Characterization of the Distance between Subtrees of a Tree by the Associated Tight Span

Hiroshi HIRAI
Research Institute for Mathematical Sciences,
Kyoto University, Kyoto 606-8502, Japan
hirai@kurims.kyoto-u.ac.jp

July 2004,
April 2005 (revised)

Abstract

A characterization is given to the distance between subtrees of a tree defined as the shortest path length between subtrees. This is a generalization of the four-point condition for tree metrics. For this, we use the theory of the tight span and obtain an extension of the famous result by A. Dress that a metric is a tree metric if and only if its tight span is a tree.

1 Introduction

Recently, mathematical treatments of phylogenetics have come to be increasingly important; see [2],[17]. The central problem in phylogenetics is reconstructing phylogenetic trees from given experimental data, e.g., DNA sequences. If the data is given as a distance matrix expressing dissimilarity between species, the problem is to search for a *tree metric* that "fits" the given distance matrix.

For a finite set X and a distance $d : X \times X \to \mathbf{R}$ with $d(x,x) = 0$ and $d(x,y) = d(y,x) \geq 0$ for $x, y \in X$, d is said to be a metric if it satisfies the triangle inequality, and a tree metric if there exists some weighted tree such that d can be expressed by the path metric between vertices of the tree. One of the most fundamental theorems in phylogenetics is the characterization of tree metrics.

Theorem 1.1 ([23], [18], [3], [4]). *A metric d is a tree metric if and only if it satisfies the four-point condition*

$$\begin{aligned} &\forall x, y, z, w \in X, \ |\{x, y, z, w\}| = 4, \\ &d(x, y) + d(z, w) \leq \max\{d(x, z) + d(y, w), d(x, w) + d(y, z)\}. \end{aligned} \quad (1.1)$$

In this paper, we generalize this characterization for the distance between subtrees of a tree. We define the distance on subtrees of a tree by the shortest path length between subtrees (see Figure 1).

Our main result is as follows:

Theorem 1.2. *A distance d can be expressed as the distance between subtrees of some tree if and only if it satisfies*

$$\begin{aligned}
&\forall x, y, z, w \in X,\ |\{x, y, z, w\}| = 4, \\
&d(x, y) + d(z, w) \leq \\
&\max \left\{ \begin{array}{l} d(x, z) + d(y, w),\ d(x, w) + d(y, z),\ d(x, y),\ d(z, w), \\ \dfrac{d(x, y) + d(y, z) + d(z, x)}{2},\ \dfrac{d(x, y) + d(y, w) + d(w, x)}{2}, \\ \dfrac{d(x, z) + d(z, w) + d(w, x)}{2},\ \dfrac{d(y, z) + d(z, w) + d(w, y)}{2} \end{array} \right\} \quad (1.2)
\end{aligned}$$

If d satisfies the triangle inequality, then it can be verified that (1.2) coincides with the four-point condition (1.1) (see Remark 2.5). Hence (1.2) is a generalization of the four-point condition.

For the proof of Theorem 1.2, we use the theory of the *tight span*, which was discovered independently by J. R. Isbell [14], A. Dress [6] and M. Chrobak and L.L. Larmore [5] and developed by A. Dress and coworkers [8]. Whereas the tight span has so far been considered essentially for a metric, in this paper, we consider the tight span for a distance that may violate the triangle inequality.

This paper is organized as follows. In Section 2, we prepare definitions and notation, and present a more general version of Theorem 1.2. In Section 3, we give the proof of the theorems.

2 Definitions, Notation and Results

2.1 Distances and partial splits

Let X be a finite set. A function $d : X \times X \to \mathbf{R}$ is said to be a *distance* on X if d satisfies $d(x, x) = 0$ and $d(x, y) = d(y, x) \geq 0$ for $x, y \in X$. A distance d is said to be a *metric* if, in addition, d satisfies $d(x, y) \leq d(x, z) + d(y, z)$ for $x, y, z \in X$. For $A, B \subseteq X$ with $A \cap B = \emptyset$ and $A, B \neq \emptyset$, the unordered pair $\{A, B\}$ is called a *partial split* on X. If a partial split $\{A, B\}$ satisfies $A \cup B = X$, then $\{A, B\}$ is called a *split* on X. For a partial split $\{A, B\}$ on X, we define a *partial split distance* $\zeta_{\{A,B\}} : X \times X \to \mathbf{R}$ by

$$\zeta_{\{A,B\}}(x, y) = \begin{cases} 1 & \text{if } x \in A,\ y \in B \text{ or } y \in A,\ x \in B \\ 0 & \text{otherwise.} \end{cases} \quad (2.1)$$

Note that $\zeta_{\{A,B\}}$ is not a metric if $A \cup B \neq X$ and is a metric, called a *split metric*, if $A \cup B = X$. A pair of partial splits $\{A, B\}$ and $\{C, D\}$ on X is said to be *compatible* if it satisfies one of the following four conditions:

$$A \subseteq C \text{ and } B \supseteq D, \quad (2.2)$$
$$A \subseteq D \text{ and } B \supseteq C, \quad (2.3)$$
$$A \supseteq C \text{ and } B \subseteq D, \quad (2.4)$$
$$A \supseteq D \text{ and } B \subseteq C. \quad (2.5)$$

A collection of partial splits $\mathcal{S}$ is said to be *compatible* if any pair of partial splits in $\mathcal{S}$ is compatible. Note that if $\mathcal{S}$ consists of splits, then compatibility in our sense coincides with compatibility of splits in the standard definition; see [3], [2], [17].

2.2 Graphs

For a weighted graph $G = (V, E, w)$ with a vertex set V, an edge set E, and a positive weight $w : E \to \mathbf{R}$ representing edge lengths, $D_G : V \times V \to \mathbf{R}$ denotes the path metric on G defined by the shortest length of a path. We also denote vertices of G by $V(G)$ and edges of G by $E(G)$.

2.3 Tight span of distances

Next we introduce the tight span and related concepts. For a distance $d : X \times X \to \mathbf{R}$, a polyhedron $P(X, d) \subseteq \mathbf{R}^X$ associated with d is defined as

$$P(X, d) = \{f \in \mathbf{R}^X \mid f(x) + f(y) \geq d(x, y) \ (x, y \in X)\}. \tag{2.6}$$

The tight span $T(X, d)$ is defined to be the union of bounded faces of $P(X, d)$, or equivalently,

$$T(X, d) = \{f \in \mathbf{R}^X \mid \forall x \in X,\ f(x) = \max_{y \in X}\{d(x, y) - f(y)\}\}. \tag{2.7}$$

The dimension of $T(X, d)$ is defined to be the maximum dimension of bounded faces of $P(X, d)$. As indicated by [6, Remark 5.4, p.370], $\dim T(X, d)$ can be characterized as follows, whether d is a metric or not.

Theorem 2.1 ([6])**.** *For a distance $d : X \times X \to \mathbf{R}$ and a positive integer n, the following two conditions are equivalent.*

(*a*) $\dim T(X, d) \geq n$.

(*b*) *There exists a $2n$-element subset $\{x_1, x_{-1}, x_2, x_{-2}, \ldots, x_n, x_{-n}\} \subseteq X$ such that*

$$\sum_{i \in I} d(x_i, x_{-i}) > \sum_{i \in I} d(x_i, x_{\sigma(i)}) \tag{2.8}$$

holds for any permutation σ of $I = \{\pm 1, \pm 2, \ldots, \pm n\}$ not satisfying $\sigma(i) = -i$ for all $i \in I$.

In the appendix, we give a simple proof of Theorem 2.1 based on standard arguments in linear programming.

Let $t^d : X \to 2^{T(X,d)}$ be defined as

$$t^d(x) = T(X, d) \cap \{f \in \mathbf{R}^X \mid f(x) = 0\} \quad (x \in X), \tag{2.9}$$

which is also the union of the bounded faces of

$$\{f \in \mathbf{R}^X \mid f(y) + f(z) \geq d(y, z) \ (y, z \in X),\ f(x) = 0\}. \tag{2.10}$$

Then $T(X, d)$ and $t^d(x)$ are contractible since the union of the bounded faces of a polyhedron is contractible; see Lemma A.5 in Appendix. Note that contractibility of $T(X, d)$ in the case that d is a metric is shown in [6, (1.10), p.332].

We define a weighted graph $G(d)$ by the 1-skeleton of $T(X, d)$ endowed with the $\| \cdot \|_\infty$ norm of $\mathbf{R}^X$. For $x \in X$, let $g^d(x)$ be defined by the graph corresponding to the 1-skeleton of $t^d(x)$, which is a connected subgraph of $G(d)$.

The following shows that in the case that d is a metric, $t^d(x)$ is a single point of $T(X, d)$ that coincides with the canonical map $X \to T(X, d)$.

Lemma 2.2. *If d is a metric, then we have $t^d(x) = \{h_x\}$ for $x \in X$, where $h_x \in \mathbf{R}^X$ is defined as*

$$h_x(y) = d(x, y) \quad (y \in X). \tag{2.11}$$

Proof. Let $f \in t^d(x)$. Then we have $f(z) \geq d(x, z)$ for $z \in X$ since $f(x) = 0$. For $y \in X$, by $f \in T(X, d)$, there exists $w \in X$ such that $f(y) + f(w) = d(y, w)$. By the triangle inequality, we have $d(y, x) + d(w, x) \leq f(y) + f(w) = d(y, w) \leq d(x, y) + d(x, w)$. Hence we obtain $f(y) = d(x, y)$. □

2.4 Results

We present a more general version of Theorem 1.2 below, which is also an extension of (a finite dimensional version of) the result of A. Dress [6] that a metric is a tree metric if and only if its tight span is a tree. In this paper, a *subtree* means a subgraph which is a tree.

Theorem 2.3. *For a distance $d : X \times X \to \mathbf{R}$, the following conditions are equivalent.*

(*a*) *There exist some weighted tree T and a family of its subtrees T_x $(x \in X)$ such that*

$$d(x, y) = \min\{D_T(u, v) \mid u \in V(T_x), v \in V(T_y)\} \quad (x, y \in X). \tag{2.12}$$

(*b*) *There exist some compatible collection of partial splits $\mathcal{S}$ on X and a positive weight $\alpha : \mathcal{S} \to \mathbf{R}$ such that*

$$d = \sum_{S \in \mathcal{S}} \alpha_S \zeta_S. \tag{2.13}$$

(*c*) *$G(d)$ is a tree.*

(*d*) *$T(X, d)$ is a tree.*

(*e*) $\dim T(X, d) \leq 1$.

(*f*) *d satisfies the condition* (1.2).

The essential part of the proof of Theorem 2.3 relies on the following, which is an extension of the fact that a finite metric space (X, d) can be isometrically embedded into $(T(X, d), \|\cdot\|_\infty)$ and realized by the 1-skeleton of $T(X, d)$ [6].

Theorem 2.4. *For a distance $d : X \times X \to \mathbf{R}$, the following holds.*

(1) $d(x, y) = \inf\{\|f - g\|_\infty \mid f \in t^d(x), g \in t^d(y)\} \quad (x, y \in X)$.

(2) $d(x, y) = \min\{D_{G(d)}(u, v) \mid u \in V(g^d(x)), v \in V(g^d(y))\} \quad (x, y \in X)$.

Remark 2.5. We show that the condition (1.2) reduces to the four-point condition (1.1) for a metric d. From the triangle inequality, we have

$$d(x, y) \leq \frac{1}{2}\{d(x, z) + d(z, y)\} + \frac{1}{2}\{d(x, w) + d(w, y)\}. \tag{2.14}$$

This implies that $d(x, y) \leq \max\{d(x, z) + d(y, w), d(x, w) + d(z, y)\}$. Similarly,

$$\{d(x, y) + d(y, z) + d(z, x)\}/2 \leq \{d(x, w) + d(w, y) + d(y, z) + d(z, x)\}/2 \quad (2.15)$$

implies

$$\{d(x, y) + d(y, z) + d(z, x)\}/2 \leq \max\{d(x, z) + d(y, w), d(x, w) + d(y, z)\}.$$

Remark 2.6. Every 3-point distance can be expressed as Theorem 2.3 (a). Let $d : \{1, 2, 3\} \times \{1, 2, 3\} \to \mathbf{R}$ be a distance on $\{1, 2, 3\}$. If d is a metric, then it is well known that d is a tree metric. Suppose that d does not satisfy the triangle inequality, say $d(1, 2) > d(1, 3) + d(2, 3)$. Consider a weighted tree $T = (\{i, j, k, l\}, \{ij, jk, kl\}, w)$ with edge length $w_{ij} = d(1, 3)$, $w_{jk} = d(1, 2) - d(1, 3) - d(2, 3)$ and $w_{kl} = d(2, 3)$, and a family of its subtrees $\{T_1 = (\{i\}, \emptyset), T_2 = (\{j, k\}, \{jk\}), T_3 = (\{l\}, \emptyset)\}$. Then they satisfy (2.12).

Remark 2.7. The split decomposition, due to Bandelt and Dress [1], has been extended in [12] for distances using partial split distances. A distance between subtrees of a tree, considered in this paper, is one of the examples of *totally split decomposable* distances in the sense of [12].

Remark 2.8. We give some remarks about the dual view of tight spans. Consider the point configuration $\mathcal{A}_{X,2} := \{\chi_x + \chi_y \mid x, y \in X\} \subseteq \mathbf{R}^X$; see the beginning of Section 3 for the definition of χ_x. Take the convex hull of $\{(\chi_x + \chi_y, d(x, y)) \mid x, y \in X\} \subseteq \mathcal{A}_{X,2} \times \mathbf{R}$, and project its upper faces to the convex hull of $\mathcal{A}_{X,2}$. Then we obtain a *regular subdivision* $\Delta(X, d)$ of $\mathcal{A}_{X,2}$. In fact, the tight span $T(X, d)$ is the union of dual faces of interior faces of $\Delta(X, d)$; see [12] and [20] for details. We see from this view point that $\dim T(X, d) \leq 1$ if and only if $\Delta(X, d)$ has no interior faces of codimension greater than 1. Furthermore, the condition (1.2) can be rephrased as follows:

> $\Delta(X, d)$ has no edge which can be represented as $[\chi_x + \chi_y, \chi_z + \chi_w]$
> for some distinct $x, y, z, w \in X$,

where $[p, q]$ denotes the closed line segment between p and q. Indeed, since the height of the upper envelope of the convex hull of $\{(\chi_x + \chi_y, d(x, y)) \mid x, y \in X\}$ at $(\chi_x + \chi_y + \chi_z + \chi_w)/2$ is given by a quarter of the optimal value of the linear program (A.7) for $Y = \{x, y, z, w\}$, we have that $[\chi_x + \chi_y, \chi_z + \chi_w]$ is an edge of $\Delta(X, d)$ if and only if the optimal value of (A.7) for $Y = \{x, y, z, w\}$ is uniquely attained by $2\chi_{\{xy,zw\}}$ if and only if $\dim T(X, d) > 1$ (see the condition (b') in Appendix). In particular, $\dim T(X, d) \leq 1$ if and only if each edge of $\Delta(X, d)$ is one of $[\chi_x + \chi_y, 2\chi_z]$, $[\chi_z + \chi_x, \chi_z + \chi_y]$, and $[2\chi_x, 2\chi_y]$. Furthermore, $[\chi_x + \chi_y, 2\chi_z]$ is an edge of $\Delta(X, d)$ if and only if $d(x, y) > d(x, z) + d(z, y)$ (consider the height of the upper envelope of the convex hull of $\{(\chi_x + \chi_y, d(x, y)) \mid x, y \in X\}$ at $(2\chi_z + \chi_x + \chi_y)/2$). Hence, d is a tree metric (d is a metric and $\dim T(X, d) \leq 1$) if and only if each edge in $\Delta(X, d)$ is parallel to $\chi_x - \chi_y$ for some $x, y \in X$. A polyhedron each of whose edges is parallel to $\chi_x - \chi_y$ is known as a *base polyhedron* or a *matroid polytope* for a $\{0, 1\}$-polytope; see [9] and [11] for base polyhedra, and this characterization by edge vectors is due to Tomizawa [21] and Gelfand, Goresky, MacPherson, and Serganova [10]. Subdivisions consisting of base polyhedra are called *matroid subdivisions*. Hence, d is a tree metric if and only if $\Delta(X, d)$ is a matroid subdivision. Matroid subdivisions appear in *tropical geometry* [19], *surgery on Grassmannians* [15], and *discrete convex analysis*; polyhedral convex functions whose lower faces induce a matroid subdivision are called *M-convex functions* in [16] (also see [13] for the relationship between M-convexity and tree metrics).

3 Proofs

In the following, let X be a finite set and $d : X \times X \to \mathbf{R}$ be a distance on X. For a set S, we denote by χ_S the characteristic vector of S defined as: $\chi_S(x) = 1$ if $x \in S$ and 0 otherwise. In particular we write simply χ_x instead of $\chi_{\{x\}}$ for a singleton $\{x\}$.

3.1 Preliminaries

For $f \in P(X, d)$, we define an undirected graph $K(f) = (X, E(f))$ by

$$xy \in E(f) \stackrel{\text{def}}{\Longleftrightarrow} f(x) + f(y) = d(x, y) \quad (x, y \in X), \tag{3.1}$$

where for $x, y \in X$, xy denotes an unordered pair, which means that xy and yx are not distinguished from each other. An edge is in $K(f)$ if f is in the facet of $P(X, d)$ corresponding to that edge. Note that $E(f)$ may contain loop edges, like xx for $x \in X$. Let $F(f)$ be the face of $P(X, d)$ that contains f in its relative interior, which is also the set of solutions to the linear inequalities

$$p(x) + p(y) = d(x, y) \quad (xy \in E(f)), \tag{3.2}$$
$$p(x) + p(y) \geq d(x, y) \quad (xy \notin E(f)). \tag{3.3}$$

By the same argument in the case that d is a metric [7], it is easy to observe that

$$\begin{aligned} f \in T(X, d) \quad &\Leftrightarrow \quad F(f) \text{ is bounded} && (3.4)\\ &\Leftrightarrow \quad K(f) \text{ does not have isolated vertices} && (3.5)\\ &\Leftrightarrow \quad \forall x \in X,\ f(x) = \max_{y \in X}\{f(y) - d(x, y)\}. && (3.6) \end{aligned}$$

For the subsequent arguments, we need a characterization of the dimension of $F(f)$. Since the dimension of $F(f)$ is given by the dimension of its affine hull (3.2), $\dim F(f)$ coincides with $|X|$ minus the rank of the matrix whose column vectors are $\{\chi_x + \chi_y \mid xy \in E(f)\}$. For a connected graph (X, E), we observe

$$\text{rank}\{\chi_x + \chi_y \mid xy \in E\} = \begin{cases} |X| - 1 & \text{if } (X, E) \text{ is bipartite,} \\ |X| & \text{if } (X, E) \text{ is nonbipartite,} \end{cases} \tag{3.7}$$

where loops are regarded as odd cycles. Therefore, if $f \in T(X, d)$, we have

$$\begin{aligned} \dim F(f) &= |X| - \text{rank}\{\chi_x + \chi_y \mid xy \in E(f)\} && (3.8)\\ &= \text{the number of bipartite components of } K(f). && (3.9) \end{aligned}$$

In particular, we have

$$F(f) \text{ is an edge} \quad \Leftrightarrow \quad K(f) \text{ has only one bipartite component}, \tag{3.10}$$
$$F(f) \text{ is a vertex} \quad \Leftrightarrow \quad K(f) \text{ has no bipartite components.} \tag{3.11}$$

The dimension of $T(X,d)$ is given by

$$\dim T(X,d) = \max_{f \in T(X,d)} \{\text{the number of bipartite components of } K(f)\}. \tag{3.12}$$

3.2 Proof of Theorem 2.4

Theorem 2.4 says

$$d(x,y) = \inf\{\|f-g\|_\infty \mid f \in t^d(x), g \in t^d(y)\}, \tag{3.13}$$
$$= \min\{D_{G(d)}(u,v) \mid u \in V(g^d(x)), v \in V(g^d(y))\}. \tag{3.14}$$

Let D_1 and D_2 be distances on X defined by the RHS in (3.13) and (3.14), respectively. We prove $d = D_1 = D_2$.

Lemma 3.1. *$d(x,y) \le D_1(x,y) \le D_2(x,y)$ holds for $x, y \in X$.*

Proof. For any $f \in t^d(x), g \in t^d(y)$, we have

$$f(x) = 0,\ f(y) \ge d(x,y),\ g(x) \ge d(x,y),\ g(y) = 0. \tag{3.15}$$

Hence we have $\|f-g\|_\infty \ge d(x,y)$. We may identify the graph $G(d)$ and the 1-skeleton of $T(X,d)$. Let $(f_0, f_1, \ldots, f_m)$ be a path of $G(d)$ with $f_0 \in V(g^d(x))$ and $f_m \in V(g^d(y))$. Hence the length of the path $(f_0, f_1, \ldots, f_m)$ is $\sum_{i=0}^{m-1} \|f_i - f_{i+1}\|_\infty \ge \|f_0 - f_m\|_\infty \ge D_1(x,y)$. □

In the following, we construct the path in $G(d)$ from $V(g^d(x))$ to $V(g^d(y))$ with its path length $d(x,y)$. This implies Theorem 2.4.

First, we take a vertex of $t^d(x)$. Let $X = \{x_1 = x, x_2 = y, x_3, \ldots, x_m\}$. Then, $f \in \mathbf{R}^X$ defined by

$$f(x_1) = 0,$$
$$f(x_i) = \max(0, \max_{k=1,\ldots,i-1}(d(x_i,x_k) - f(x_k))) \quad (i = 2, \ldots, m)$$

is a vertex of $t^d(x)$. Indeed, define $\{f^k\}_{k=1,\ldots,m} \subseteq \mathbf{R}^X$ by $f^k(x_i) = f(x_i)$ for $i \le k$ and $f^k(x_i) = +\infty$ (sufficiently large) for $i > k$. By induction on k, we see that $f^k \in P(X,d)$, $E(f^k) \subseteq E(f)$, and x_k is covered by some edge in $E(f^k)$ which is a loop ($f(x_k) = 0$), or is connected to some nonbipartite component ($f(x_k) = d(x_k,x_j) - f(x_j)$ for some $j < k$). Hence, $f = f^m$ is a vertex of $t^d(x)$ by (3.11). In particular, we have $xx, xy \in E(f)$, $f(y) = d(x,y)$, and $f(x) = 0$.

Next we try to move f toward $t^d(y)$ through edges of $T(X,d)$. If $yy \in E(f)$, then we have $f \in t^d(y)$ and $D_2(x,y) = D_1(x,y) = 0 = d(x,y)$. Hence we suppose $yy \notin E(f)$, i.e., $f(y) > 0$.

To move f in $T(X,d)$, we use stable sets of $K(f)$, where a vertex set $S \subseteq X$ is called a *stable set* of $K(f)$ if for any $x, y \in S$ we have $xy \notin E(f)$. For a subset $S \subseteq X$, we define the *neighborhood* $N(S)$ by $\{z \in X \setminus S \mid \exists w \in S,\ zw \in E(f)\}$. If $S \subseteq X$ is a stable set of $K(f)$, for sufficiently small $\epsilon > 0$, a vector

$f + \epsilon(\chi_{N(S)} - \chi_S)$ is also in $P(X, d)$. In particular, $\chi_{N(S)} - \chi_S$ is a feasible direction of $P(X, d)$ at f. We use this fact.

Let $S_y \subseteq X$ be a stable set of $K(f)$ constructed according to the following process, where $N(S_y \cup N(S_y))$ is the set of vertices at distance exactly 2 to S_y,

(S0) $S_y = \{y\}$.

(S1) If there is no loopless vertex in $N(S_y \cup N(S_y))$ then output S_y and stop.

(S2) Take a loopless vertex $z \in N(S_y \cup N(S_y))$.

(S3) $S_y \leftarrow S_y \cup \{z\}$ and go to (S1).

By this construction, we see that the graph

$$G_{S_y} = (S_y \cup N(S_y), \{zw \in E(f) \mid z \in S_y, w \in N(S_y)\}) \tag{3.16}$$

is a connected bipartite subgraph of $K(f)$. For $\epsilon \geq 0$, let $f^\epsilon \in \mathbf{R}^X$ be defined as

$$f^\epsilon = f + \epsilon(\chi_{N(S_y)} - \chi_{S_y}). \tag{3.17}$$

Let $\epsilon_0 > 0$ be defined by the maximum of $\epsilon \geq 0$ such that $f^\epsilon \in P(X, d)$. Then ϵ_0 is given by

$$\min \left\{ \begin{array}{c} \min\limits_{z,w \in S_y} (f(z) + f(w) - d(z,w))/2, \\ \min\limits_{z \in S_y, w \notin S_y \cup N(S_y)} f(z) + f(w) - d(z,w) \end{array} \right\}. \tag{3.18}$$

Then it is seen that

(1) $f^\epsilon \in T(X, d)$ for $0 \leq \epsilon \leq \epsilon_0$,

(2) $K(f^\epsilon)$ has one bipartite component G_{S_y} for $0 < \epsilon < \epsilon_0$, and

(3) $K(f^{\epsilon_0})$ has no bipartite components.

Indeed, each $z \notin S_y \cup N(S_y)$ is covered by some edge zw with $w \notin S_y \cup N(S_y)$ and each $z \in S_y \cup N(S_y)$ is covered by some edges of G_{S_y}. These edges remain in $K(f^\epsilon)$ for $0 \leq \epsilon \leq \epsilon_0$. This implies (1). For $0 < \epsilon < \epsilon_0$, any edge $zw \in E(f)$ with $z \in N(S_y), w \notin S_y$ vanishes in $(X, E(f^\epsilon))$, and each edge in G_{S_y} remains. This implies (2). In $K(f^{\epsilon_0})$, there exists some new edge $zw \in E(f^{\epsilon_0})$ such that $z, w \in S_y$ or $z \in S_y, w \notin S_y \cup N(S_y)$. In the former case, an odd cycle appears in the subgraph induced by $S_y \cup N(S_y)$. In the latter case, the bipartite component G_{S_y} is connected to some nonbipartite component. This implies (3).

By (3.10) and (3.11), the move $f \to f^{\epsilon_0}$ is on the edge of $T(X, d)$, f^{ϵ_0} is a vertex of $T(X, d)$, and we have

$$\|f^{\epsilon_0} - f\|_\infty = f^{\epsilon_0}(x) - f(x) = f(y) - f^{\epsilon_0}(y) = \epsilon_0 \tag{3.19}$$

by $y \in S_y$ and $x \in N(S_y)$. Put $f_1 = f^{\epsilon_0}$ and repeat this process for f_1. Note that $y \in S_y$ and $x \in N(S_y)$ always hold in each step of this process. Then we have the path $(f = f_0, f_1, f_2, \ldots)$ of $G(d)$. By (3.19), we have $f_0(y) > f_1(y) > \cdots$. After finitely many steps, we have $f_l(y) = 0$, $f_l(x) = d(x, y)$, and $f_l \in t^d(y)$. Therefore the path length of $(f = f_0, f_1, f_2, \ldots, f_l = g)$ is $\sum_{i=0}^{l-1} \|f_{i+1} - f_i\|_\infty = f(y) - g(y) = g(x) - f(x) = d(x, y)$.

3.3 Proof of Theorem 2.3

We restate six conditions of Theorem 2.3 as follows:

(a) There exist some weighted tree T and a family of its subtrees T_x $(x \in X)$ such that

$$d(x,y) = \min\{D_T(u,v) \mid u \in V(T_x), v \in V(T_y)\} \quad (x,y \in X).$$

(b) There exist some compatible collection of partial splits $\mathcal{S}$ on X and a positive weight $\alpha : \mathcal{S} \to \mathbf{R}$ such that

$$d = \sum_{S \in \mathcal{S}} \alpha_S \zeta_S.$$

(c) $G(d)$ is a tree.

(d) $T(X,d)$ is a tree.

(e) $\dim T(X,d) \leq 1$.

(f) d satisfies the condition (1.2).

We prove the equivalence of these conditions by showing the following:

$$\begin{array}{ccccc} (a) & \Leftarrow & (c) & \Leftarrow & (d) \\ \Downarrow & & & & \Updownarrow \\ (b) & \Rightarrow & (f) & \Leftrightarrow & (e) \end{array} \tag{3.20}$$

$(c) \Leftarrow (d)$ is obvious. $(a) \Leftarrow (c)$ follows from Theorem 2.4. $(d) \Leftrightarrow (e)$ follows from the contractibility of $T(X,d)$.

We show $(f) \Leftrightarrow (e)$ from Theorem 2.1 for $n = 2$. Recall the fact that every permutation can be uniquely decomposed to disjoint cyclic permutations. For a permutation σ of a 4-point set X, $d^\sigma := \sum_{i \in X} d(i, \sigma(i))$ is given as

$$d^\sigma = \begin{cases} 0 & \text{if } \sigma = \text{identity}, \\ 2d(x,y) & \text{if } \sigma = (x\ y), \\ 2d(x,y) + 2d(z,w) & \text{if } \sigma = (x\ y)(z\ w), \\ d(x,y) + d(y,z) + d(z,x) & \text{if } \sigma = (x\ y\ z), \\ d(x,y) + d(y,z) + d(z,w) + d(w,x) & \text{if } \sigma = (x\ y\ z\ w), \end{cases} \tag{3.21}$$

where $x, y, z, w \in X$ and $\sigma = (x_0\ x_1\ \cdots\ x_{m-1})$ means a cyclic permutation $\sigma(x_i) = x_{i+1 \bmod m}$. Note that $d^{(x_0 \cdots x_{m-1})} = d^{(x_{m-1} \cdots x_0)}$. Hence, Theorem 2.1 for $n = 2$ says that $\dim T(X,d) \leq 1$ if and only if

$$\forall x,y,z,w \in X \text{ (all distinct)}$$
$$d^{(xy)(zw)} \leq \max \left\{ \begin{array}{l} d^{\mathrm{id}}, \\ d^{(xy)}, d^{(xz)}, d^{(xw)}, d^{(yz)}, d^{(yw)}, d^{(zw)}, \\ d^{(xz)(yw)}, d^{(xw)(yz)}, \\ d^{(xyz)}, d^{(xyw)}, d^{(xzw)}, d^{(yzw)}, \\ d^{(xyzw)}, d^{(xywz)}, d^{(xzyw)} \end{array} \right\}. \tag{3.22}$$

Clearly, (1.2) implies (3.22). We show the converse. Since $d \geq 0$, $d^{(xz)(yw)} = d^{(xz)} + d^{(yw)}$ and $d^{(xw)(yz)} = d^{(xw)} + d^{(yz)}$, the terms d^{id}, $d^{(xz)}$, $d^{(yw)}$, $d^{(xw)}$ and

$d^{(yz)}$ are redundant in (3.22). Similarly, $d^{(xzyw)} = (d^{(xz)(yw)} + d^{(xw)(yz)})/2$ implies that $d^{(xzyw)}$ is also redundant. Suppose that d satisfies (3.22) and violates (1.2). Then we have $d^{(xz)(yw)} < d^{(xy)(zw)} \leq d^{(xywz)}$ or $d^{(xw)(yz)} < d^{(xy)(zw)} \leq d^{(xyzw)}$. Both inequalities contradict $d^{(xywz)} = (d^{(xz)(yw)} + d^{(xy)(zw)})/2$ and $d^{(xzyw)} = (d^{(xw)(yz)} + d^{(xz)(yw)})/2$. Hence we obtain the equivalence between (1.2) and (3.22).

Next we show $(a) \Rightarrow (b)$. Deletion of each edge e of T separates T into two trees T_e^A and T_e^B. From this, we have a disjoint pair $\{A_e, B_e\}$ defined as

$$A_e = \{x \in X \mid T_x \text{ is a subtree of } T_e^A\}, \tag{3.23}$$
$$B_e = \{x \in X \mid T_x \text{ is a subtree of } T_e^B\}. \tag{3.24}$$

For two edges $e, f \in E(T)$, we may assume that T_e^A is a subtree of T_f^A and T_f^B is a subtree of T_e^B. This implies the compatibility of $\{A_e, B_e\}$ and $\{A_f, B_f\}$. Hence we define the compatible collection of partial splits $\mathcal{S}$ on X and its positive weight $\alpha : \mathcal{S} \to \mathbf{R}$ by

$$\mathcal{S} = \{\{A_e, B_e\} \mid e \in E(T),\ \{A_e, B_e\} \text{ is a partial split}\}, \tag{3.25}$$
$$\alpha_{\{A_e, B_e\}} = \text{the length of edge } e\ . \tag{3.26}$$

Let $d' = \sum_{S \in \mathcal{S}} \alpha_S \zeta_S$. We show $d = d'$. Let $e \in E(T)$ be an edge with $\{A_e, B_e\} \in \mathcal{S}$. For $x \in A_e$ and $y \in B_e$, any path between T_x and T_y must contain e. This implies $d \geq d'$. Next we show $d \leq d'$. For $x, y \in X$, if T_x and T_y have a common vertex, i.e., $d(x, y) = 0$, then there is no edge in T that separates T_x and T_y. Hence we have $d(x, y) = d'(x, y) = 0$. Suppose that $d > 0$. Let $e \in E(T)$ be an edge of the shortest path between T_x and T_y. Neither T_x or T_y contains the edge e. Since both T_x and T_y are trees, it must be $x \in A_e,\ y \in B_e$ or $y \in A_e,\ x \in B_e$. Hence we have $\{A_e, B_e\} \in \mathcal{S}$. This implies $d \leq d'$.

$(b) \Rightarrow (f)$. It is sufficient to show this in the case that d is a distance on 4-point set. For this, we classify maximal compatible families of partial splits on the 4-point set $\{1, 2, 3, 4\}$. All partial splits on $\{1, 2, 3, 4\}$ are listed below, where we denote a partial split $\{\{1, 2\}, \{3\}\}$ simply by $12|3$:

(S1): $1|234$, $2|134$, $3|124$, $4|123$,

(S2): $12|34$, $13|24$, $23|14$,

(S3): $1|2$, $1|3$, $1|4$, $2|3$, $2|4$, $3|4$,

(S4): $1|23$, $2|13$, $3|12$, $1|24$, $2|14$, $4|12$, $1|34$, $3|14$, $4|13$, $2|34$, $3|24$, $4|23$.

The next proposition shows that maximal compatible families of partial splits on $\{1, 2, 3, 4\}$ are classified into six types. We illustrates this six types and their tree representations in Figure 2, where the line corresponding to a partial split $\{A, B\}$ separates points of A and B and meets points of $\{1, 2, 3, 4\} \setminus A \cup B$.

Two families of partial splits $\mathcal{S}_1$ and $\mathcal{S}_2$ on X are said to be *isomorphic* if there exists some bijection $\sigma : X \to X$ such that $\mathcal{S}_2 = \{\{\sigma(A), \sigma(B)\} \mid \{A, B\} \in \mathcal{S}_1\}$.

Proposition 3.2. *Any maximal compatible family of partial splits on* $\{1, 2, 3, 4\}$ *is isomorphic to one of the following:*

Type 1: $\{1|234, 2|134, 12|34, 3|124, 4|123\}$,

Type 2: $\{1|234, 2|134, 12|34, 12|4, 4|123\}$,

Type 3: $\{1|234, 1|34, 12|34, 12|4, 4|123\}$,

Type 4: $\{1|234, 1|34, 1|4, 13|4, 4|123\}$,

Type 5: $\{1|234, 1|34, 1|4, 12|4, 4|123\}$,

Type 6: $\{1|23, 2|13, 3|12, 1|234, 2|134, 3|124\}$.

Proof. For a family of partial splits $\mathcal{S}'$, the *incompatibility graph* of $\mathcal{S}'$ is defined to be a graph whose vertex set is $\mathcal{S}'$ and edge set is

$$\{ST \mid \ S \in \mathcal{S}' \text{ and } T \in \mathcal{S}' \text{ are not compatible}\}. \tag{3.27}$$

Then ${\mathcal{S}'}_0 \subseteq \mathcal{S}'$ is compatible if and only if ${\mathcal{S}'}_0$ is a stable set of the incompatibility graph of $\mathcal{S}'$.

Let $\mathcal{S}$ be a maximal compatible family of partial splits on $\{1, 2, 3, 4\}$. Suppose that $\mathcal{S}$ has a partial split of (S2), say $12|34$. The set of all partial splits compatible to $12|34$ is given by

$$\mathcal{S}_1 = \{12|34, 1|234, 2|134, 3|124, 4|123, 1|34, 2|14, 12|4, 12|3\}. \tag{3.28}$$

Then the incompatibility graph of $\mathcal{S}_1$ is (a) of Figure 3. From maximal stable sets of this graph, we see that $\mathcal{S}$ is of Type 1, Type 2, or Type 3.

Suppose that $\mathcal{S}$ has a partial split of (S3), say $1|2$. The set of all partial splits compatible to $1|2$ is given by

$$\mathcal{S}_2 = \{1|2, 1|234, 2|134, 1|24, 1|23, 2|34, 2|13\}. \tag{3.29}$$

Then the incompatibility graph of $\mathcal{S}_2$ is (b) of Figure 3. From maximal stable sets of this graph, we see that $\mathcal{S}$ is of Type 4 or Type 5.

Suppose that $\mathcal{S}$ has no partial splits of (S2) and (S3). If $\mathcal{S}$ consists of partial splits of (S1), $\mathcal{S}$ is not maximal compatible. Suppose that $\mathcal{S}$ has a partial split of (S4), say $1|23$. The set of all partial splits of (S1) and (S4) compatible to $1|23$ is given by

$$\mathcal{S}_3 = \{1|23, 2|13, 3|12, 1|234, 2|134, 3|124, 2|14, 3|14\}. \tag{3.30}$$

Then the incompatibility graph of $\mathcal{S}_3$ is (c) of Figure 3. Hence all maximal stable sets of this graph are

(1) $\{1|23, 2|13, 3|12, 1|234, 2|134, 3|124\}$,

(2) $\{1|23, 2|14, 1|234, 2|134\}$, and

(3) $\{1|23, 3|14, 1|234, 3|124\}$.

Neither (2) nor (3) is maximal compatible. Hence $\mathcal{S}$ must be (1) and is of Type 6.

□

Finally, we can confirm the condition (1.2) for each type in Proposition 3.2 as follows:

(Type 1, 2, 3) $\max\{d^{(12)(34)}, d^{(13)(24)}, d^{(14)(23)}\}$ is attained at least twice,

(Type 4, 5) $\max\{d^{(12)(34)}, d^{(13)(24)}, d^{(14)(23)}\} = d^{(14)}$,

(Type 6) $\max\{d^{(12)(34)}, d^{(13)(24)}, d^{(14)(23)}\} \leq d^{(123)}$,

where we use the notation in (3.21) and the labelling corresponds to Figure 2.

Acknowledgement

The author thanks Satoru Fujishige, Kazuo Murota, and Akihisa Tamura for helpful comments, and Bernd Sturmfels for suggesting a connection between the present work and tropical geometry, and anonymous referees for suggestive comments.

A Appendix

Proof of Theorem 2.1

Our proof is based on the fundamental duality principle in the theory of linear programming; see [22] for example for linear programming.

Lemma A.1. *Let $A = (a_1\ a_2\ \dots\ a_m)$ be an $n \times m$ matrix with n-dimensional column vectors $\{a_i \mid i = 1, 2, \dots, m\} \subseteq \mathbf{R}^n$. For $b \in \mathbf{R}^n$, consider the polyhedron*

$$Q = \{u \in \mathbf{R}^m \mid Au = b, u \geq 0\}. \tag{A.1}$$

Then $u \in Q$ is a vertex of Q if and only if the vectors $\{a_i \mid u_i > 0\}$ are linearly independent.

Let E_X denote the set of unordered pairs defined as

$$E_X = \{xy \mid x \in X, y \in X\}. \tag{A.2}$$

The following is an easy consequence of the previous lemma.

Lemma A.2. *Let $Q(X)$ be the set of nonnegative weights on E_X such that the sum of the weights around each vertex is equal to 2, i.e.,*

$$Q(X) = \{\lambda \in \mathbf{R}^{E_X} \mid \sum_{xy \in E_X} (\chi_x + \chi_y)\lambda_{xy} = 2\chi_X, \lambda_{xy} \geq 0\ (xy \in E_X)\}. \tag{A.3}$$

Then $\lambda \in Q(X)$ is a vertex of $Q(X)$ if and only if there exists some edge cover E of (X, E_X) consisting of a matching and odd cycles, pairwise vertex disjoint, such that

$$\lambda_{xy} = \begin{cases} 2 & \text{if } xy \text{ is an edge of matching of } E, \\ 1 & \text{if } xy \text{ is an edge of some odd cycle of } E, \\ 0 & \text{otherwise,} \end{cases} \qquad (xy \in E_X). \tag{A.4}$$

Considering the facts that a permutation of X can be decomposed as disjoint cyclic permutations, that a cyclic permutation can be regarded as a cycle of graph (X, E_X) and that an even cycle is the union two edge-disjoint matchings, the optimal value of the linear program

$$\max. \sum_{xy \in E_X} \lambda_{xy} d(x,y) \quad \text{s.t.} \quad \lambda \in Q(X) \tag{A.5}$$

is given by

$$\max\{\sum_{x \in X} d(x, \sigma(x)) \mid \sigma \text{ is a permutation of } X\}. \tag{A.6}$$

Hence, the condition (b) of Theorem 2.1 can be rephrased as follows:

(b') There exist a $2n$-element subset $Y \subseteq X$ and a perfect matching M of (Y, E_Y) such that $2\chi_M \in \mathbf{R}^{E_Y}$ is the unique optimal solution to the linear program

$$\max. \sum_{xy \in E_Y} \lambda_{xy} d(x,y) \quad \text{s.t.} \quad \lambda \in Q(Y). \tag{A.7}$$

In the following, we often use the dimension formula (3.12).

Lemma A.3. *The following holds, where $d^Y : Y \times Y \to \mathbf{R}$ denotes the restriction of d to Y.*

(1) $\dim T(Y, d^Y) \leq \dim T(X, d)$ *for* $Y \subseteq X$.

(2) *If* $\dim T(X, d) \geq n$, *there exists* $Y \subseteq X$ *with* $|Y| = 2n$ *such that* $\dim T(Y, d^Y) = n$.

Proof. For $f \in \mathbf{R}^X$ and $Y \subseteq X$, let $f^Y : Y \to \mathbf{R}$ denote the restriction of f to Y.

(1). It is sufficient to show the case $Y = X \setminus \{z\}$ for some $z \in X$. Suppose that $\dim T(Y, d^Y) = n$. Then there exists $f \in T(Y, d^Y)$ such that a graph $(Y, E(f^Y))$ has n bipartite components $(A_1 \cup B_1, E_1), \ldots, (A_n \cup B_n, E_n)$ with $A_i \cap B_i = \emptyset$ and $E_i \subseteq \{xy \mid x \in A_i, y \in B_i\}$ for $i = 1, \ldots, n$. We use the notation and the method in Subsection 3.2. Let $f' \in \mathbf{R}^X$ be defined as

$$f'(x) = \begin{cases} \max\{0, \max_{y \in Y}(d(z,y) - f(y))\} & \text{if } x = z, \\ f(x) & \text{otherwise.} \end{cases} \tag{A.8}$$

Then some edges connecting z appear in $(X, E(f'))$ and we have $f' \in T(X, d)$. If $(X, E(f'))$ has no edges connecting $\{z\}$ and $A_1 \cup B_1 \cup \cdots \cup A_n \cup B_n$, then $(X, E(f'))$ also has n bipartite components.

We suppose that there exists $y \in A_1$ with $zy \in E(f')$. Let S and S' be stable sets of $(X, E(f'))$ defined as $S = A_1 \cup A_2 \cup \cdots \cup A_n$ and $S' = A_1 \cup B_2 \cup \cdots \cup B_n$. Let $g \in \mathbf{R}^X$ be defined as

$$g = f' + \epsilon(\chi_{N(S)} - \chi_S) + \epsilon'(\chi_{N(S')} - \chi_{S'}) \tag{A.9}$$

for sufficiently small $\epsilon, \epsilon' > 0$. Then we have $g \in T(X, d)$. Furthermore all edges in $(X, E(f'))$ connecting $\{z\}$ and $X \setminus A_1$ vanish in $(X, E(g))$. This implies that $(X, E(g))$ has n bipartite components.

(2). Since $\dim T(X, d) \geq n$, there exists $f \in T(X, d)$ such that $(X, E(f))$ has n bipartite components. Take n edges from each bipartite component, say $\{x_1y_1, x_2y_2, \ldots, x_ny_n\}$ and put $Y = \{x_1, x_2, \ldots, x_n, y_1, \ldots, y_n\}$. Then it is easy to check that f^Y is in $T(Y, d^Y)$ and $(Y, E(f^Y))$ has n bipartite components. □

Hence, it is sufficient to show the following.

Theorem A.4. *Suppose $|X| = 2n$. The following conditions are equivalent.*

(a) $\dim T(X, d) = n$.

(b) *There exists some perfect matching M of (X, E_X) such that $\lambda^* = 2\chi_M \in \mathbf{R}^{E_X}$ is the unique optimal solution to linear program (A.5).*

Proof. (a) $\Rightarrow$ (b). There exists $f^* \in P(X, d)$ such that $K(f^*)$ has n bipartite components. Hence $E(f^*)$ must be a perfect matching of (X, E_X). Consider the dual program of (A.5):

$$\min . \sum_{x \in X} f(x) \quad \text{s.t.} \quad f \in P(X, d). \tag{A.10}$$

Then $\lambda^* = 2\chi_{E(f^*)}$ and f^* satisfies the (strict) complementary slackness condition

$$\lambda^*_{xy} > 0 \Leftrightarrow f^*(x) + f^*(y) = d(x, y) \quad (xy \in E_X). \tag{A.11}$$

Hence λ^* and f^* are optimal solutions to (A.5) and (A.10), respectively. Conversely, any optimal solution $\tilde{\lambda}$ of (A.5) satisfies

$$\tilde{\lambda}_{xy} = 0 \quad (xy \notin E(f^*)). \tag{A.12}$$

Since $\{\chi_x + \chi_y \mid xy \in E(f^*)\}$ is linearly independent, we have $\tilde{\lambda} = \lambda^*$. Hence λ^* is the unique optimal solution of linear program (A.5).

(b) $\Rightarrow$ (a). By the strict complementary slackness theorem, there exist optimal solutions $\tilde{\lambda}$ and f^* of (A.5) and (A.10) such that

$$\tilde{\lambda}_{xy} > 0 \Leftrightarrow f^*(x) + f^*(y) = d(x, y) \quad (xy \in E_X). \tag{A.13}$$

By the condition (b), we have $\tilde{\lambda} = \lambda^*$. Hence it must be that $E(f^*) = M$. This implies $\dim T(X, d) = n$. □

Contractibility of the Union of Bounded Faces of a Polyhedron

Lemma A.5. *The union of bounded faces of a pointed polyhedron is contractible.*

Proof. Let $P \subseteq \mathbf{R}^n$ be a pointed polyhedron and B the union of bounded faces of P. We construct a continuous map (retraction) $r : P \to B$ satisfying $r(x) = x$ for $x \in B$. If such a retraction exists, r is homotopic to the identity map by a homotopy $h : P \times [0, 1] \to P$ defined as $h(x, t) = tr(x) + (1 - t)x$. Hence, B is homotopic to P which is contractible by convexity.

We may assume that P is represented as

$$P = \{x \in \mathbf{R}^n \mid \langle a_j, x\rangle \leq b_j \ (j = 1, \dots, m)\} \tag{A.14}$$

for some $\{(a_1, b_1), (a_2, b_2), \dots, (a_m, b_m)\} \subseteq \mathbf{R}^{n+1}$, where $\langle \cdot, \cdot \rangle$ denotes the standard inner product of $\mathbf{R}^n$. Let $\{u_1, u_2, \dots, u_k\} \subseteq \mathbf{R}^n$ be the set of extreme rays of a pointed cone $\{u \in \mathbf{R}^n \mid \langle a_j, u\rangle \leq 0 \ (j = 1, \dots, m)\}$. Then, a face $F \subseteq P$ is bounded if and only if F does not contain each ray u_i $(i = 1, \dots, k)$, where we say "F *contains* a ray u_i" if it satisfies $F + tu_i \subseteq F$ for $t \geq 0$. For a ray u_i, we define a map $\phi_{u_i} : P \to P$ as

$$\begin{aligned} \phi_{u_i}(x) &:= x - u_i \sup\{t \in \mathbf{R} \mid x - tu_i \in P\} \\ &= x - u_i \sup\{t \in \mathbf{R} \mid \langle a_j, x - tu_i \rangle \leq b_j \ (j = 1, \ldots m)\} \\ &= x - u_i \inf_{j : \langle a_j, u_i \rangle < 0} \{(\langle a_j, x \rangle - b_j)/\langle a_j, u_i \rangle\} \quad (x \in P). \end{aligned} \quad \text{(A.15)}$$

Since P is pointed, the infimum of (A.15) is attained. In particular, ϕ_{u_i} is continuous. Furthermore, ϕ_{u_i} is a retraction from P to the union of faces which do not contain the ray u_i. Indeed, this immediately follows from the fact that for $x \in P$, the unique minimal face F containing x does not contain the ray u_i if and only if there exists $j \in \{1, \ldots, m\}$ with $\langle a_j, u_i \rangle < 0$ such that $\langle a_j, x \rangle = b$. Furthermore, $\phi_{u_i}(F) \subseteq F$ holds for any face F since $\langle a_j, x \rangle = b_j$ implies $\langle a_j, \phi_{u_i}(x) \rangle = b_j$. Hence, we obtain a desired retraction $r : P \to B$ defined as

$$r = \phi_{u_k} \circ \phi_{u_{k-1}} \circ \cdots \circ \phi_{u_1}. \quad \text{(A.16)}$$

□

References

[1] H.-J. Bandelt and A. W. M. Dress, A canonical decomposition theory for metrics on a finite set, Adv. Math. 92 (1992), 47–105.

[2] J.-P. Barthélémy and A. Guénoche, Trees and Proximity Representations, Translated from the French by Gregor Lawden, Wiley, Chichester, 1991.

[3] P. Buneman, The recovery of trees from measures of dissimilarity, in: Mathematics in the Archaeological and Historical Sciences, F. R. Hodson et al., Eds., Edinburgh University Press, Edinburgh, 1971, pp. 387–395.

[4] P. Buneman, A note on metric properties of trees, J. Comb. Theory Ser. B 17 (1974) 48–50.

[5] M. Chrobak and L. L. Larmore, Generosity helps or an 11-competitive algorithm for three servers, J. Algorithms 16 (1994), 234–263.

[6] 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 (1984), 321–402.

[7] A. W. M. Dress, Towards a classification of transitive group actions on finite metric spaces, Adv. Math. 74 (1989), 163–189.

[8] A. Dress, V. Moulton and W. Terhalle, T-theory: an overview, Eur. J. Comb. 17 (1996), 161–175.

[9] J. Edmonds, Submodular functions, matroids, and certain polyhedra, in: Combinatorial Structures and their Applications, R. Guy et al., Eds, Gordon and Breach, New York, 1970, pp. 69–87.

[10] I. M. Gelfand, R. M. Goresky, R. D. MacPherson, and V. V. Serganova, Combinatorial geometries, convex polyhedra, and Schubert cells, Adv. in Math. 63 (1987), 301–316.

[11] S. Fujishige, *Submodular Functions and Optimization*, Annals of Discrete Mathematics 47, North-Holland, Amsterdam, 1991, 2nd ed. (to appear).

[12] H. Hirai, Geometric study on the split decomposition of finite metrics, RIMS preprint 1459, Kyoto University, May 2004.

[13] H. Hirai and K. Murota, M-convex functions and tree metrics, Japan J. Indust. Appl. Math. 21 (2004), 391–403.

[14] J. R. Isbell, Six theorems about injective metric spaces, Comment. Math. Helv. 39 (1964), 65–76.

[15] L. Lafforgue, Chirurgie des grassmanniennes, CRM Monograph Series, 19. AMS, Providence, RI, 2003.

[16] K. Murota, *Discrete Convex Analysis*, SIAM, Philadelphia, PA, 2003.

[17] C. Semple and M. Steel, Phylogenetics, Oxford University Press, Oxford, 2003.

[18] J. M. S. Simões-Pereira, A note on the tree realizability of a distance matrix, J. Comb. Theory 6 (1969), 303–310.

[19] D. Speyer, Tropical Linear Spaces, `arXiv:math.CO/0410455`.

[20] B. Sturmfels and J. Yu, Classification of six-point metrics, Electron. J. Combin. 11 (2004).

[21] N. Tomizawa, Theory of hyperspaces (XVI) — On the structure of hedrons (in Japanese), Papers of the Technical Group on Circuit and System Theory, Institute of Electronics and Communication Engineer of Japan, CAS82–174, 1983.

[22] R. J. Vanderbei, Linear programming, Second edition, Kluwer, Boston, 2001.

[23] K. A. Zareckiĭ, Constructing a tree on the basis of a set of distances between the hanging vertices, Usp. Mat. Nauk 20 (1965), 90–92.

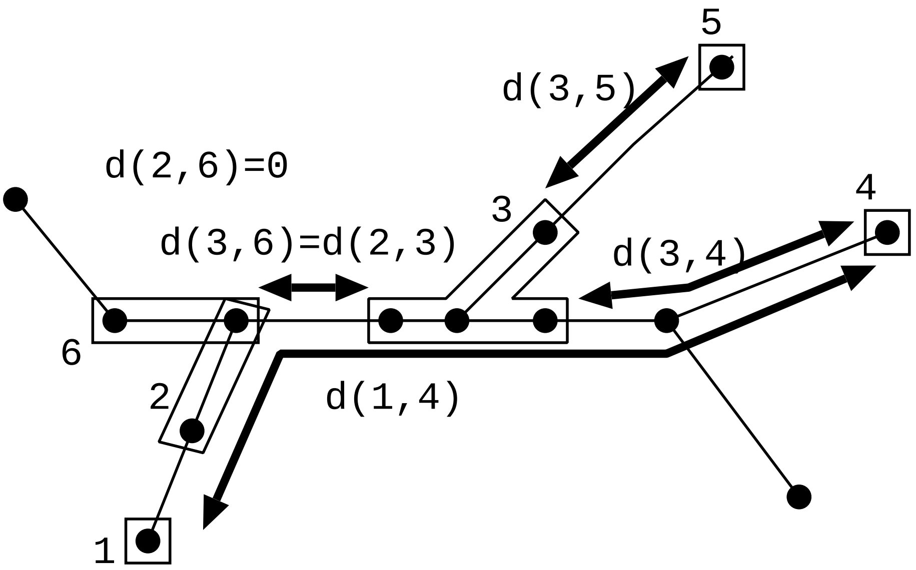

Figure 1: Shortest path lengths between six subtrees of a tree

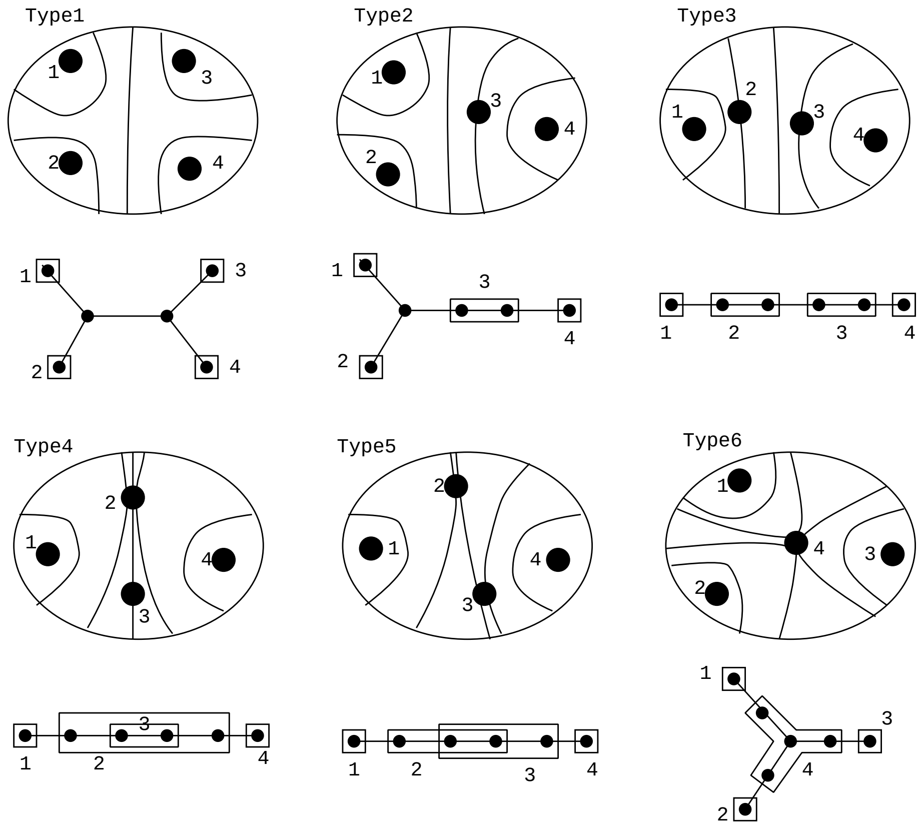

Figure 2: All types of maximal compatible families of $\{1, 2, 3, 4\}$ and their tree representations

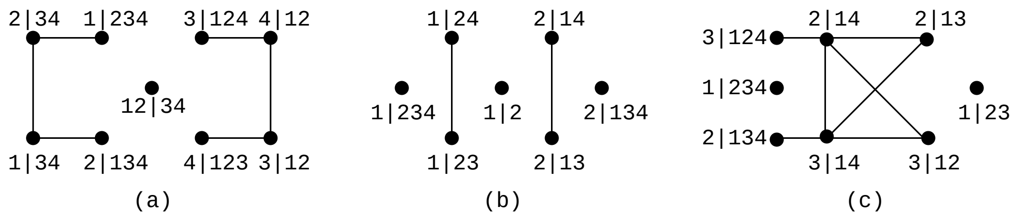

Figure 3: Incompatibility graphs