$-good pair $(S',\mathcal{A}')$. Then, $F' \subseteq F$ if and only if*

$$S_{\mathcal{A}} \subseteq_{\mathrm{st}} S' \subseteq_{\mathrm{st}} S \text{ and } \mathcal{A}' \subseteq \mathcal{A}.$$

□

Let us now describe P'_d in greater detail.

Theorem 3.4 *Let $i \in [m-1] \cup \{0\}$. The matroid M_F of the face F of P'_d corresponding to an (m,i)-good pair $(S,\mathcal{A})$ of T has a collection of bases*

$$B_F = B^1 \cap B^2 \cap B^3,$$

where:

$B^1 := \{A \in \binom{[n]}{m} \mid |A \cap H| = 1 \text{ for all } H \in \mathcal{H}_{(S,\mathcal{A})}\}$,

$B^2 := \{A \in \binom{[n]}{m} \mid |A \cap H| \geq 1 \text{ for all } H \in \mathcal{H}_{(S,\mathcal{L}(S)\backslash\mathcal{A})}\}$,

$B^3 := \{A \in \binom{[n]}{m} \mid |A \cap R| \leq 1$ *for all* $R \in \mathcal{R}_S\}$.

If F *is bounded, then its face lattice is isomorphic to the face lattice of the* i*-dimensional cube in* $\mathbb{R}^n$.

Proof Sketch: We only describe the lattice isomorphism when F is bounded. Using Proposition 3.3, we see that the faces of F are described in a natural way as the set of triples of sets $\mathcal{C} \subseteq \mathcal{B} \subseteq \mathcal{A}$, via the good pair $\left(S^{\mathcal{B}}_{\mathcal{A}}, \mathcal{C}\right)$. Now, consider the map

$$\begin{aligned}\phi : \{\text{Faces of } F\} &\mapsto \{\text{Faces of the 0-1 } i\text{-dimensional cube in } \mathbb{R}^{\mathcal{A}}\},\\ \mathcal{C} \subseteq \mathcal{B} \subseteq \mathcal{A} &\mapsto \text{Convex hull of } \textbf{all} \text{ 0-1 vectors with 0's in } \mathcal{A}\backslash\mathcal{B} \text{ and 1's in } \mathcal{B}\backslash C.\end{aligned}$$

Then, ϕ is a lattice isomorphism. □

Consider the $(m-1)$-dimensional **pyrope** $\mathscr{P}_{m-1} :=$ conv $\left([-1,0]^{m-1} \cup [1,0]^{m-1}\right)$, *i.e.* the convex hull of the 0-1 and 0-(-1) cubes in $\mathbb{R}^{m-1}$. See Joswig and Kulas [JK10] for a detailed description of this tropical polytope.

Theorem 3.5 *The matroid* M_F *of an* $(m-1)$*-dimensional face* F *of* P'_d *corresponding to a full subtree* S *of* T *with* m *leaves has a collection of bases*

$B_F := \{A \in \binom{[n]}{m} \mid |A \cap H| = 1$ *for all* $H \in \mathcal{H}_{(S,\mathcal{L}(S))}\}$*, so it is a transversal matroid.*

If F *is bounded, then its face lattice is isomorphic to the face lattice of the* $(m-1)$*-dimensional pyrope.*

Proof Sketch: We only describe the lattice isomorphism when F is bounded. Fix a choice of $l \in \mathcal{L}(S)$. Per Proposition 3.3, the proper faces of F can be described as pairs of sets $(\mathcal{A}, \mathcal{B})$ satisfying $\emptyset \subsetneq \mathcal{A} \subsetneq \mathcal{L}(S)$ and $\mathcal{B} \subsetneq \mathcal{A}$, via the good pair $\left(\mathfrak{in}(S)^{\mathcal{A}}, \mathcal{B}\right)$. For one such pair $(\mathcal{A}, \mathcal{B})$, consider the finite set $V_{\mathcal{A},\mathcal{B}} \subseteq \mathbb{R}^{\mathcal{L}(S)\backslash l}$ that we now construct. For each $\mathcal{C} \subseteq \mathcal{B}$:

- If $l \in \mathcal{L}(S)\backslash \left(\mathcal{A}\backslash\mathcal{C}\right)$, define $v \in \mathbb{R}^{\mathcal{L}(S)\backslash l}$ as the vector of 1's in $\mathcal{A}\backslash\mathcal{C}$ and 0's everywhere else. Then, let $v \in V_{\mathcal{A},\mathcal{B}}$.
- If $l \in \mathcal{A}\backslash\mathcal{C}$, define $v \in \mathbb{R}^{\mathcal{L}(S)\backslash l}$ as the vector of 0's in $\mathcal{A}\backslash\mathcal{C}$ and -1's everywhere else. Then, let $v \in V_{\mathcal{A},\mathcal{B}}$.

Consider the map

$$\phi : \{\text{ Faces of } \partial(F)\ \} \mapsto \{\text{ Faces of } \partial(\mathscr{P}_{m-1}) \text{ in } \mathbb{R}^{\mathcal{L}(S)\backslash l}\},$$

under which the face of $\partial(F)$ corresponding to $(\mathcal{A}, \mathcal{B})$ gets mapped to the convex hull of the set $V_{\mathcal{A},\mathcal{B}}$.
Then, ϕ is a lattice isomorphism. □

In Figures 1, 2 and 3 we present an example of the tight spans of a tree with 9 leaves for $m = 2, 3, 4$.
Notice that Theorems 3.4 and 3.5 describe explicitly the family of matroids

$$\mathfrak{M}^{loop}_d = \{B_F \mid F \text{ is a face of } P'_d\}.$$

Of main interest are the matroids of the $(m-1)$-dimensional faces of P'_d. These turn out to be transversal matroids which encode the combinatorial structure of T in a natural way, as we now explain. Note that an $(m-1)$-dimensional face F of P'_d corresponds to a tree $S \subseteq_{\mathrm{st}} T$ such that:

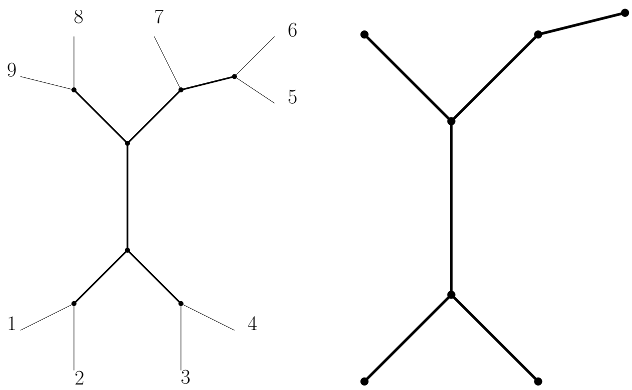

Fig. 1: On the left we have a tree T with 9 leaves. On the right we show its tight span for $m = 2$. Note that it is isomorphic to $\mathfrak{in}(T)$.

- **(case 1)** either S has $m-1$ leaves and at least $2(m-1)$ vertices, so $(S, \mathcal{L}(S))$ is an $(m, m-1)$-good pair of T,
- **(case 2)** or S has m leaves and exactly $2(m-1)$ vertices, so S is a full subtree of T with m leaves.

For both cases, M_F is a transversal matroid and B_F is the set of transversals of the collection $\mathcal{H}_{(S,\mathcal{L}(S))} \cup \{H_S^c\}$.

Conjecture 3.6 *Assume $n \geq 2m-1$. Then, the collection of bases of matroids*

$$\{B_F \mid F \textit{ is an } (m-1)\textit{-dimensional face of } P'_d\}$$

recovers uniquely the shape of T.

In particular, the collection $\mathcal{H}_{(S,\mathcal{L}(S))} \cup \{H_S^c\}$ can be recovered by computing the 2-circuits of M_F.

We propose the following problems:

Problem 3.7 *Characterize the family of matroids $\mathfrak{M}_d^{loop}$ when d is the m-dissimilarity vector of a tree T.*

And more ambitiously,

Problem 3.8 *Characterize the family of matroids $\mathfrak{M}_p^{loop}$ when $p \in \mathcal{G}_{m,n}$.*

We now discuss some aspects of these results. From the point of view of $\mathcal{D}_d$, the most relevant faces of P'_d are its vertices. Let us briefly describe how we found these.

Note that an $(m, 0)$-good pair $(S, \emptyset)$ of T is equivalent to a tree $S \subseteq_{\mathrm{st}} \mathfrak{in}(T)$ satisfying the inequalities $|\mathcal{L}(S)| \leq m-1 \leq |\mathcal{V}(S)|$.

Now, the following result holds:

Proposition 3.9 *Let $S \subseteq_{\mathrm{st}} \mathfrak{in}(T)$ satisfy the inequalities $|\mathcal{L}(S)| \leq m-1 \leq |\mathcal{V}(S)|$, and define $x \in \mathbb{R}^n$ by:*

$$x_i = \omega(i \rightharpoonup S) + \frac{\omega(S)}{m} \textit{ for all } i \in [n].$$

Then, x is a vertex of P'_d and $B_x = B^1 \cap B^2$, where:

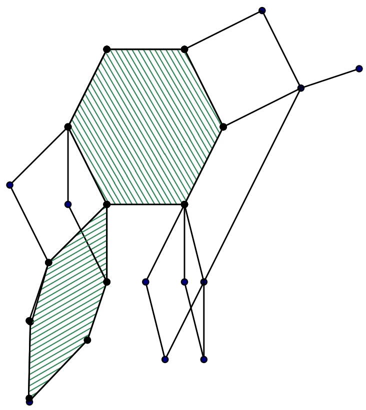

Fig. 2: The tight span of the tree T in Figure 1 for $m = 3$. Hexagons (which correspond to 2-dimensional pyropes) are shown in green. Enclosed regions are 2-dimensional faces and they are 2-dimensional cubes.

$$B^1 := \{A \in \tbinom{[n]}{m} \mid |A \cap H| \geq 1 \text{ for all } H \in \mathcal{H}_{(S,\mathcal{L}(S))}\},$$

$$B^2 := \{A \in \tbinom{[n]}{m} \mid |A \cap R| \leq 1 \text{ for all } R \in \mathcal{R}_S\}.$$

□

The subpolytope of $\mathscr{H}_m$ corresponding to the vertex x of P'_d found in Proposition 3.9 is then $\mathcal{P}_x := \text{conv}\{e^A \in \mathbb{R}^n \mid A \in B_x\}$. We would like an inequality description of $\mathcal{P}_x$.

Lemma 3.10 *We have:*

$$\mathcal{P}_x = \{y \in \mathscr{H}_m \mid \sum_{j \in H} y_j \geq 1 \text{ for all } H \in \mathcal{H}_{(S,\mathcal{L}(S))} \text{ and } \sum_{j \in R} y_j \leq 1 \text{ for all } R \in \mathcal{R}_{\mathcal{S}}\}.$$

□

We then use Lemma 3.10 to justify why all vertices of P'_d are obtained in this way. This result was a key step in our study, so we present the full proof.

Proposition 3.11 *Let y be a generic point in the interior of $\mathscr{H}_m$. Then, there exists a vertex x of P'_d corresponding to an $(m, 0)$-good pair $(S, \emptyset)$ of T, such that $y \in \mathcal{P}_x$.*

Proof: Let $v \in \mathcal{V}(T)$. The vertex v defines a partition Π_v on the set of leaves $[n]$ of T, under which two different leaves $i, j \in [n]$ belong to the same class if and only if the minimal paths from i to v and from j to v contain a common edge adjacent to v.

For each $\overline{a} \in \Pi_v$, let $\sigma_{\overline{a}} = \sum_{i \in \overline{a}} y_i$, and introduce a subset V_y of the set of internal vertices of T:

$$V_y = \{v \in \mathcal{V}(\mathtt{in}(T)) \mid m - \sigma_{\overline{a}} > 1 \text{ for all } \overline{a} \in \Pi_v\}.$$

We prove that $V_y \neq \emptyset$. To begin, notice that if we take a vertex $v \in \mathcal{V}(\mathtt{in}(T)) \setminus V_y$ and there exist two different classes $\overline{a}, \overline{b} \in \Pi_v$ such that $\sigma_{\overline{a}} > m-1$ and $\sigma_{\overline{b}} > m-1$, then $m \geq \sigma_{\overline{a}} + \sigma_{\overline{b}} > 2m-2$ or $m < 2$.

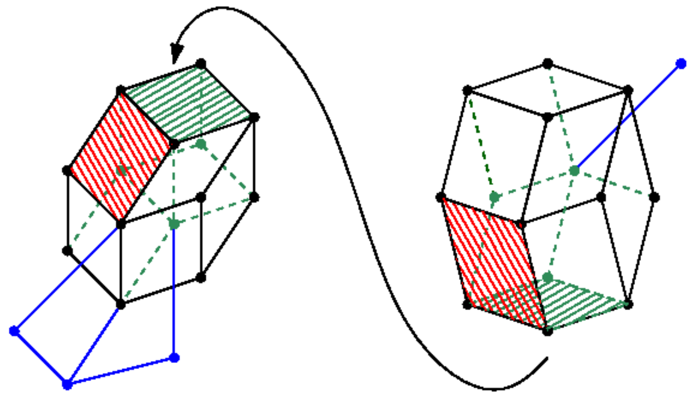

Fig. 3: The tight span of the tree T in Figure 1 for $m = 4$. On the left we have a complex of three 3-dimensional cubes pasted together, and two 2-dimensional cubes (in blue). On top of this, we paste as indicated a rhombic dodecahedron (which corresponds to a 3-dimensional pyrope) with a 1-dimensional cube appended to it (in blue).

This contradicts one of the earliest assumptions made in this section. Thus, for each $v \in \mathcal{V}(\mathrm{in}(T)) \setminus V_y$ there exists a unique $\overline{a} \in \Pi_v$ such that $\sigma_{\overline{a}} > m - 1$.

If $V_y = \emptyset$, among all pairs $(v, \sigma_{\overline{a}})$ with $v \in \mathcal{V}(\mathrm{in}(T))$ and $\overline{a} \in \Pi_v$ such that $\sigma_{\overline{a}} > m - 1$, take the one with $\sigma_{\overline{a}}$ minimal. The class $\overline{a}$ corresponds to a unique edge e of T adjacent to v. Let u be the other vertex adjacent to the edge e. If $u \in \mathcal{V}(\mathrm{in}(T))$ so that $\deg(u) = 3$, then by the minimality of $\sigma_{\overline{a}}$ and because y is an interior point of $\mathscr{H}_m$, it can only be the case that $\sigma_{\overline{b}} > m - 1$, where $\overline{b} \in \Pi_u$ is the class corresponding to the edge e. However, in this latter case we obtain $m = \sigma_{\overline{a}} + \sigma_{\overline{b}} > 2m - 2$ or $m < 2$, a contradiction. Finally, if $u \notin \mathcal{V}(\mathrm{in}(T))$, then u is a leaf of T. But then $1 > y_u = \sigma_{\overline{a}} > m - 1$ or $m < 2$, the first inequality coming from the fact that $y \in \mathscr{H}_m$.

Therefore, $V_y \neq \emptyset$.

It is not difficult to check that V_y defines a tree $S \subseteq_{\mathrm{st}} \mathrm{in}(T)$, so that $V_y = \mathcal{V}(S)$. More precisely, if we have three different vertices $u, v \in V_y$ and $w \in \mathcal{V}(\mathrm{in}(T))$ such that w lies in the minimal path from u to v, then $w \in V_y$. For example, if $\sigma_{\overline{c}} > m - 1$ for some $\overline{c} \in \Pi_w$, then that implies that either $\sigma_{\overline{a}} > m - 1$ or $\sigma_{\overline{b}} > m - 1$ holds for some $\overline{a} \in \Pi_v$ or some $\overline{b} \in \Pi_u$.

Now, $|\mathcal{L}(S)| \leq m - 1$. Otherwise, for every leaf l of S, let $\overline{a_l} \in \Pi_l$ correspond to the edge of S adjacent to l. Then:

$$m \geq \sum_{l \in \mathcal{L}(S)} (m - \sigma_{\overline{a_l}}) > \sum_{l \in \mathcal{L}(S)} 1 = |\mathcal{L}(S)| \geq m.$$

Also, $|\mathcal{V}(S)| = |V_y| \geq m - 1$. To see this, note that $\mathcal{L}(\overline{S}) \cap V_y = \emptyset$. For each $l \in \mathcal{L}(\overline{S})$, let $\overline{a_l} \in \Pi_l$ correspond to the edge e adjacent to S. It has to be the case that $\sigma_{\overline{a_l}} > m - 1$ or $1 > m - \sigma_{\overline{a_l}}$. Otherwise, if we let v be the vertex of V_y adjacent to e and $\overline{a} \in \Pi_v$ be the class corresponding to the edge e, then we would have $\sigma_{\overline{a}} > m - 1$, contradicting the fact that $v \in V_y$. But then, as the set of leaves $\mathcal{L}(\overline{S})$ induces a partition of $[n]$, we see that:

$$m = \sum_{l \in \mathcal{L}(\overline{S})} (m - \sigma_{\overline{a_l}}) < \sum_{l \in \mathcal{L}(\overline{S})} 1 = |\mathcal{L}(\overline{S})| \leq m.$$

Hence, $(S, \emptyset)$ is an $(m, 0)$-good pair of T and corresponds to a vertex x of P'_d.

If $l \in \mathcal{L}(S)$ and $\overline{a} \in \Pi_l$ is the class corresponding to the edge of S adjacent to l, then $m - \sigma_{\overline{a}} > 1$. On the other hand, if $l \in \mathcal{L}(\overline{S})$ and $\overline{a} \in \Pi_l$ is the class corresponding to the edge adjacent to both S and l, then $m - \sigma_{\overline{a}} < 1$. Therefore, $y \in \mathcal{P}_x$ per the description of $\mathcal{P}_x$ in terms of inequalities. $\square$

Acknowledgements

This project was developed under the advising of Federico Ardila at San Francisco State University. Special thanks go to Federico for several helpful discussions and careful readings, and to Serkan Hosten and Matthias Beck for careful readings. Thanks to Bernd Sturmfels, Lior Pachter and Filip Cools for helpful discussions.

References

[Bun74] P. Buneman. A note on metric properties of trees. *J. Combin. Theory Ser. B*, 17:48–50, 1974.

[Coo09] Filip Cools. On the relation between weighted trees and tropical grassmannians. http://arxiv.org/abs/0903.2010, 2009.

[Dre84] A.W.M. Dress. Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces. *Adv. Math.*, 53:321–402, 1984.

[Hir06] Hiroshi Hirai. Characterization of the distance between subtrees of a tree by the associated tight span. *Annals of Combinatorics*, 10:111–128, 2006.

[HJ08] Sven Herrmann and Michael Joswig. Splitting polytopes. *Münster J. of Math.*, 1:109–142, 2008.

[Iri10] Benjamin Iriarte. Dissimilarity vectors of trees are contained in the tropical grassmannian. *The Electronic Journal of Combinatorics*, 17:N1, 2010.

[JK10] Michael Joswig and Katja Kulas. Tropical and ordinary convexity combined. *Adv. Geometry*, 10:333–352, 2010.

[PS03] Lior Pachter and David Speyer. Reconstructing trees from subtree weights. http://arxiv.org/abs/math/0311156, 2003.

[PS05] Lior Pachter and Bernd Sturmfels. Algebraic statistics for computational biology. *Cambridge University Press, New York*, 2005.

[Spe08] David Speyer. Tropical linear spaces. *SIAM J. Discrete Math.*, 22, Issue 4:1527–1558, 2008.

[SS04] David Speyer and Bernd Sturmfels. The tropical grassmannian. *Adv. Geom.*, 4, No. 3:389–411, 2004.