Discrete Applied Mathematics 369 (2025) 28–44

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Buneman graphs, partial splits and subtree distances

David Bryant [a], Katharina T. Huber [b], Vincent Moulton [b], Andreas Spillner [c,*]

[a] *Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand*
[b] *School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK*
[c] *Merseburg University of Applied Sciences, 06217 Merseburg, Germany*

ARTICLE INFO

Article history:
Received 3 November 2023
Received in revised form 31 October 2024
Accepted 20 February 2025
Available online xxxx

Dedicated to the memory of
Andreas W.M. Dress

Keywords:
Partial split
Network representation
Subtree distance

ABSTRACT

In phylogenetics and other areas of classification, the *Buneman graph* is commonly used to represent a collection of bipartitions or *splits* of a (finite) set X in order to display evolutionary relationships. The set X usually corresponds to a set of taxa (or species), and the splits are usually derived from molecular sequence data associated to the taxa. One issue with this approach is that missing molecular data can lead to bipartitions of subsets of X or *partial splits*, instead of splits of the full set X. In this paper, we show that the definition of the Buneman graph can be naturally extended to collections of partial splits of a set X. Just as with splits, we show that the graph so obtained is an X-labeled median graph but, in contrast to the usual Buneman graph, the elements in X are represented by convex subsets of the vertex set of the graph instead of single vertices. We also show that the Buneman graph for a collection of partial splits is closely related to subtree distances. In particular, for a collection $\mathcal{S}$ of weighted partial splits that satisfies a certain pairwise compatibility condition, we show that the corresponding edge-weighted Buneman graph is the unique minimal tree that represents the subtree distance d corresponding to $\mathcal{S}$. Moreover, we show that in this special situation the Buneman graph can also be considered as a type of configuration space for the set of all tree-metrics that minimally extend the subtree distance d.

1. Introduction

In phylogenetics, evolutionary relationships can be represented by a certain type of undirected graph that was introduced by Barthélemy [7] under the name of a *median graph with latent vertices*. This graph is also commonly known as the *Buneman graph* [13] since, as mentioned in [7,13], one way in which it can be constructed is as an extension of a method for reconstructing phylogenetic trees presented in [10]. One of the main applications of Buneman graphs is to analyze mitochondrial DNA data [6], in which setting they are also known as *median networks*. This is because they are examples of *median graphs*, an important class of graphs that has applications in areas such as lattices, concurrency theory and cubical complexes (see e.g. [3] and [17, Chapter 12] for reviews of median graphs and some of their applications). The key difference between Buneman graphs and median graphs is that a subset of their vertices is labeled by some underlying (finite) set X (corresponding, for example, to a collection of species or individuals), which allows the interrelationships between the elements in X to be displayed graphically.

One way in which a Buneman graph can be constructed is as follows: Starting with a set of species X, a collection $\mathcal{S}$ of bipartitions or *splits* $S = \{A, B\}$ of X (i.e. $A \cup B = X$ and $A \cap B = \emptyset$) is derived, for example, from the columns of an

* Corresponding author.
E-mail addresses: david.bryant@otago.ac.nz (D. Bryant), k.huber@uea.ac.uk (K.T. Huber), v.moulton@uea.ac.uk (V. Moulton), andreas.spillner@hs-merseburg.de (A. Spillner).

https://doi.org/10.1016/j.dam.2025.02.035
0166-218X/

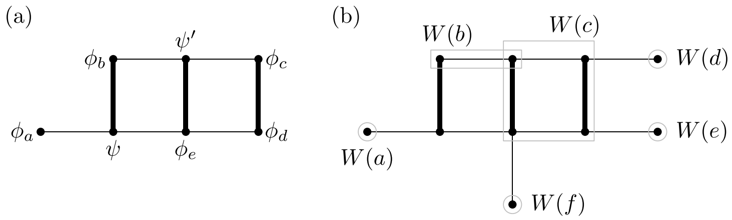

Fig. 1. (a) The Buneman graph $B(\mathcal{S})$ for the collection $\mathcal{S}$ of splits $S_1 = \{\{a\}, \{b, c, d, e\}\}$ $S_2 = \{\{a, b\}, \{c, d, e\}\}$, $S_3 = \{\{a, b, e\}, \{c, d\}\}$ and $S_4 = \{\{a, d, e\}, \{b, c\}\}$ of the set $X = \{a, b, c, d, e\}$. The edges in $E(S_4)$ are drawn bold. We have $\psi(S_1) = \psi'(S_1) = \{b, c, d, e\}$, $\psi(S_2) = \{a, b\} \neq \{c, d, e\} = \psi'(S_2)$, $\psi(S_3) = \psi'(S_3) = \{a, b, e\}$ and $\psi(S_4) = \{a, d, e\} \neq \{b, c\} = \psi'(S_4)$, implying that $\{\psi, \psi'\}$ is not an edge. In contrast, $\psi'(S_i) \neq \phi_e(S_i)$ precisely for $i = 4$, implying that $\{\psi', \phi_e\}$ is an edge. (b) The Buneman graph $B(\mathcal{S})$ for the collection $\mathcal{S}$ of partial splits $\{\{x\}, X - \{x\}\}$ for $x \in \{a, d, e, f\}$, $\{\{a\}, \{c, d, e, f\}\}$, $\{\{a, b, f\}, \{d, e\}\}$ and $S = \{\{a, e, f\}, \{b, d\}\}$ of $X = \{a, b, c, d, e, f\}$. For each $x \in X$ the convex set $W(x)$ is enclosed by a gray line. The bold edges represent the partial split S.

alignment of molecular sequences sampled from the species in X [6]. Let $\mathcal{P}(X)$ denote the power set of X. The Buneman graph $B(\mathcal{S})$ associated to $\mathcal{S}$ is defined to be the graph whose vertex set consists of those maps $\phi : \mathcal{S} \to \mathcal{P}(X)$ that satisfy

(i) $\phi(S) \in S$ for all $S \in \mathcal{S}$, and
(ii) $\phi(S) \cap \phi(S') \neq \emptyset$ for all $S, S' \in \mathcal{S}$,

and whose edge set consists of those pairs $\{\phi, \phi'\}$ for which there is precisely one split S in $\mathcal{S}$ with $\phi(S) \neq \phi'(S)$. Intuitively, Condition (i) implies that $B(\mathcal{S})$ is a subgraph of a hypercube and Condition (ii) tells us which vertices of the hypercube are actually used. For example, in Fig. 1(a) we depict a collection of splits and its Buneman graph. Note that the set X is embedded in $B(\mathcal{S})$, by associating to each $x \in X$ the map $\phi_x : \mathcal{S} \to \mathcal{P}(X)$ which is defined by putting $\phi_x(S)$ to be the set in S that contains x. We clearly have $x \in \phi_x(S) \cap \phi_x(S')$ for all $S, S' \in \mathcal{S}$, implying (ii). But not all maps $\phi : \mathcal{S} \to \mathcal{P}(X)$ satisfy (ii). For the collection of splits in Fig. 1(a), for example, we obtain such a map ϕ by putting $\phi(S_1) = \{a\}$, $\phi(S_2) = \{a, b\}$, $\phi(S_3) = \{c, d\}$ and $\phi(S_4) = \{b, c\}$, for which $\phi(S_2) \cap \phi(S_3) = \emptyset$.

The Buneman graph $B(\mathcal{S})$ enjoys several nice properties (for reviews, see e.g. [14, Chapter 4] and [30, Section 3.8]). For example, it is a tree (or, more specifically, an *X-tree*) precisely if every pair of splits in $\mathcal{S}$ satisfies a certain compatibility condition [10] (see also Section 6). Moreover, $B(\mathcal{S})$ represents the underlying collection $\mathcal{S}$ of splits as follows: For each split $S = \{A, B\} \in \mathcal{S}$, consider the set $E(S)$ of those edges $\{\phi, \phi'\}$ of $B(\mathcal{S})$ with $\phi(S) \neq \phi'(S)$. Then the removal of $E(S)$ from $B(\mathcal{S})$ gives rise to two connected components, one containing the vertices ϕ_x for $x \in A$ and the other containing the vertices ϕ_x for $x \in B$ (see Fig. 1(a)).

One problem with using this approach for real data sets is that molecular sequence data sometimes gives rise to *partial splits* $S = \{A, B\}$ of X, that is, we still have $A \cap B = \emptyset$ but need not have $A \cup B = X$ (for clarity, if we require $A \cup B = X$ from now on we will say that S is a *full split* of X) (see e.g. [21,29]). Thus it is of interest to find an analogue of the construction of the usual Buneman graph outlined above that can be used for collections of partial splits. In this paper, we show that for any collection $\mathcal{S}$ of partial splits we can still define the Buneman graph $B(\mathcal{S})$ and that it has similar properties as the usual Buneman graph (e.g. it is a median graph). In particular our definition of $B(\mathcal{S})$ yields the usual Buneman graph if all splits in $\mathcal{S}$ are full splits of X.

In Fig. 1(b) we give an example of the Buneman graph $B(\mathcal{S})$ for a collection $\mathcal{S}$ of partial splits. Interestingly, as we shall see, in the Buneman graph $B(\mathcal{S})$ the elements $x \in X$ are no longer necessarily represented by single vertices but instead by a non-empty *convex* set $W(x)$ of vertices (convexity is formally defined in Section 3). Moreover, each partial split $S = \{A, B\} \in \mathcal{S}$ is still represented by a set $E(S)$ of edges, the difference being that the removal of $E(S)$ from $B(\mathcal{S})$ now gives rise to two connected components, one containing the sets $W(x)$ for $x \in A$, the other containing the sets $W(x)$ for $x \in B$ and all sets $W(x)$ for $x \in X - (A \cup B)$ having a non-empty intersection with both connected components. This is illustrated in Fig. 1(b), which also shows that we may have $W(x) \cap W(x') \neq \emptyset$ for two distinct $x, x' \in X$ and that there may exist vertices in $B(\mathcal{S})$ that are not contained in the set $W(x)$ for any $x \in X$. We note that our construction of $B(\mathcal{S})$ was motivated by results of Hirai in [18] concerning *subtree distances*, and, as we shall see, there are several close connections with Hirai's results.

The rest of this paper is structured as follows. In Section 2, we formally define the Buneman graph $B(\mathcal{S})$ for a collection $\mathcal{S}$ of partial splits and establish some of its basic properties. In Section 3, we present some useful results concerning certain convex sets of vertices in $B(\mathcal{S})$. Then, after defining the convex sets $W(x)$ mentioned above, we employ these results to show that for a collection $\mathcal{S}$ of partial splits on X, the distance on X given by these convex sets relative to $B(\mathcal{S})$ is precisely the distance on X that is naturally induced by $\mathcal{S}$ (Theorem 4.5). In Section 5, we shed some light on how the combinatorial properties of the partial splits in $\mathcal{S}$ influence the size and the structure of $B(\mathcal{S})$. In Section 6, we give a characterization for when the Buneman graph $B(\mathcal{S})$ is a tree (Theorem 6.3). In Section 7, we present a relationship between Buneman graphs and subtree distances (Theorem 7.1). In Section 8, in case the Buneman graph for a given collection $\mathcal{S}$ of (weighted) partial splits on X is a tree, we show that it can be considered as a type of configuration space of all those X-trees whose underlying collection of full splits is a certain type of extension of $\mathcal{S}$ (Corollary 8.4). We conclude in Section 9 with mentioning some potential directions for future work.

2. Basic properties of the Buneman graph for partial splits

In this section, we shall show that, as with collections of full splits, the Buneman graph for a collection of partial splits is a median graph. To do this, we begin by presenting some basic terminology for partial splits and for distances corresponding to collections of partial splits.

Let X be a finite, non-empty set. As informally introduced in the introduction, a *partial split* of X is an unordered pair $\{A, B\}$ with $A, B \subseteq X$, $A, B \neq \emptyset$ and $A \cap B = \emptyset$. We denote such an unordered pair by $A|B$. For a partial split $S = A|B$ of X we define $S^c = X - (A \cup B)$, $A_S^c = B$ and $B_S^c = A$. If S is clear from the context we omit the subscript in the latter two expressions and just write A^c and B^c, respectively. A *split system*[1] on X is a non-empty set $\mathcal{S}$ of partial splits of X.

Example 2.1. The partial splits

$$S_i = \{x_i\}|X - \{x_i\} \text{ for } 1 \leq i \leq 5 \qquad S_7 = \{x_1, x_2, x_5, x_6\}|\{x_3, x_4\}$$
$$S_6 = \{x_1, x_2\}|\{x_3, x_4, x_5, x_7\} \qquad S_8 = \{x_1, x_2, x_3, x_6\}|\{x_4, x_5\}$$

of $X = \{x_1, \ldots, x_7\}$ form the split system $\mathcal{S} = \{S_1, \ldots, S_8\}$ on X.

A *distance* on X is a map $d : X \times X \to \mathbb{R}_{\geq 0}$ with $d(x, y) = d(y, x)$ and $d(x, x) = 0$ for all $x, y \in X$. A distance d on X is a *metric* if, in addition, $d(x, z) \leq d(x, y) + d(y, z)$ holds for all $x, y, z \in X$. We write $d_1 \leq d_2$ for distances d_1 and d_2 on X if $d_1(x, y) \leq d_2(x, y)$ for all $x, y \in X$. For a partial split $S = A|B$ of X, we define the distance $d_S : X \times X \to \{0, 1\}$ on X by putting

$$d_S(x, y) = |A \cap \{x, y\}| \cdot |B \cap \{x, y\}|$$

for all $x, y \in X$. In addition, for a split system $\mathcal{S}$ on X, we define the distance $d_{\mathcal{S}} : X \times X \to \mathbb{N}$ by putting $d_{\mathcal{S}} = \sum_{S \in \mathcal{S}} d_S$.

Example 2.2. For the split system $\mathcal{S}$ on $X = \{x_1, \ldots, x_7\}$ in Example 2.1 we obtain $d_{S_6}(x_1, x_7) = 1$, $d_{S_7}(x_1, x_7) = 0$ and $d_{\mathcal{S}}(x_1, x_7) = 2$.

The following concepts were introduced in [18, p. 113]. Two distinct partial splits S_1 and S_2 of X are *compatible* if there is some $A_1 \in S_1$ and some $A_2 \in S_2$ such that $A_1 \subseteq A_2^c$ and $A_2 \subseteq A_1^c$. Note that if such sets A_1 and A_2 exist then they are unique and we put $N(S_1, S_2) = N(S_2, S_1) = \{A_1, A_2\}$. Otherwise we say that S_1 and S_2 are *incompatible* and put $N(S_1, S_2) = N(S_2, S_1) = \emptyset$.

For a finite set M, let $\mathcal{P}(M)$ denote the power set of M. Given a split system $\mathcal{S}$ on X, a map $\phi : \mathcal{S} \to \mathcal{P}(X)$ is an *$\mathcal{S}$-map* if $\phi(S) \in S$ for all $S \in \mathcal{S}$. We let $V^*(\mathcal{S})$ denote the set of $\mathcal{S}$-maps and for $\phi, \phi' \in V^*(\mathcal{S})$ we put

$$\Delta(\phi, \phi') = \{S \in \mathcal{S} \,:\, \phi(S) \neq \phi'(S)\}.$$

The graph $B^*(\mathcal{S})$ with vertex set $V^*(\mathcal{S})$ and edge set consisting of those pairs $\{\phi, \phi'\}$ of $\mathcal{S}$-maps with $|\Delta(\phi, \phi')| = 1$ is isomorphic to a hypercube of dimension $|\mathcal{S}|$. We define the *Buneman graph $B(\mathcal{S})$ (of $\mathcal{S}$)* to be the subgraph of $B^*(\mathcal{S})$ induced on the set $V(\mathcal{S}) \subseteq V^*(\mathcal{S})$ consisting of those $\mathcal{S}$-maps ϕ that satisfy the following condition:

(BG) For all $S, S' \in \mathcal{S}$ with $S \neq S'$, we have $\{\phi(S), \phi(S')\} \neq N(S, S')$.

In the remainder of the paper, slightly abusing notation, we just use $B(\mathcal{S})$ again to denote the Buneman graph for any split system $\mathcal{S}$.

Example 2.3. We consider again the split system $\mathcal{S}$ on $X = \{x_1, \ldots, x_7\}$ in Example 2.1. It can be checked that the partial splits S_i and S_j, $1 \leq i < j \leq 8$, are compatible, except for S_7 and S_8 which are incompatible. The set $V(\mathcal{S})$ consists of the $\mathcal{S}$-maps given in the appendix. To illustrate Condition (BG), consider the splits S_6 and S_7, for which we have $N(S_6, S_7) = \{\{x_1, x_2\}, \{x_3, x_4\}\}$. This excludes all $\mathcal{S}$-maps ϕ with $\phi(S_6) = \{x_1, x_2\}$ and $\phi(S_7) = \{x_3, x_4\}$ from $V(\mathcal{S})$. As can be seen for the $\mathcal{S}$-map ϕ_3 contained in $V(\mathcal{S})$, we have $\{\phi_3(S_6), \phi_3(S_7)\} = \{\{x_1, x_2\}, \{x_1, x_2, x_5, x_6\}\} \neq N(S_6, S_7)$. The resulting Buneman graph $B(\mathcal{S})$ is shown in Fig. 2. We will return to this example in Example 4.1.

Condition (BG) plays a role similar to condition (ii) stated in the introduction for full splits in that it filters vertices from the hypercube $B^*(\mathcal{S})$ to obtain a connected graph which represents the partial splits as described in Fig. 1(b). The next lemma makes this more precise.

Lemma 2.4. *Let $\mathcal{S}$ be a split system consisting of full splits of X, and let ϕ be an $\mathcal{S}$-map. Then Condition (ii) given in the introduction is equivalent to Condition (BG).*

[1] We caution the reader that in the literature the term split system usually refers to a collection of *full* splits. Here, the splits in a split system may (but need not) be full. If we require the splits in a split system to be full we will always explicitly say so.

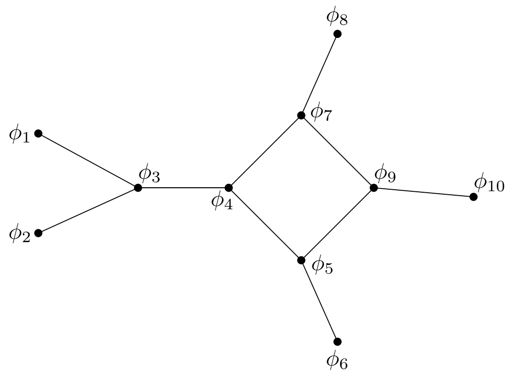

Fig. 2. The Buneman graph $B(\mathcal{S})$ with the vertex set $V(\mathcal{S}) = \{\phi_1, \ldots, \phi_{10}\}$ given in Example 2.3.

Proof. Let $S = A|B$ and $S' = A'|B'$ be two distinct full splits of X. Then the definition of compatibility of two partial splits given above is equivalent to the condition that precisely one of the four intersections

$$A \cap A', \quad A \cap B', \quad B \cap A', \quad B \cap B' \tag{1}$$

is empty and the two sets involved in this empty intersection are those contained in $N(S, S')$. Hence, S and S' are incompatible if and only if all four intersections in (1) are non-empty.

Now assume that ϕ satisfies (BG). Let $S, S' \in \mathcal{S}$. If $S = S'$ then we clearly have $\phi(S) \cap \phi(S') \neq \emptyset$. Assume $S \neq S'$. If S and S' are incompatible then $\phi(S) \cap \phi(S') \neq \emptyset$ since all four intersections in (1) are non-empty. If S and S' are compatible then $\{\phi(S), \phi(S')\} \neq N(S, S')$ implies $\phi(S) \cap \phi(S') \neq \emptyset$. In summary, $\phi(S) \cap \phi(S') \neq \emptyset$ for all $S, S' \in \mathcal{S}$, as required.

Next assume that ϕ satisfies (i). Let $S, S' \in \mathcal{S}$ with $S \neq S'$. If S and S' are incompatible, we clearly have $\{\phi(S), \phi(S')\} \neq \emptyset = N(S, S')$. If S and S' are compatible, then $N(S, S')$ consists of two disjoint subsets of X while, by (ii), $\phi(S) \cap \phi(S') \neq \emptyset$. Hence, for all $S, S' \in \mathcal{S}$ with $S \neq S'$ we have $\{\phi(S), \phi(S')\} \neq N(S, S')$, as required. □

Note that if $S, S' \in \mathcal{S}$, $S \neq S'$, are incompatible then $N(S, S') = \emptyset \neq \{\phi(S), \phi(S')\}$ for all $\phi \in V^*(\mathcal{S})$. Thus, $N(S, S') = \{\phi(S), \phi(S')\}$ for some $\phi \in V^*(\mathcal{S})$ can only hold if the partial splits S and S' are compatible. The next lemma follows immediately from the definitions above.

Lemma 2.5. *Let $\mathcal{S}$ be a split system on X, $\phi \in V^*(\mathcal{S})$ and $S, T \in \mathcal{S}$, $S \neq T$. Then the following statements are equivalent:*

(a) S and T are compatible with $N(S, T) = \{\phi(S), \phi(T)^c\}$.
(b) $\phi(S) \subseteq \phi(T)$ and $\phi(T)^c \subseteq \phi(S)^c$.
(c) $\phi(S) \subseteq \phi(T)$ and $S^c \cap \phi(T)^c = \emptyset$.

The next technical lemma implies that $V(\mathcal{S})$ is always non-empty, and it will be used in Section 4. For a partial split $S = A|B$ of X and $x \in X$, we put $S(x) = A$ if $x \in A$, $S(x) = B$ if $x \in B$ and $S(x) = \emptyset$ otherwise.

Lemma 2.6. *Let $\mathcal{S}$ be a split system on X. Then, for all $x \in X$, there exists some $\phi \in V(\mathcal{S})$ with $x \in \phi(S) \cup S^c$ for all $S \in \mathcal{S}$.*

Proof. Consider $x \in X$ and put $x_0 = x$, $\mathcal{S}_0 = \{S \in \mathcal{S} : S(x) \neq \emptyset\}$ and $\mathcal{R}_0 = \{S \in \mathcal{S} : S(x) = \emptyset\}$. For $i \geq 1$, if $\mathcal{R}_{i-1} \neq \emptyset$, we select $x_i \in X$ such that there exists at least one $S \in \mathcal{R}_{i-1}$ with $S(x_i) \neq \emptyset$. In addition, we put $\mathcal{S}_i = \{S \in \mathcal{R}_{i-1} : S(x_i) \neq \emptyset\}$ and $\mathcal{R}_i = \{S \in \mathcal{R}_{i-1} : S(x_i) = \emptyset\}$. Let $k \in \mathbb{N}$ be such that $\mathcal{R}_k = \emptyset$. Note that k is well-defined and $\mathcal{R}_i \neq \emptyset$ for $0 \leq i < k$. Moreover, the sets $\mathcal{S}_i$, $0 \leq i \leq k$, are pairwise disjoint and $\bigcup_{i=0}^{k} \mathcal{S}_i = \mathcal{S}$, but $\mathcal{S}_0$ may be empty.

Define $\phi \in V^*(\mathcal{S})$ by putting $\phi(S) = S(x_i)$ for all $S \in \mathcal{S}_i$, $0 \leq i \leq k$. By construction, we have $x \in \phi(S) \cup S^c$ for all $S \in \mathcal{S}$. Thus, it remains to show that $\phi \in V(\mathcal{S})$. Consider two splits $S, T \in \mathcal{S}$, $S \neq T$, such that S and T are compatible. If there exists some $0 \leq i \leq k$ with $S, T \in \mathcal{S}_i$ then, by construction, $x_i \in \phi(S) \cap \phi(T)$, implying $\phi(S) \cap \phi(T) \neq \emptyset$ and, thus, $\{\phi(S), \phi(T)\} \neq N(S, T)$. Otherwise, there exist, without loss of generality, $0 \leq i < j \leq k$ with $S \in \mathcal{S}_i$ and $T \in \mathcal{S}_j$, implying that $S(x_i) = \phi(S) \not\subseteq \phi(T)$ and $S(x_i) = \phi(S) \not\subseteq \phi(T)^c$ since, by construction, $T(x_i) = \emptyset$. But then we cannot have $\{\phi(S), \phi(T)\} = N(S, T)$ since, by the definition of $N(S, T)$, this would imply $\phi(S) \subseteq \phi(T)^c$ and $\phi(T) \subseteq \phi(S)^c$. Thus, we also have $\{\phi(S), \phi(T)\} \neq N(S, T)$ in this case, implying that $\phi \in V(\mathcal{S})$, as required. □

Now we prove an analogue of [30, Lemma 3.8.2].

Lemma 2.7. *Let $\mathcal{S}$ be a split system on X, $\phi \in V(\mathcal{S})$ and $T \in \mathcal{S}$. Define a map $\phi' \in V^*(\mathcal{S})$ by setting $\phi'(S) = \phi(S)$ for all $S \neq T$ and $\phi'(T) = \phi(T)^c$. Then $\phi' \in V(\mathcal{S})$ if and only if, for all $S \in \mathcal{S}$, $\phi(S) \subseteq \phi(T)$ and $S^c \cap \phi(T)^c = \emptyset$ implies $S = T$.*

Proof. Assume that there is some $S \in \mathcal{S}, S \neq T$, such that $\phi(S) \subseteq \phi(T)$ and $S^c \cap \phi(T)^c = \emptyset$. By Lemma 2.5, this is equivalent to the existence of some $S \in \mathcal{S}, S \neq T$, such that S and T are compatible with $N(S, T) = \{\phi(S), \phi(T)^c\} = \{\phi'(S), \phi'(T)\}$. In view of (BG), this is equivalent to $\phi' \notin V(\mathcal{S})$. □

For a graph G we denote by $d_G(u, v)$ the length of a shortest path in G from vertex u to vertex v. A subgraph H of G is an *isometric subgraph* if $d_H(u, v) = d_G(u, v)$ for all vertices u and v of H. A graph G is a *median graph* if, for all vertices u, v and w of G, there exists a unique vertex $m = m_G(u, v, w)$ of G with $d_G(u, v) = d_G(u, m) + d_G(m, v)$, $d_G(u, w) = d_G(u, m) + d_G(m, w)$ and $d_G(v, w) = d_G(v, m) + d_G(m, w)$. The vertex $m_G(u, v, w)$ is called the *median* of u, v and w in G. Note that, for all split systems $\mathcal{S}$ on X, the graph $B^*(\mathcal{S})$, being isomorphic to a hypercube, is a median graph. More specifically, the median of $\phi_1, \phi_2, \phi_3 \in V^*(\mathcal{S})$ equals the vertex $\phi \in V^*(\mathcal{S})$ with $|\{i \in \{1, 2, 3\} : \phi_i(S) = \phi(S)\}| \geq 2$ for all $S \in \mathcal{S}$. We now prove the main result of this section, which can be regarded as a generalization of [13, (c), p. 330] (see also [7, Thm. 1]).

Theorem 2.8. *Let $\mathcal{S}$ be a split system on X. Then $B(\mathcal{S})$ is a median graph, embedded as an isometric subgraph of the hypercube $B^*(\mathcal{S})$.*

Proof. We first establish that $d_{B(\mathcal{S})}(\phi_1, \phi_2) = |\Delta(\phi_1, \phi_2)|$ for all $\phi_1, \phi_2 \in V(\mathcal{S})$, which immediately implies that $B(\mathcal{S})$ is a connected, isometric subgraph of $B^*(\mathcal{S})$. We use induction on $k = |\Delta(\phi_1, \phi_2)|$. The base case $k \in \{0, 1\}$ holds by the definition of $B(\mathcal{S})$. So assume that $k \geq 2$. Let $\mathcal{S}'$ denote the set of those $S' \in \Delta(\phi_1, \phi_2)$ with $\phi_2(S')$ minimal with respect to set inclusion. Select $T \in \mathcal{S}'$ such that $\phi_2(T)^c$ is maximal with respect to set inclusion. Define $\phi' \in V^*(\mathcal{S})$ by putting $\phi'(S) = \phi_2(S)$ if $S \neq T$ and $\phi'(T) = \phi_2(T)^c$.

We first establish $\phi' \in V(\mathcal{S})$ using Lemma 2.7. Consider $S \in \mathcal{S}$ and assume that $\phi_2(S) \subseteq \phi_2(T)$ and $S^c \cap \phi_2(T)^c = \emptyset$. Then we must have $S \in \Delta(\phi_1, \phi_2)$ since otherwise, by Lemma 2.5, S and T are compatible with $N(S, T) = \{\phi_2(S), \phi_2(T)^c\} = \{\phi_1(S), \phi_1(T)\}$, in contradiction to $\phi_1 \in V(\mathcal{S})$. By the choice of T and Lemma 2.5, $S \in \Delta(\phi_1, \phi_2)$ implies $\phi_2(S) = \phi_2(T)$ and $\phi_2(T)^c = \phi_2(S)^c$ and, thus, $S = T$, as required.

By construction we have $|\Delta(\phi_1, \phi')| = k - 1$ and obtain

$$\begin{aligned} d_{B(\mathcal{S})}(\phi_1, \phi_2) &\leq d_{B(\mathcal{S})}(\phi_1, \phi') + d_{B(\mathcal{S})}(\phi', \phi_2) = |\Delta(\phi_1, \phi')| + 1 = |\Delta(\phi_1, \phi_2)| \\ &\leq d_{B(\mathcal{S})}(\phi_1, \phi_2), \end{aligned}$$

where the first inequality holds by the triangle inequality of the shortest path metric, the first equality holds by induction and the last inequality holds since $B(\mathcal{S})$ is a subgraph of $B^*(\mathcal{S})$. This finishes the inductive proof of $d_{B(\mathcal{S})}(\phi_1, \phi_2) = |\Delta(\phi_1, \phi_2)|$.

It remains to show that $B(\mathcal{S})$ is a median graph. By [28, Theorem 5.6], since we have already established that $B(\mathcal{S})$ is an isometric subgraph of the hypercube $B^*(\mathcal{S})$, it suffices to show that $m_{B^*(\mathcal{S})}(\phi_1, \phi_2, \phi_3) \in V(\mathcal{S})$ for all $\phi_1, \phi_2, \phi_3 \in V(\mathcal{S})$. So, consider $S, S' \in \mathcal{S}, S \neq S'$. By the construction of $\phi = m_{B^*(\mathcal{S})}(\phi_1, \phi_2, \phi_3)$ we have

$$\{\phi(S), \phi(S')\} \in \{\{\phi_1(S), \phi_1(S')\}, \{\phi_2(S), \phi_2(S')\}, \{\phi_3(S), \phi_3(S')\}\}.$$

Thus, in view of $\phi_1, \phi_2, \phi_3 \in V(\mathcal{S})$ we have $\{\phi(S), \phi(S')\} \neq N(S, S')$. But this implies $\phi \in V(\mathcal{S})$, as required. □

3. Convex subsets of $V(\mathcal{S})$

As mentioned in the introduction, the Buneman graph $B(\mathcal{S})$ of a split system $\mathcal{S}$ on X comes naturally equipped with a map W which takes each element in X to a subset of the vertex set $V(\mathcal{S})$ of $B(\mathcal{S})$. As we will see in Section 4, a key property of the subsets that occur as images of the map W is that they are *convex*. Recall, that a subset C of the vertex set of a graph G is *convex* if, for all $c, c' \in C$, all shortest paths from c to c' in G only contain vertices in C. In this section, we shall present some results concerning certain special convex subsets of Buneman graphs that will be useful for understanding the map W, as well as having some independent interest.

The reader who is familiar with median graphs and the concept of *gated sets* in metric spaces [16] will probably have no difficulty relating our results in this and the next section to properties that have been well-studied for median graphs, some of which are considered as folklore. Even so, since we are considering labeled median graphs, in order to make this paper self-contained and to help the reader less familiar with median graphs, we will provide proofs for our results and use specific references to the literature on median graphs that we found helpful for translating properties of median graphs to the case of the Buneman graph $B(\mathcal{S})$. As mentioned in the introduction, more details concerning the general theory of median graphs and their origins can be found in, for example, [3], [8, Sec. 3], and [17, Chapter 12].

We begin by defining some of the convex subsets of interest. Let $\mathcal{S}$ be a split system on X, $S \in \mathcal{S}$ and $A \in S$. We define $V(S, A) = \{\phi \in V(\mathcal{S}) : \phi(S) = A\}$. Note that Lemma 2.6 implies that, for all $S = A|B \in \mathcal{S}$, the sets $V(S, A)$ and $V(S, B)$ are both non-empty. Thus, by the definition of an $\mathcal{S}$-map, $\{V(S, A), V(S, B)\}$ is a bipartition of $V(\mathcal{S})$. Moreover, since $d_{B(\mathcal{S})}(\phi_1, \phi_2) = |\Delta(\phi_1, \phi_2)|$ for all $\phi_1, \phi_2 \in V(\mathcal{S})$, $V(S, A)$ and $V(S, B)$ are both convex subsets of $V(\mathcal{S})$. In addition, we define

$$U(S, A) = \{\phi \in V(S, A) : \text{ there exists } \phi' \in V(S, B) \text{ with } \Delta(\phi, \phi') = \{S\}\}.$$

Note that $U(S, A) \neq \emptyset$. Our notation is inspired by [22, Sec. 2] where, for any two adjacent vertices a and b in a graph $G = (V, E)$, the sets

$$W_{ab} = \{w \in V : d_G(w, a) < d_G(w, b)\}$$
$$W_{ba} = \{w \in V : d_G(w, b) < d_G(w, a)\} \text{ and}$$
$$U_{ab} = \{u \in W_{ab} : u \text{ is adjacent to some } w \in W_{ba}\}$$

are defined. To show that $U(S, A)$ is convex, we use the fact that, for G a median graph, the set U_{ab} is convex [22, Thm. 2.7].

Lemma 3.1. *Let $\mathcal{S}$ be a split system on X, $S \in \mathcal{S}$ and $A \in S$. Then $U(S, A)$ is a convex subset of $V(\mathcal{S})$.*

Proof. First note that for $\phi \in U(S, A)$ and $\phi' \in U(S, B)$ with $\Delta(\phi, \phi') = \{S\}$ we have, by definition, that $\{\phi, \phi'\}$ is an edge of $B(\mathcal{S})$. Thus, by Theorem 2.8, we have

$$V(S, A) = \{\phi'' \in V(\mathcal{S}) : d_{B(\mathcal{S})}(\phi'', \phi) < d_{B(\mathcal{S})}(\phi'', \phi')\} = W_{\phi\phi'}$$

and, by symmetry, $V(S, B) = W_{\phi'\phi}$. Hence, $U(S, A) = U_{\phi\phi'}$ and, using the fact that $B(\mathcal{S})$ is a median graph, it follows from [22, Thm. 2.7] that $U(S, A)$ is a convex subset of $V(\mathcal{S})$. □

We remark that it can be shown that, for any $S = A|B \in \mathcal{S}$, the sets $V(S, A)$ and $V(S, B)$ form *half-spaces* in the median graph $B(\mathcal{S})$. Then $U(S, A)$ is the *boundary* of the half-space $V(S, A)$ and it is established in [27] that the boundary of a half-space in a median graph is always convex.

We now show how the sets $U(S, A)$ interact with split compatibility.

Lemma 3.2. *Let $\mathcal{S}$ be a split system on X, $S = A|B \in \mathcal{S}$ and $\phi, \psi \in U(S, A)$ with $\Delta(\phi, \psi) = \{T\}$. Then S and T are incompatible.*

Proof. By the definition of $U(S, A)$ there exist $\phi', \psi' \in U(S, B)$ with $\{S\} = \Delta(\phi, \phi')$ and $\{S\} = \Delta(\psi, \psi')$. This implies $\Delta(\phi', \psi') = \{T\}$. Hence, assuming $T = C|D$, we have

$$\{\{\phi(S), \phi(T)\}, \{\phi'(S), \phi'(T)\}, \{\psi(S), \psi(T)\}, \{\psi'(S), \psi'(T)\}\}$$
$$= \{\{A, C\}, \{A, D\}, \{B, C\}, \{B, D\}\}.$$

Thus, in view of (BG), we must have $N(S, T) = \emptyset$, implying that S and T are incompatible. □

Lemma 3.3. *Let $\mathcal{S}$ be a split system on X, $S = A|B \in \mathcal{S}$ and $\phi \in V(S, A)$. Suppose that $\phi' \in U(S, B)$ is such that $d_{B(\mathcal{S})}(\phi, \phi')$ is minimum over all $\phi' \in U(S, B)$ and that $\phi = \phi_0, \phi_1, \ldots, \phi_k = \phi'$, $k \geq 1$, is a path in $B(\mathcal{S})$ with $k = d_{B(\mathcal{S})}(\phi, \phi')$. Put $\Delta(\phi_i, \phi_{i+1}) = \{T_i\}$, $0 \leq i \leq k - 1$. Then S and T_i are compatible for all $0 \leq i \leq k - 2$.*

Proof. We use induction on k. The base case $k = 1$ is established by noting that the set $\{T_i : 0 \leq i \leq k - 2\}$ is empty.

Now consider $k \geq 2$. By induction, T_i and $S = T_{k-1}$ are compatible for all $1 \leq i \leq k-2$. It remains to show that T_0 and T_{k-1} are compatible.

First note that if T_i and T_j are compatible for some $0 \leq i < j \leq k - 1$ then we must have $N(T_j, T_i) = \{\phi_k(T_j), \phi_k(T_i)^c\}$, since

- $N(T_j, T_i) = \{\phi_k(T_j), \phi_k(T_i)\}$ cannot hold since $\phi_k \in V(\mathcal{S})$,
- $N(T_j, T_i) = \{\phi_k(T_j)^c, \phi_k(T_i)^c\} = \{\phi_i(T_j), \phi_i(T_i)\}$ cannot hold since $\phi_i \in V(\mathcal{S})$, and
- $N(T_j, T_i) = \{\phi_k(T_j)^c, \phi_k(T_i)\} = \{\phi_j(T_j), \phi_j(T_i)\}$ cannot hold since $\phi_j \in V(\mathcal{S})$.

Hence, by Lemma 2.5, if T_i and T_j are compatible for some $0 \leq i < j \leq k-1$ we have $\phi_k(T_j) \subseteq \phi_k(T_i)$ and $\phi_k(T_i)^c \subseteq \phi_k(T_j)^c$.

Assume for a contradiction that T_0 and T_{k-1} are incompatible. Then T_0 and T_i, $1 \leq i \leq k-2$, must also be incompatible since otherwise we would have

$$\phi_k(T_{k-1}) \subseteq \phi_k(T_i) \subseteq \phi_k(T_0) \text{ and } \phi_k(T_0)^c \subseteq \phi_k(T_i)^c \subseteq \phi_k(T_{k-1})^c,$$

in contradiction to our assumption that T_0 and T_{k-1} are incompatible.

Next, we define $\psi_i \in V^*(\mathcal{S})$, $1 \leq i \leq k$, by putting $\psi_1 = \phi_0$ and, for $i \geq 2$,

$$\psi_i(S') = \begin{cases} \psi_{i-1}(S') & \text{if } S' \neq T_{i-1} \\ \psi_{i-1}(S')^c & \text{if } S' = T_{i-1}. \end{cases}$$

Note that $\psi_i(S') = \phi_i(S')$, $1 \leq i \leq k$, for all $S' \neq T_0$. We use induction on i to show that $\psi_i \in V(\mathcal{S})$. The base case $i = 1$ holds in view of $\psi_1 = \phi_0 \in V(\mathcal{S})$. Consider $i \geq 2$. We need to show that $\{\psi_i(S'), \psi_i(S'')\} \neq N(S', S'')$ for all $S', S'' \in \mathcal{S}$, $S' \neq S''$. Indeed,

- for $S', S'' \in \mathcal{S} - \{T_{i-1}\}$ we have $\{\psi_i(S'), \psi_i(S'')\} = \{\psi_{i-1}(S'), \psi_{i-1}(S'')\} \neq N(S', S'')$ since, by induction, $\psi_{i-1} \in V(\mathcal{S})$,

- for $S' \in \mathcal{S} - \{T_{i-1}, T_0\}$ we have $\{\psi_i(S'), \psi_i(T_{i-1})\} = \{\phi_i(S'), \phi_i(T_{i-1})\} \neq N(S', T_{i-1})$ since $\phi_i \in V(\mathcal{S})$, and
- $\{\psi_i(T_0), \psi_i(T_{i-1})\} \neq N(T_0, T_{i-1})$ since T_0 and T_{i-1} are incompatible.

This finishes the inductive argument on i.

To finish the inductive argument on k, note that $\phi = \psi_1, \psi_2, \ldots, \psi_k$ is a path in $B(\mathcal{S})$ with $\psi_k \in U(S, B)$ and $d_{B(\mathcal{S})}(\psi_1, \psi_k) = |\Delta(\psi_1, \psi_k)| = k - 1$. But this is in contradiction to the choice of ϕ'. Hence, T_0 and $T_{k-1} = S$ must be compatible, as required. □

Lemma 3.4. *Let $\mathcal{S}$ be a split system on X, $S = A|B \in \mathcal{S}$, $\phi \in V(S, A)$ and $\psi \in U(S, B)$. In addition, let $\phi' \in U(S, B)$ be such that $d_{B(\mathcal{S})}(\phi, \phi')$ is minimum over all $\phi' \in U(S, B)$ and let $\phi = \phi_0, \phi_1, \ldots, \phi_k = \phi', \phi_{k+1}, \ldots, \phi_\ell = \psi$ be a path in $B(\mathcal{S})$ consisting of a shortest path $\phi_0, \phi_1, \ldots, \phi_k$ from ϕ to ϕ' and a shortest path $\phi_k, \phi_{k+1}, \ldots, \phi_\ell$ from ϕ' to ψ. Then $\phi_0, \phi_1, \ldots, \phi_\ell$ is a shortest path from ϕ to ψ in $B(\mathcal{S})$.*

Proof. Since $U(S, B)$ is a convex subset of $V(\mathcal{S})$ we have $\phi_i \in U(S, B)$ for $k \leq i \leq \ell$. Thus, by Lemma 3.2, all $T \in \Delta(\phi_k, \phi_\ell)$ are incompatible with S. Moreover, by Lemma 3.3, all $T \in \Delta(\phi_0, \phi_{k-1})$ are compatible with S. Hence, $\Delta(\phi, \psi)$ is the disjoint union of $\Delta(\phi_0, \phi_{k-1})$, $\{S\}$ and $\Delta(\phi_k, \phi_\ell)$, implying that $\phi_0, \phi_1, \ldots, \phi_\ell$ is a shortest path from ϕ to ψ in $B(\mathcal{S})$. □

The following lemma can be viewed as a special case of Property S_4 considered for *convexity structures* in [31] (see also [11]).

Lemma 3.5. *Let $\mathcal{S}$ be a split system on X and W, W'' be disjoint, non-empty, convex subsets of $V(\mathcal{S})$. In addition, let $\phi \in W$ and $\phi'' \in W''$ be such that $d_{B(\mathcal{S})}(\phi, \phi'')$ is minimum over all $\phi \in W$, $\phi'' \in W''$. Then $d_{B(\mathcal{S})}(\phi, \phi'') > 0$ and, for all $S \in \Delta(\phi, \phi'')$, there exists some $A \in S$ with $W \subseteq V(S, A)$ and $W'' \subseteq V(S, A_S^c)$.*

Proof. Since $W \cap W'' = \emptyset$ we clearly have $d = d_{B(\mathcal{S})}(\phi, \phi'') > 0$. Consider $S = A|B \in \Delta(\phi, \phi'')$ and assume for a contradiction that $W \cap V(S, A) \neq \emptyset$ and $W \cap V(S, B) \neq \emptyset$. Assume without loss of generality that $\phi \in V(S, A)$. Since W is convex, it contains a shortest path from ϕ to some vertex in $V(S, B)$, implying that there exists some $\psi \in U(S, B) \cap W$. Let $\phi' \in U(S, B)$ be such that $d_{B(\mathcal{S})}(\phi, \phi')$ is minimum over all $\phi' \in U(S, B)$. By Lemma 3.4, there exists a shortest path $\phi = \phi_0, \phi_1, \ldots, \phi_d = \phi''$ in $B(\mathcal{S})$ with $\phi' = \phi_i$ for some $1 \leq i \leq d$. In particular, $\phi' \notin W$ by the choice of ϕ and ϕ''. Again by Lemma 3.4, there also exists a shortest path in $B(\mathcal{S})$ from ϕ to ψ that contains ϕ'. Since W is convex, this implies $\phi' \in W$, a contradiction. Hence, we must have either $W \subseteq V(S, A)$ and, by symmetry, $W'' \subseteq V(S, B)$ or $W \subseteq V(S, B)$ and, by symmetry, $W'' \subseteq V(S, A)$, as required. □

4. Labeling maps in Buneman graphs

In this section, we show that, for any split system $\mathcal{S}$ on X, there is a natural labeling map W that takes each $x \in X$ to a convex subset $W(x)$ of the vertex set $V(\mathcal{S})$ of the Buneman graph $B(\mathcal{S})$. In the main result of this section, we show that we can represent the distance $d_{\mathcal{S}}$ on X associated to $\mathcal{S}$ in terms of the distance in $B(\mathcal{S})$ between the subsets given by W.

Let $\mathcal{S}$ be a split system on X. Then we obtain the *labeling map* $W : X \to \mathcal{P}(V(\mathcal{S}))$ by setting

$$W(x) = \{\phi \in V(\mathcal{S}) \,:\, x \in \phi(S) \cup S^c \text{ for all } S \in \mathcal{S}\}$$

for all $x \in X$.

Example 4.1. For the Buneman graph $B(\mathcal{S})$ considered in Example 2.3 we obtain the following labeling map W:

$$W(x_1) = \{\phi_1\},\ W(x_2) = \{\phi_2\},\ W(x_3) = \{\phi_6\},\ W(x_4) = \{\phi_{10}\},\ W(x_5) = \{\phi_8\}$$
$$W(x_6) = \{\phi_3, \phi_4\},\ W(x_7) = \{\phi_4, \phi_5, \phi_7, \phi_9\}$$

We first present some key properties of the labeling map W.

Lemma 4.2. *Let $\mathcal{S}$ be a split system on X. Then the following hold:*

(a) $W(x) \neq \emptyset$ for all $x \in X$.
(b) $|W(x)| = 1$ for all $x \in X$ if and only if all $S \in \mathcal{S}$ are full splits of X.
(c) $W(x)$ is a convex subset of $V(\mathcal{S})$ for all $x \in X$.

Proof. (a) This is just Lemma 2.6 rephrased in terms of the labeling map W.

(b) First assume that $\mathcal{S}$ is a set of full splits of X. Then, by definition, $W(x)$ contains precisely the $\mathcal{S}$-map $\phi \in V(\mathcal{S})$ with $\phi(S) = S(x)$ for all $S \in \mathcal{S}$, implying $|W(x)| = 1$.

Next assume that there exists some $S = A|B \in \mathcal{S}$ such that S is not a full split of X. Consider $x \in S^c$. For this $x = x_0$, recall the construction of an $\mathcal{S}$-map $\phi \in W(x)$ in (a): We have $S \in \mathcal{R}_0$ and, thus, for x_1 we can select some element in A or some element in B. Depending on this choice of x_1 we obtain two distinct $\mathcal{S}$-maps in $W(x)$, implying $|W(x)| > 1$.

(c) Consider $x \in X$ and assume for a contradiction that there exist $\phi_1, \phi_2 \in W(x)$ such that there is some shortest path in $B(\mathcal{S})$ from ϕ_1 to ϕ_2 whose vertices are not all contained in $W(x)$. Let ϕ be the first vertex on this path after ϕ_1 that is not in $W(x)$ and let $\phi' \in W(x)$ be the vertex immediately before ϕ on this path. Let $\{S\} = \Delta(\phi', \phi)$. Then, in view of $\phi' \in W(x)$ and $\phi \notin W(x)$, we have $x \notin \phi(S) \cup S^c$. Moreover, by Theorem 2.8, we have $S \in \Delta(\phi_1, \phi_2)$ and, therefore, $\phi_1(S)^c \cup S^c = \phi'(S)^c \cup S^c = \phi(S) \cup S^c = \phi_2(S) \cup S^c$. But then $x \notin \phi_2(S) \cup S^c$ in contradiction to $\phi_2 \in W(x)$. □

Proposition 4.3. *Let $\mathcal{S}$ be a split system on X, $S = A|B \in \mathcal{S}$ and $x \in X$. Then $W(x) \subseteq V(S, A)$ if and only if $x \in A$.*

Proof. First assume $x \in A$ and consider $\phi \in W(x)$. Then, by the definition of $W(x)$, we have $\phi(S) = A$ and, therefore, $\phi \in V(S, A)$, as required.

Conversely, assume $W(x) \subseteq V(S, A)$. Consider $\phi \in W(x)$ and $\phi' \in U(S, B)$ such that $d = d_{B(\mathcal{S})}(\phi, \phi')$ is minimum over all $\phi \in W(x)$ and $\phi' \in U(S, B)$. Then we have $d \geq 1$. Let $\phi = \phi_0, \phi_1, \ldots, \phi_d = \phi'$ be a shortest path in $B(\mathcal{S})$ from ϕ to ϕ'. By the choice of ϕ and ϕ' we have $\phi_i \notin W(x)$ for $1 \leq i \leq d$. Put $\{T_i\} = \Delta(\phi_i, \phi_{i+1})$ for $0 \leq i \leq d-1$. Then we have $T_{d-1} = S$. Moreover, $x \in \phi_0(T_0)$ must hold in view $\phi_0 \in W(x)$ and $\phi_1 \notin W(x)$. By Lemma 3.3, T_0 and S either coincide or are compatible. More specifically, from the proof of Lemma 3.3 we have $\phi_d(S) \subseteq \phi_d(T_0)$ and $\phi_d(T_0)^c \subseteq \phi_d(S)^c$. Hence, $x \in \phi_0(T_0) = \phi_d(T_0)^c \subseteq \phi_d(S)^c = A$, as required. □

Let $\mathcal{S}$ be a split system on X and consider the Buneman graph $B(\mathcal{S})$ with labeling map W. We define the distance $d_{(B(\mathcal{S}),W)} : X \times X \to \mathbb{N}$ by putting

$$d_{(B(\mathcal{S}),W)}(x, x') = \min\{d_{B(\mathcal{S})}(\phi, \phi') \,:\, \phi \in W(x) \text{ and } \phi' \in W(x')\}$$

for all $x, x' \in X$.

Example 4.4. For the Buneman graph $B(\mathcal{S})$ with labeling map W considered in Example 4.1 we have $d_{(B(\mathcal{S}),W)}(x_1, x_7) = 2$, which coincides with the value $d_{\mathcal{S}}(x_1, x_7)$ given in Example 2.2.

Theorem 4.5. *Let $\mathcal{S}$ be a split system on X and $B(\mathcal{S})$ be the associated Buneman graph with labeling map W. Then $d_{(B(\mathcal{S}),W)} = d_{\mathcal{S}}$.*

Proof. Consider $x, x' \in X$. If $x = x'$ we clearly have $d_{(B(\mathcal{S}),W)}(x, x') = d_{\mathcal{S}}(x, x') = 0$. So, assume $x \neq x'$ and put

$$\mathcal{S}^* = \{S \in \mathcal{S} : S(x) \neq \emptyset,\ S(x') \neq \emptyset \text{ and } S(x) \neq S(x')\}.$$

Then $d_{\mathcal{S}}(x, x') = |\mathcal{S}^*|$. Select $\phi \in W(x)$, $\phi' \in W(x')$ such that $d_{(B(\mathcal{S}),W)}(x, x') = d_{B(\mathcal{S})}(\phi, \phi')$. By Theorem 2.8, it suffices to show that $\Delta(\phi, \phi') = \mathcal{S}^*$.

First consider $S \in \mathcal{S}^*$. By the definition of $\mathcal{S}^*$ and in view of $\phi \in W(x)$ and $\phi' \in W(x')$, we have $\phi(S) = S(x) \neq S(x') = \phi'(S)$, implying $S \in \Delta(\phi, \phi')$. Hence $\mathcal{S}^* \subseteq \Delta(\phi, \phi')$.

Conversely, consider $S = A|B \in \Delta(\phi, \phi')$. Assume without loss of generality that $\phi(S) = A$ and $\phi'(S) = B$. Note that, by Lemma 4.2, $W(x)$ and $W(x')$ are non-empty, convex subsets of $V(\mathcal{S})$ and, in view of $S \in \Delta(\phi, \phi')$, we must have $W(x) \cap W(x') = \emptyset$. Thus, by the choice of $\phi \in W(x)$ and $\phi' \in W(x')$, Lemma 3.5 implies $W(x) \subseteq V(S, A)$ and $W(x') \subseteq V(S, B)$. Hence, by Proposition 4.3, we have $x \in A$ and $x' \in B$, implying $S \in \mathcal{S}^*$. It follows that also $\Delta(\phi, \phi') \subseteq \mathcal{S}^*$, as required. □

5. Crossing graphs

In this section, we show that, we can count induced subgraphs of $B(\mathcal{S})$ that are isomorphic to hypercubes of a given dimension directly in terms of the structure of $\mathcal{S}$, for any split system $\mathcal{S}$ on X. We note that the results that we present in this section are closely related to well-known results on median graphs which are reviewed in [24]; the original references for the results that we use can be found in that review.

We first define the *crossing graph* $B(\mathcal{S})^\#$ of $B(\mathcal{S})$ to be the graph that has vertex set $\mathcal{S}$ and edge set consisting of those 2-element subsets $\{S = A|B,\ S' = A'|B'\} \subseteq \mathcal{S}$ for which all four intersections

$$V(S, A) \cap V(S', A'), \quad V(S, A) \cap V(S', B'),$$
$$V(S, B) \cap V(S', A') \quad \text{and} \quad V(S, B) \cap V(S', B')$$

are non-empty. In the literature, the vertices of the crossing graph of a graph G are also referred to as the *colors* of G (see e.g. [23]).

Theorem 5.1. *Let $\mathcal{S}$ be a split system on X and $S, S' \in \mathcal{S}$ with $S \neq S'$. Then $\{S, S'\}$ is an edge of $B(\mathcal{S})^\#$ if and only if S and S' are incompatible.*

Proof. Assume that $\{S = A|B,\ S' = A'|B'\}$ is an edge of $B(\mathcal{S})^\#$. Then, by the definition of $B(\mathcal{S})^\#$, there exist $\phi \in V(S, A) \cap V(S', A')$ and $\psi \in V(S, A) \cap V(S', B')$. Consider a shortest path in $B(\mathcal{S})$ from ϕ to ψ. Using the fact that $V(S, A)$ is a convex subset of $V(\mathcal{S})$, this shortest path must contain some $\phi' \in V(S, A) \cap U(S', B')$. By symmetry, there also exists some

$\phi'' \in V(S, B) \cap U(S', B')$. Since, by Lemma 3.1, also $U(S', B')$ is a convex subset of $V(\mathcal{S})$, any shortest path from ϕ' to ϕ'' in $B(\mathcal{S})$ is contained in $U(S', B')$. Moreover, any such path must contain $\psi', \psi'' \in U(S', B')$ with $\Delta(\psi', \psi'') = \{S\}$. Hence, by Lemma 3.2, S and S' are incompatible.

Conversely, assume that $\{S = A|B,\ S' = A'|B'\}$ is not an edge of $B(\mathcal{S})^{\#}$. Then, again by the definition of $B(\mathcal{S})^{\#}$, we may assume without loss of generality that $V(S, A) \subseteq V(S', A')$ and $V(S', B') \subseteq V(S, B)$. Consider $x \in A$. By Proposition 4.3, we have $W(x) \subseteq V(S, A) \subseteq V(S', A')$. Thus, again by Proposition 4.3, $x \in A'$. Hence, $A \subseteq A'$. By symmetry, we also have $B' \subseteq B$. It follows that S and S' are compatible. □

By Theorem 5.1, the crossing graph $B(\mathcal{S})^{\#}$ for a split system $\mathcal{S}$ on X coincides with the *incompatibility graph* $Incomp(\mathcal{S})$ that also has vertex set $\mathcal{S}$ and whose edge set consists of all 2-element subsets $\{S, S'\} \subseteq \mathcal{S}$ with S and S' incompatible. Using results presented in [23] concerning the relationship between median graphs and their associated crossing graphs (see that paper also for more details about the origins of crossing graphs), we immediately obtain some structural information for the Buneman graph $B(\mathcal{S})$ in terms of $Incomp(\mathcal{S})$. We call a hypercube of dimension i an *i-cube*, for short.

Corollary 5.2. *Let $\mathcal{S}$ be a split system on X.*

(a) $B(\mathcal{S})$ is a tree if and only if $Incomp(\mathcal{S})$ is a coclique.
(b) $B(\mathcal{S})$ is an $|\mathcal{S}|$-cube if and only if $Incomp(\mathcal{S})$ is a clique.

Proof. In view of the fact that $B(\mathcal{S})^{\#}$ and $Incomp(\mathcal{S})$ coincide, (a) is an immediate consequence of [23, Corollary 4.3] and (b) is an immediate consequence of [23, Proposition 4.1]. □

Using results presented in [24], we now show how to count the number of cubes in the Buneman graph of a split system $\mathcal{S}$ on X. For $i \in \mathbb{N}$, we denote by $\beta_i(\mathcal{S})$ the number of i-element subsets of $\mathcal{S}$ whose elements are pairwise incompatible. Note that $\beta_0(\mathcal{S}) = 1$ and $\beta_1(\mathcal{S}) = |\mathcal{S}|$. In addition, we denote by $\alpha_i(\mathcal{S})$ the number of induced subgraphs of $B(\mathcal{S})$ that are i-cubes. In particular, $\alpha_0(\mathcal{S})$ is the number of vertices and $\alpha_1(\mathcal{S})$ is the number of edges of $B(\mathcal{S})$.

Theorem 5.3. *Let $\mathcal{S}$ be a split system on X. Then*

$$\alpha_i(\mathcal{S}) = \sum_{k \geq i} \binom{k}{i} \beta_k(\mathcal{S}) \tag{2}$$

holds for all $i \in \mathbb{N}$.

Proof. (2) restates formula (13.7) given in [24, p. 331] for the case that all $S \in \mathcal{S}$ are full splits. However, as pointed out in the proof of formula (13.7) in [24], the crucial fact used is that the Buneman graph of a split system is a hypercube if and only if the splits in the split system are pairwise incompatible, a fact that, by Corollary 5.2(b), also holds if the split system contains partial splits. □

Example 5.4. For the split system $\mathcal{S}$ considered in Example 2.3 we have $\beta_0(\mathcal{S}) = 1$, $\beta_1(\mathcal{S}) = 8$, $\beta_2(\mathcal{S}) = 1$ and $\beta_k(\mathcal{S}) = 0$ for all $k \geq 3$. Thus, with formula (2) we obtain $\alpha_0(\mathcal{S}) = 10$, $\alpha_1(\mathcal{S}) = 10$, $\alpha_2(\mathcal{S}) = 1$ and $\alpha_i(\mathcal{S}) = 0$ for all $i \geq 3$, which coincide with the values obtained directly from Fig. 2.

6. Tree representations of split systems

An *X-tree* is a tree T with vertex set V together with a map $\phi : X \to V$ such that, for all vertices $v \in V$ with degree at most 2, v is contained in $\phi(X)$ [30, p. 16]. Such trees play an important role in the theory of phylogenetics. For example, a fundamental result in phylogenetics states that split systems on X consisting of full splits that are pairwise compatible are in one-to-one correspondence with X-trees [10] (see also [30, Theorem 3.1.4]). In this section, we generalize this result to split systems that may contain partial splits. In particular, we characterize those set-labeled trees which correspond to split systems consisting of pairwise compatible partial splits.

To this end, we first collect some key properties of Buneman graphs that are a tree.

Lemma 6.1. *Let $\mathcal{S}$ be a split system on X such that $G = B(\mathcal{S}) = (V, E)$ is a tree. Then the labeling map $W : X \to \mathcal{P}(V)$ has the following properties:*

(W1) For all $x \in X$, $W(x) \neq \emptyset$ and the subgraph of G induced by $W(x)$ is a tree.
(W2) For all vertices $v \in V$ with $\deg(v) = 1$ there exists some $x \in X$ such that $W(x) = \{v\}$.
(W3) For all vertices $v \in V$ with $\deg(v) = 2$ and adjacent vertices u and w, $u \neq w$, there exists some $x \in X$ such that $v \in W(x)$ and $|\{u, w\} \cap W(x)| \leq 1$.

Proof. (W1): Consider $x \in X$. Then, by Lemma 4.2(a), $W(x) \neq \emptyset$ and, by Lemma 4.2(c), $W(x)$ is a convex subset of V, implying that $W(x)$ induces a subtree in G.

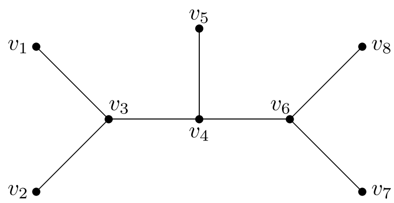

Fig. 3. The tree referred to in Example 6.2.

(W2): Consider $v \in V$ with $\deg(v) = 1$. Let u be the vertex that is adjacent to v in G and $\{S = A|B\} = \Delta(u, v)$. Without loss of generality we assume that $\{v\} = V(S, A)$. Let $x \in A$. Then, by Lemma 4.2(a) and Proposition 4.3, we have $\emptyset \neq W(x) \subseteq V(S, A) = \{v\}$, implying $W(x) = \{v\}$.

(W3): Consider $v \in V$ with $\deg(v) = 2$. Let u and w be the vertices that are adjacent to v in G. Put $\{S = A|B\} = \Delta(u, v)$ and $\{S' = A'|B'\} = \Delta(v, w)$. By Theorem 2.8, we have $2 = d_G(u, w) = |\Delta(u, w)| = |\{S, S'\}|$ and, thus, $S \neq S'$. Without loss of generality we assume $v \in V(S, B) \cap V(S', A')$. Assume for a contradiction that, for all $x \in X$, $v \notin W(x)$ or $\{u, v, w\} \subseteq W(x)$. Then, by Proposition 4.3, for all $x \in X$, we have $x \in A$ if and only if $W(x) \subseteq V(S, A)$ if and only if $W(x) \subseteq V(S', A')$ if and only if $x \in A'$. Hence, $A = A'$. By symmetry, we also have $B = B'$, implying $S = S'$, a contradiction. Thus, there exists $x \in X$ with $v \in W(x)$ and $|\{u, w\} \cap W(x)| \leq 1$. □

We call an ordered pair (G, W) consisting of a tree $G = (V, E)$ and a map $W : X \to \mathcal{P}(V)$ a *weak X-tree* if W satisfies properties (W1)-(W3) stated in Lemma 6.1. Note that if $|W(x)| = 1$ for all $x \in X$ then a weak X-tree corresponds to a usual X-tree as defined above[2].

Now, for any weak X-tree $(G = (V, E), W)$ define the split system $\mathcal{S}_{(G,W)} = \{S_e : e \in E\}$ where, for all $e = \{u, w\} \in E$, the partial split S_e of X is obtained by removing e from G. More precisely, let $V_{e,u}$ and $V_{e,w}$ denote the vertex sets of the two connected components of $G - e$ that contain u and w, respectively, and put $A_{e,u} = \{x \in X : W(x) \subseteq V_{e,u}\}$ and $A_{e,w} = \{x \in X : W(x) \subseteq V_{e,w}\}$. In view of (W2) we have $A_{e,u} \neq \emptyset$ and $A_{e,w} \neq \emptyset$. We put $S_e = A_{e,u}|A_{e,w}$.

Example 6.2. Consider the tree $G = (V, E)$ with vertex set $V = \{v_1, \ldots, v_8\}$ depicted in Fig. 3, the set $X = \{x_1, \ldots, x_7\}$ and the map $W : X \to \mathcal{P}(V)$ with

$$W(x_1) = \{v_1\}, \; W(x_2) = \{v_2\}, \; W(x_3) = \{v_7\}, \; W(x_4) = \{v_8\}, \; W(x_5) = \{v_5\}$$
$$W(x_6) = \{v_3, v_4, v_6\}, \; W(x_7) = \{v_4, v_6, v_7, v_8\}.$$

Then (G, W) is a weak X-tree and for the edge $e = \{v_3, v_4\} \in E$ we obtain the partial split $S_e = \{x_1, x_2\}|\{x_3, x_4, x_5, x_7\}$.

As mentioned at the beginning of this section, it is known (see e.g. [30, Thm. 3.1.4]) that for a split system $\mathcal{S}$ consisting of full splits of X there exists an X-tree (G, W) with $\mathcal{S} = \mathcal{S}_{(G,W)}$ if and only if the splits in $\mathcal{S}$ are pairwise compatible. Moreover, this X-tree is unique up to isomorphism. The next theorem gives a generalization of this result to split systems that may contain partial splits. We say that two weak X-trees $(G = (V, E), W)$ and $(G' = (V', E'), W')$ are *isomorphic* if there exists a graph isomorphism $f : V \to V'$ such that $f(W(x)) = W'(x)$ for all $x \in X$.

Theorem 6.3. *Let $\mathcal{S}$ be a split system on X.*

(a) If the Buneman graph $B(\mathcal{S}) = (V, E)$ is a tree with labeling map $W : X \to \mathcal{P}(V)$ then $(B(\mathcal{S}), W)$ is a weak X-tree with $\mathcal{S} = \mathcal{S}_{(B(\mathcal{S}),W)}$.
(b) If there exists a weak X-tree (G', W') with $\mathcal{S} = \mathcal{S}_{(G',W')}$ then (G', W') is isomorphic to $(B(\mathcal{S}), W)$.
(c) There exists a weak X-tree (G, W) with $\mathcal{S} = \mathcal{S}_{(G,W)}$ if and only if the partial splits in $\mathcal{S}$ are pairwise compatible.

Proof. (a): If the Buneman graph $B(\mathcal{S})$ is a tree then, by Lemma 6.1, the labeling map W satisfies (W1)-(W3). To show that $\mathcal{S} = \mathcal{S}_{(B(\mathcal{S}),W)}$, first consider $S = A|B \in \mathcal{S}$. Let $x \in A$ and $x' \in B$, implying that $S(x) \neq \emptyset$, $S(x') \neq \emptyset$ and $S(x) \neq S(x')$. Select $\phi \in W(x)$ and $\phi' \in W(x')$ such that $d_{B(\mathcal{S})}(\phi, \phi') = d_{(B(\mathcal{S}),W)}(x, x')$. In the proof of Theorem 4.5 we have shown that then $S \in \Delta(\phi, \phi')$. Hence, there exists an edge $e = \{\psi, \psi'\}$ in $B(\mathcal{S})$ with $\Delta(\psi, \psi') = \{S\}$.

To establish $\mathcal{S} = \mathcal{S}_{(B(\mathcal{S}),W)}$, it remains to show that, for all edges $e = \{\phi, \phi'\}$ in $B(\mathcal{S})$, we have $\Delta(\phi, \phi') = \{S_e\}$. Let $\Delta(\phi, \phi') = \{S = A|B\}$. Then $V(S, A)$ and $V(S, B)$ are the vertex sets of the two connected components of $B(\mathcal{S}) - e$. Thus, by Proposition 4.3, we have $S_e = A|B$, as required.

[2] In [20] the term weak X-tree is also used but refers to a different generalization of usual X-trees than considered here.

with $x \in A_{e,u} \cap A_{e',w'}$, $z \in A_{e,w} \cap A_{e',u'}$ and either $y \in A_{e,w} - A_{e',u'}$ or $y \in A_{e',w'} - A_{e,u}$. Thus $S_e \neq S_{e'}$, implying that for each $S \in \mathcal{S}$ there is a unique $e = e_S = \{u_S, w_S\} \in E'$ with $S_e = S$.

We define for each $v \in V'$ an $\mathcal{S}$-map ϕ_v by putting

$$\phi_v(S) = \begin{cases} A_{e_S,u_S} & \text{if } v \in V'_{e_S,u_S} \\ A_{e_S,w_S} & \text{if } v \in V'_{e_S,w_S} \end{cases}$$

for all $S \in \mathcal{S}$. It follows immediately that $\phi_v = \phi_{v'}$ for $v, v' \in V'$ implies $v = v'$, $\Delta(\phi_{u_S}, \phi_{w_S}) = \{S\}$ for all $S \in \mathcal{S}$, and $d_{G'}(v, v') = |\Delta(\phi_v, \phi_{v'})|$ for all $v, v' \in V'$. In addition, the partial splits in $\mathcal{S}$ are pairwise compatible and we have $\phi_v \in V(\mathcal{S})$ for all $v \in V'$. To see this, consider $S, S' \in \mathcal{S}$, $S \neq S'$. Without loss of generality we assume that $u_{S'} \in V'_{e_S,w_S}$ and $u_S \in V'_{e_{S'},w_{S'}}$. Then $A_{e_S,u_S} \subseteq A_{e_{S'},w_{S'}}$ and $A_{e_{S'},u_{S'}} \subseteq A_{e_S,w_S}$ and, therefore, $N(S, S') = \{A_{e_S,u_S}, A_{e_{S'},u_{S'}}\}$. In particular, S and S' are compatible. Moreover, in view of $V'_{e_S,u_S} \cap V'_{e_{S'},u_{S'}} = \emptyset$, we have $\{\phi_v(S), \phi_v(S')\} \neq N(S, S')$, as required.

The fact that the partial splits in $\mathcal{S}$ are pairwise compatible implies, by Corollary 5.2(a) and Theorem 5.3, that the Buneman graph $B(\mathcal{S})$ is a tree with $|\mathcal{S}| + 1 = |V'|$ vertices. Thus, we have $\{\phi_v : v \in V'\} = V(\mathcal{S})$ and $f : V' \to V(\mathcal{S})$ with $f(v) = \phi_v$ is a graph isomorphism between G' and $B(\mathcal{S})$. It remains to show that $f(W'(x)) = W(x)$ for all $x \in X$. Consider $v \in V'$ and $x \in X$. For each $S \in \mathcal{S}$ assume without loss of generality that $v \in V'_{e_S,u_S}$. Then have $v \in W'(x)$ if and only if $x \in X - A_{e_S,w_S}$ for all $S \in \mathcal{S}$ if and only if $x \in \phi_v(S) \cup S^c$ for all $S \in \mathcal{S}$ if and only if $\phi_v \in W(x)$, as required.

(c): It follows from (a) and (b), that there exists a weak X-tree (G, W) with $\mathcal{S} = \mathcal{S}_{(G,W)}$ if and only if the Buneman graph $B(\mathcal{S})$ is a tree which, by Corollary 5.2(a), is the case if and only if the partial splits in $\mathcal{S}$ are pairwise compatible. □

7. Subtree distances

A subtree distance is essentially a distance that can be represented by taking the distance between a collection of subtrees of an edge-weighted tree [18]. It is known that any subtree distance has a certain unique minimal representation that can be efficiently computed [1,25]. In this section, we prove that this minimal representation can also be obtained in terms of the Buneman graph.

We first recall some definitions and facts concerning subtree distances. An ordered pair (G, ω) consisting of a tree $G = (V, E)$ and a map $\omega : E \to \mathbb{R}_{>0}$ is called an *edge-weighted tree*. For all $u, v \in V$, we denote by $d_{(G,\omega)}(u, v)$ the sum of the weights of those edges of G that lie on the path from u to v in G. A *subtree realization* of a distance d on X is an ordered pair $((G, \omega), W)$ consisting of an edge-weighted tree $(G = (V, E), \omega)$ and a map $W : X \to \mathcal{P}(V)$ such that

(SR1) W has property (W1) stated in Lemma 6.1 and
(SR2) $d(x, y) = \min\{d_{(G,\omega)}(u, v) : u \in W(x) \text{ and } v \in W(y)\}$ for all $x, y \in X$.

In [18] it is shown that for distances d on X the following are equivalent:

(D1) d has a subtree realization.
(D2) There exist a set $\mathcal{S}$ of pairwise compatible partial splits of X and a map $\alpha : \mathcal{S} \to \mathbb{R}_{>0}$ such that $d = \sum_{S\in\mathcal{S}} \alpha(S) \cdot d_S$.
(D3) d satisfies a certain 4-point condition given in [18, Theorem 1.2].

We call distances d on X for which any of (D1)-(D3) hold *subtree distances*, for short. By [19, Thm. 4.13 and Rem. 4.18], the set $\mathcal{S}$ of partial splits and the map α in (D2) are uniquely determined by d. In particular, $\mathcal{S} = \emptyset$ if and only if $d(x, y) = 0$ for all $x, y \in X$.

We now consider minimal subtree realizations. To this end, let $(G = (V, E), \omega)$ be an edge-weighted tree and $W : X \to \mathcal{P}(V)$. Then

- *suppressing* a vertex $v \in V$ having precisely two adjacent vertices u and w means we put $V_{\sim v} = V - \{v\}$, $E_{\sim v} = (E - \{\{u, v\}, \{v, w\}\}) \cup \{\{u, w\}\}$, $\omega_{\sim v} : E_{\sim v} \to \mathbb{R}_{>0}$ with

$$\omega_{\sim v}(e) = \begin{cases} \omega(e) & \text{if } e \neq \{u, w\} \\ \omega(\{u, v\}) + \omega(\{v, w\}) & \text{if } e = \{u, w\} \end{cases}$$

  and $W_{\sim v} : X \to \mathcal{P}(V_{\sim v})$ with $W_{\sim v}(x) = W(x) - \{v\}$ for all $x \in X$.
- *contracting* an edge $e = \{u, v\} \in E$ means we put $V_{\sim e} = (V - \{u, v\}) \cup \{w\}$ with some new vertex $w \notin V$,

$$\begin{aligned} E_{\sim e} =& (E - \{e \in E : \{u, v\} \cap e \neq \emptyset\}) \\ & \cup \{\{a, w\} : a \in V_{\sim e} \text{ and } \{\{a, v\}, \{a, u\}\} \cap E \neq \emptyset\}, \end{aligned}$$

  $\omega_{\sim e} : E_{\sim e} \to \mathbb{R}_{>0}$ with and $W_{\sim e} : X \to \mathcal{P}(V_{\sim e})$ with

$$W_{\sim e}(x) = \begin{cases} W(x) & \text{if } \{u, v\} \cap W(x) = \emptyset \\ (W(x) - \{u, v\}) \cup \{w\} & \text{if } \{u, v\} \cap W(x) \neq \emptyset \end{cases}$$

for all $x \in X$.

A subtree realization $((G = (V, E), \omega), W)$ of a distance d on X is *minimal* if, for all $v \in V$ with $\deg(v) = 2$, $(((V_{\sim v}, E_{\sim v}), \omega_{\sim v}), W_{\sim v})$ is not a subtree realization of d and, for all $e \in E$, $(((V_{\sim e}, E_{\sim e}), \omega_{\sim e}), W_{\sim e})$ is not a subtree realization of d. Two subtree realizations $((G = (V, E), \omega), W)$ and $((G' = (V', E'), \omega'), W')$ are *isomorphic* if there exists a graph isomorphism $f : V \to V'$ such that $\omega(\{u, v\}) = \omega'(\{f(u), f(v)\})$ for all $\{u, v\} \in E$ and $f(W(x)) = W'(x)$ for all $x \in X$. In [25, Theorem 2] it is shown that if a distance d on X has a subtree realization then d has a minimal subtree realization that is unique up to isomorphism (see also [1] for related results).

We call a subtree realization $((G, \omega), W)$ of a distance d on X a *subtree representation* of d if (G, W) is a weak X-tree. It is known that a distance d on X is a metric on X characterized by a certain 4-point condition if and only if d has a subtree representation $((G, \omega), W)$ with (G, W) an X-tree (see e.g. [30, Thm. 7.1.8 and Thm. 7.2.6]). The next theorem naturally generalizes this result by giving a characterization of the minimal subtree realization of a subtree distance on X.

Theorem 7.1. *Let d be a subtree distance on X such that $d = \sum_{S \in \mathcal{S}} \alpha(S) \cdot d_S$ for some non-empty set $\mathcal{S}$ of pairwise compatible partial splits of X and some map $\alpha : \mathcal{S} \to \mathbb{R}_{>0}$. Consider the Buneman graph $B(\mathcal{S}) = (V(\mathcal{S}), E)$ with labeling map W and put $\omega_\alpha : E \to \mathbb{R}_{>0}$ such that $\omega_\alpha(\{\phi, \phi'\}) = \alpha(S)$ for all $\phi, \phi' \in V(\mathcal{S})$ with $\Delta(\phi, \phi') = \{S\}$. Then the following holds:*

(a) $((B(\mathcal{S}), \omega_\alpha), W)$ is a subtree representation of d.
(b) Every subtree representation of d is a minimal subtree realization of d.
(c) All subtree representations of d are isomorphic to $((B(\mathcal{S}), \omega_\alpha), W)$.

Proof. (a) Since the partial splits in $\mathcal{S}$ are pairwise compatible, $(B(\mathcal{S}), W)$ is a weak X-tree by Theorem 6.3. Put $\omega = \omega_\alpha$ and consider $x, x' \in X$. Select $\psi \in W(x)$ and $\psi' \in W(x')$ such that

$$d_{(B(\mathcal{S}),\omega)}(\psi, \psi') = \min\{d_{(B(\mathcal{S}),\omega)}(\phi, \phi') : \phi \in W(x) \text{ and } \phi' \in W(x')\}.$$

Put $\mathcal{S}^* = \{S \in \mathcal{S} : S(x) \neq \emptyset,\ S(x') \neq \emptyset \text{ and } S(x) \neq S(x')\}$. Then, by Theorem 4.5, we have

$$d_{(B(\mathcal{S}),\omega)}(\psi, \psi') = \sum_{S \in \Delta(\psi,\psi')} \alpha(S) = \sum_{S \in \mathcal{S}^*} \alpha(S) = \sum_{S \in \mathcal{S}} \alpha(S) \cdot d_S(x, y) = d(x, y).$$

Hence, $((B(\mathcal{S}), \omega_\alpha), W)$ is a subtree representation of d.

(b) Let $((G' = (V', E'), \omega'), W')$ be a subtree representation of d. First consider a vertex $v \in V'$ that has precisely two adjacent vertices u and w. In view of (W3), there exists some $x \in X$ with $v \in W'(x)$ and $|\{u, w\} \cap W'(x)| \leq 1$. If $W'(x) = \{v\}$ we have $W'_{\sim v}(x) = \emptyset$ and, thus, $(((V'_{\sim v}, E'_{\sim v}), \omega'_{\sim v}), W'_{\sim v})$ is not a subtree realization of d because (W1) is violated. So, assume without loss of generality that $\{u, w\} \cap W'(x) = \{u\}$. Let $t \in V'$ be a vertex with degree 1 such that the path from u to t in G' contains v. By (W2) there exists $y \in X$ with $W'(y) = \{t\}$. Put $G'_{\sim v} = (V'_{\sim v}, E'_{\sim v})$. Then, by the definition of suppressing a vertex, we have

$$\begin{aligned} d(x, y) &= d_{(G',\omega')}(v, t) < d_{(G'_{\sim v},\omega'_{\sim v})}(u, t) \\ &= \min\{d_{(G'_{\sim v},\omega'_{\sim v})}(a, b) : a \in W'_{\sim v}(x) \text{ and } b \in W'_{\sim v}(y)\}, \end{aligned}$$

implying again that $(((V'_{\sim v}, E'_{\sim v}), \omega'_{\sim v}), W'_{\sim v})$ is not a subtree realization of d.

Next consider an edge $e \in E'$. Let $u, v \in V'$ be two vertices with degree 1 such that e lies on the path from u to v in G'. By (W2) there exist $x, y \in X$ with $W'(x) = \{u\}$ and $W'(y) = \{v\}$. Put $G'_{\sim e} = (V'_{\sim e}, E'_{\sim e})$, $\{u'\} = W'_{\sim e}(x)$ and $\{v'\} = W'_{\sim e}(y)$. Then, by the definition of contracting an edge, we have

$$d(x, y) = d_{(G',\omega')}(u, v) > d_{(G'_{\sim e},\omega'_{\sim e})}(u', v'),$$

implying that $(((V'_{\sim e}, E'_{\sim e}), \omega'_{\sim e}), W'_{\sim e})$ is not a subtree realization of d.

Thus, suppressing a vertex $v \in V'$ with degree 2 or contracting an edge $e \in E'$ does not yield a subtree realization of d, implying that $((G' = (V', E'), \omega'), W')$ is a minimal subtree realization of d.

(c) This follows immediately from (a), (b) and the fact that the minimal subtree realization of d is unique up to isomorphism by [25, Theorem 2]. □

8. Compatible extensions

To motivate the main result in this section, we first give some additional definitions. We call an ordered pair $(\mathcal{S}, \alpha)$ consisting of a split system $\mathcal{S}$ on X and a weighting map $\alpha : \mathcal{S} \to \mathbb{R}_{>0}$ a *weighted split system* on X and put $d_{(\mathcal{S},\alpha)} = \sum_{S \in \mathcal{S}} \alpha(S) \cdot d_S$. In addition, we say that a weighted split system $(\mathcal{U}, \beta)$ is a *compatible extension* of $(\mathcal{S}, \alpha)$ if

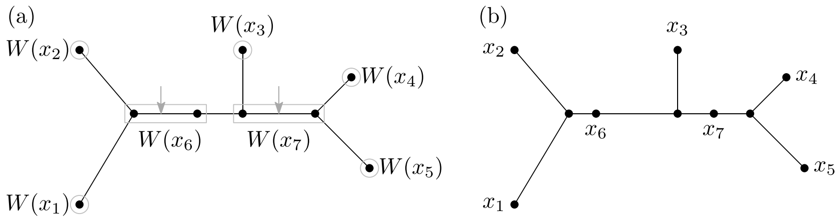

Fig. 4. (a) A geometric realization of the subtree representation $((B(\mathcal{S}), \omega_\alpha), W)$ of the distance $d_{(\mathcal{S},\alpha)}$ on $X = \{x_1, \ldots, x_7\}$ in Example 8.1. For each $x \in X$ the set $W(x)$ is enclosed by a gray line and the length of each edge e drawn in the figure corresponds to the weight $\omega_\alpha(e)$. The gray arrows indicate the points we select from $W(x_6)$ and $W(x_7)$, respectively, to obtain a geometric realization of an edge-weighted X-tree in (b). Unlabeled vertices of degree 2 are suppressed after selecting the points from W.

(CE1) every $U \in \mathcal{U}$ is a full split of X,
(CE2) the full splits in $\mathcal{U}$ are pairwise compatible, and
(CE3) $d_{(\mathcal{S},\alpha)} \leq d_{(\mathcal{U},\beta)}$.

A compatible extension $(\mathcal{U}, \beta)$ of $(\mathcal{S}, \alpha)$ is *minimal* if, for all compatible extensions $(\mathcal{U}', \beta')$ of $(\mathcal{S}, \alpha)$, $d_{(\mathcal{U}',\beta')} \leq d_{(\mathcal{U},\beta)}$ implies $(\mathcal{U}', \beta') = (\mathcal{U}, \beta)$. We let $\mathfrak{E}_{\min}(\mathcal{S}, \alpha)$ be the set of all possible minimal compatible extensions of $(\mathcal{S}, \alpha)$.

In the last section, we saw that in case d is a subtree distance on X with $d = d_{(\mathcal{S},\alpha)}$ for some non-empty set $\mathcal{S}$ of pairwise compatible partial splits of X and some weighting map $\alpha : \mathcal{S} \to \mathbb{R}_{>0}$, then the Buneman graph $((B(\mathcal{S}), \omega_\alpha), W)$ is, up to isomorphism, the unique subtree representation of d. Interestingly, in this situation we can also obtain an edge-weighted X-tree from $((B(\mathcal{S}), \omega_\alpha), W)$ by selecting, for each $x \in X$, some point in the set $W(x)$ (considered as a continuous object) and defining the edge-weights to be those naturally given by the subtree representation. In this section, we show that the set of weighted split systems that correspond to an edge-weighted X-tree that can be obtained in this way is equal to $\mathfrak{E}_{\min}(\mathcal{S}, \alpha)$ (Theorem 8.3). In particular, we can consider $((B(\mathcal{S}), \omega_\alpha), W)$ as being the configuration space of all those metrics that tightly bound $d_{(\mathcal{S},\alpha)}$ from above (Corollary 8.4).

Example 8.1. Consider the weighted split system $(\mathcal{S}, \alpha)$ on $X = \{x_1, \ldots, x_7\}$ consisting of the partial splits

$S_i = \{x_i\}|X - \{x_i\}$ for $1 \leq i \leq 5$

$S_6 = \{x_1, x_2\}|\{x_3, x_4, x_5, x_7\}$

$S_7 = \{x_1, x_2, x_6\}|\{x_3, x_4, x_5, x_7\}$

$S_8 = \{x_1, x_2, x_3, x_6\}|\{x_4, x_5\}$

and the following weighting:

i	1	2	3	4	5	6	7	8
$\alpha(S_i)$	1.5	1.3	1.0	0.8	1.2	1.0	0.7	1.1

The partial splits in $\mathcal{S}$ are pairwise compatible and the subtree representation $((B(\mathcal{S}), \omega_\alpha), W)$ of $d_{(\mathcal{S},\alpha)}$ is shown in Fig. 4(a). An X-tree resulting from a selection of a point in $W(x)$, for each $x \in X$, is shown in Fig. 4(b).

Before continuing with proving Theorem 8.3, we introduce some additional notation concerning realizations. A *geometric realization* of an edge-weighted tree $(G = (V, E), \omega)$ (in the plane) is a map $\rho : V \to \mathbb{R}^2$ such that

(GR1) for all $\{u, v\} \in E$ the Euclidean distance between the points $\rho(u)$ and $\rho(v)$ is $\omega(\{u, v\})$, and
(GR2) for any two distinct edges $\{u, v\}, \{u', v'\} \in E$ the straight line segment with endpoints $\rho(u)$ and $\rho(v)$ and the straight line segment with endpoints $\rho(u')$ and $\rho(v')$ are disjoint, except for when the two edges are incident to the same vertex $u = u'$ in which case the intersection of the straight line segments is the point $\rho(u) = \rho(u')$.

Every edge-weighted tree has some geometric realization and there exist algorithms for computing one (see e.g. [2]). We denote by $P_\rho(G, \omega)$ the set of those points $p \in \mathbb{R}^2$ for which there exists some edge $\{u, v\} \in E$ such that p lies on the straight line segment with endpoints $\rho(u)$ and $\rho(v)$. Moreover, for all $p, q \in P_\rho(G, \omega)$, we define $d_\rho(p, q)$ as the length of the shortest curve from p to q in $P_\rho(G, \omega)$. Note that, by definition, we have $d_{(G,\omega)}(u, v) = d_\rho(\rho(u), \rho(v))$ for all $u, v \in V$.

Now, let d be a subtree distance on X, $((G = (V, E), \omega), W)$ a subtree representation of d and ρ a geometric realization of (G, ω). For all $x \in X$ we denote by $P_\rho(x)$ the set of those points $p \in P_\rho(G, \omega)$ for which there exists some edge $\{u, v\} \in E$ with $u, v \in W(x)$ such that p lies on the straight line segment with endpoints $\rho(u)$ and $\rho(v)$. A map $\sigma : X \to P_\rho(G, \omega)$ with $\sigma(x) \in P_\rho(x)$ for all $x \in X$ is called a *selection* from the labeling map W with respect to the geometric realization ρ. To prove Theorem 8.3 we use the following technical lemma concerning selections.

Lemma 8.2. *Let d be a subtree distance on X, $((G = (V, E), \omega), W)$ a subtree representation of d and ρ a geometric realization of (G, ω). In addition, let D be a metric on X with $d \leq D$. Then there exists a selection σ from W with respect to ρ such that $d(x, y) \leq d_\rho(\sigma(x), \sigma(y)) \leq D(x, y)$ for all $x, y \in X$.*

Proof. Let L be the set of those $x \in X$ with $W(x) = \{v\}$ such that v has degree 1. In view of (W2), we have $L \neq \emptyset$. For all $x \in L$, put $p_x = \rho(v)$, where $\{v\} = W(x)$. Let $x_1, x_2, \ldots, x_k$ be an arbitrary ordering of the elements in $X - L$. Put $X_0 = L$ and, for all $i \in \{1, 2, \ldots, k\}$, put $X_i = X_{i-1} \cup \{x_i\}$.

We recursively define a map σ on X_i, for $0 \leq i \leq k$, and use induction on i to show that σ satisfies

(1) $\sigma(x) \in P_\rho(x)$ for all $x \in X_i$,
(2) $d(x, y) \leq d_\rho(\sigma(x), \sigma(y)) \leq D(x, y)$, for all $x, y \in X_i$, and
(3) $\min_{s \in P_\rho(x)} d_\rho(s, \sigma(y)) \leq D(x, y)$, for all $x \in X - X_i$ and all $y \in X_i$,

which will complete the proof since $X_k = X$. Note that (1) and (2) are the properties of the map σ stated in the lemma whereas (3) is an auxiliary property used in the induction to prove the lemma.

We begin by putting $\sigma(x) = p_x$ for all $x \in X_0$. Then we have $\sigma(x) \in P_\rho(x)$ for all $x \in X_0$, $d_\rho(\sigma(x), \sigma(y)) = d(x, y) \leq D(x, y)$, for all $x, y \in X_0$, and $\min_{s \in P_\rho(x)} d_\rho(s, \sigma(y)) = d(x, y) \leq D(x, y)$, for all $x \in X - X_0$ and all $y \in X_0$, establishing the base case of the induction.

Now, suppose that for some $i \geq 0$ we have defined a map σ on X_i with properties (1)-(3). We put

$$Q_y = \{q \in P_\rho(G, \omega) : d_\rho(\sigma(y), q) \leq D(y, x_{i+1})\},$$

for all $y \in X_i$,

$$R_y = \{r \in P_\rho(G, \omega) : \min_{s \in P_\rho(y)} d_\rho(s, r) \leq D(y, x_{i+1})\},$$

for all $y \in X - X_{i+1}$, and

$$I = P_\rho(x_{i+1}) \cap \left(\bigcap_{y \in X_i} Q_y \right) \cap \left(\bigcap_{y \in X - X_{i+1}} R_y \right).$$

Then, selecting any point in I as $\sigma(x_{i+1})$, it follows immediately that σ satisfies properties (1)-(3) also on X_{i+1}. Hence, it suffices to show that $I \neq \emptyset$.

By construction, the point sets $P_\rho(x_{i+1})$, Q_y, for all $y \in X_i$, and R_y, for all $y \in X - X_{i+1}$, are closed and connected subsets of $P_\rho(G, \omega)$. Therefore, by the Helly property (see e.g. [9, p. 21]), it suffices to show that these point sets have a pairwise non-empty intersection:

- $Q_y \cap Q_{y'} \neq \emptyset$, for all $y, y' \in X_i$, since
$$d_\rho(\sigma(y), \sigma(y')) \leq D(y, y') \leq D(y, x_{i+1}) + D(x_{i+1}, y'),$$
where the first inequality holds by induction and the second inequality holds by the fact that D is a metric.
- $R_y \cap R_{y'} \neq \emptyset$, for all $y, y' \in X - X_{i+1}$, since
$$d(y, y') = \min_{\substack{s \in P_\rho(y) \\ s' \in P_\rho(y')}} d_\rho(s, s') \leq D(y, y') \leq D(y, x_{i+1}) + D(x_{i+1}, y'),$$
where the equality holds by the definition of a subtree representation, the first inequality holds by the assumption $d \leq D$ and the second inequality holds by the fact that D is a metric.
- $Q_y \cap P_\rho(x_{i+1}) \neq \emptyset$, for all $y \in X_i$, since
$$\min_{s \in P_\rho(x_{i+1})} d_\rho(s, \sigma(y)) \leq D(x_{i+1}, y),$$
which holds by induction.
- $R_y \cap P_\rho(x_{i+1}) \neq \emptyset$, for all $y \in X - X_{i+1}$, since
$$d(y, x_{i+1}) = \min_{\substack{s \in P_\rho(y) \\ s' \in P_\rho(x_{i+1})}} d_\rho(s, s') \leq D(y, x_{i+1}),$$
where the equality holds by the definition of a subtree representation and the inequality holds by the assumption $d \leq D$.
- $Q_y \cap R_{y'} \neq \emptyset$, for all $y \in X_i$ and all $y' \in X - X_{i+1}$, since
$$\min_{s \in P_\rho(y')} d_\rho(s, \sigma(y)) \leq D(y, y'),$$
which holds by induction. □

As before, let d be a subtree distance on X, $((G = (V, E), \omega), W)$ a subtree representation of d and ρ a geometric realization of (G, ω). As informally explained in Example 8.1, every selection σ from W with respect to ρ yields an X-tree $(G_\sigma = (V_\sigma, E_\sigma), W_\sigma)$ together with an edge-weighting ω_σ by putting:

- $V_\sigma = \{\rho(v) : v \in V,\ \deg(v) \geq 3\} \cup \{\sigma(x) : x \in X\}$
- E_σ to be the set of those $\{p, q\}$ with $p, q \in V_\sigma$, $p \neq q$, for which the shortest curve from p to q in $P_\rho(G, \omega)$ does not contain any point in $V_\sigma - \{p, q\}$,
- $W_\sigma(x) = \{\sigma(x)\}$ for all $x \in X$, and
- $\omega_\sigma(\{p, q\})$ to be the Euclidean distance between p and q for all $\{p, q\} \in E_\sigma$.

In addition, we define the weighted split system $(\mathcal{U}_\sigma, \beta_\sigma)$ by putting $\mathcal{U}_\sigma = \mathcal{S}_{(G_\sigma, W_\sigma)}$ and $\beta_\sigma(S_e) = \omega_\sigma(e)$ for all $e \in E_\sigma$. As an immediate consequence of these definitions we have

$$d_\rho(\sigma(x), \sigma(y)) = d_{(G_\sigma, \omega_\sigma)}(\sigma(x), \sigma(y)) = d_{(\mathcal{U}_\sigma, \beta_\sigma)}(x, y) \tag{3}$$

for all $x, y \in X$.

Theorem 8.3. *Let $(\mathcal{S}, \alpha)$ be a weighted split system consisting of pairwise compatible partial splits of X and let $((B(\mathcal{S}), \omega_\alpha), W)$ be the unique subtree representation of $d_{(\mathcal{S},\alpha)}$. Then, for all geometric realizations ρ of $(B(\mathcal{S}), \omega_\alpha)$, we have*

$$\mathfrak{E}_{\min}(\mathcal{S}, \alpha) = \{(\mathcal{U}_\sigma, \beta_\sigma) : \sigma \text{ is a selection from } W \text{ with respect to } \rho\}.$$

Proof. Let ρ be a geometric realization of $(B(\mathcal{S}), \omega_\alpha)$.

First consider $(\mathcal{U}, \beta) \in \mathfrak{E}_{\min}(\mathcal{S}, \alpha)$. Then $d_{(\mathcal{U},\beta)}$ is a metric on X and, by the definition of a compatible extension, $d_{(\mathcal{S},\alpha)} \leq d_{(\mathcal{U},\beta)}$. So, by Lemma 8.2, there exists a selection σ from W with respect to ρ such that $d_{(\mathcal{S},\alpha)}(x, y) \leq d_\rho(\sigma(x), \sigma(y)) \leq d_{(\mathcal{U},\beta)}(x, y)$ for all $x, y \in X$. Hence, by Eq. (3), we have $d_{(\mathcal{U}_\sigma,\beta_\sigma)} \leq d_{(\mathcal{U},\beta)}$. Thus, since $(\mathcal{U}, \beta)$ is minimal, $(\mathcal{U}, \beta) = (\mathcal{U}_\sigma, \beta_\sigma)$. This establishes

$$\mathfrak{E}_{\min}(\mathcal{S}, \alpha) \subseteq \{(\mathcal{U}_\sigma, \beta_\sigma) : \sigma \text{ is a selection from } W \text{ with respect to } \rho\}.$$

To establish the other inclusion, consider a selection σ from W with respect to ρ. Then, by the construction of the X-tree (G_σ, W_σ) and the edge-weighting ω_σ, we have

$$\begin{aligned} d_{(\mathcal{S},\alpha)}(x, y) &= \min\{d_\rho(p, q) : p \in P_\rho(x),\ y \in P_\rho(y)\} \leq d_\rho(\sigma(x), \sigma(y)) \\ &= d_{(\mathcal{U}_\sigma,\beta_\sigma)}(x, y) \end{aligned}$$

for all $x, y \in X$. Hence, $(\mathcal{U}_\sigma, \beta_\sigma)$ is a compatible extension of $(\mathcal{S}, \alpha)$. To show that $(\mathcal{U}_\sigma, \beta_\sigma) \in \mathfrak{E}_{\min}(\mathcal{S}, \alpha)$, consider a compatible extension $(\mathcal{U}, \beta)$ of $(\mathcal{S}, \alpha)$ such that $d_{(\mathcal{U},\beta)} \leq d_{(\mathcal{U}_\sigma,\beta_\sigma)}$. Then, if $x, y \in X$ are such that $|W(x)| = |W(y)| = 1$, we have $d_{(\mathcal{S},\alpha)}(x, y) = d_\rho(\sigma(x), \sigma(y)) = d_{(\mathcal{U}_\sigma,\beta_\sigma)}(x, y) = d_{(\mathcal{U},\beta)}(x, y)$. Hence, it remains to consider $x', y' \in X$ with $|W(x')| \geq 2$ or $|W(y')| \geq 2$. In view of (W2), there exist $x, y \in X$ with $|W(x)| = |W(y)| = 1$ such that $\sigma(x')$ and $\sigma(y')$ lie on the path from $\sigma(x)$ to $\sigma(y)$ in G_σ. Then we have

$$\begin{aligned} d_{(\mathcal{S},\alpha)}(x, y) &= d_{(\mathcal{U}_\sigma,\beta_\sigma)}(x, y) = d_\rho(\sigma(x), \sigma(y)) \\ &= d_\rho(\sigma(x), \sigma(x')) + d_\rho(\sigma(x'), \sigma(y')) + d_\rho(\sigma(y'), \sigma(y)) \\ &= d_{(\mathcal{U}_\sigma,\beta_\sigma)}(x, x') + d_{(\mathcal{U}_\sigma,\beta_\sigma)}(x', y') + d_{(\mathcal{U}_\sigma,\beta_\sigma)}(y', y) \\ &\geq d_{(\mathcal{U},\beta)}(x, x') + d_{(\mathcal{U},\beta)}(x', y') + d_{(\mathcal{U},\beta)}(y', y) \\ &\geq d_{(\mathcal{U},\beta)}(x, y) = d_{(\mathcal{S},\alpha)}(x, y), \end{aligned}$$

where the first line holds since $|W(x)| = |W(y)| = 1$, the second line holds since $\sigma(x')$ and $\sigma(y')$ lie on the path from $\sigma(x)$ to $\sigma(y)$ in G_σ, the third line holds by Eq. (3), the fourth line holds by our assumption that $d_{(\mathcal{U},\beta)} \leq d_{(\mathcal{U}_\sigma,\beta_\sigma)}$ and the last line holds since $d_{(\mathcal{U},\beta)}$ is a metric. It follows that also $d_{(\mathcal{U}_\sigma,\beta_\sigma)}(x', y') = d_{(\mathcal{U},\beta)}(x', y')$ and, thus, the metrics $d_{(\mathcal{U}_\sigma,\beta_\sigma)}$ and $d_{(\mathcal{U},\beta)}$ on X coincide. Since the splits in $\mathcal{U}_\sigma$ and also in $\mathcal{U}$ are pairwise compatible, this implies, as pointed out in the discussion of Property (D2) in Section 7, that $(\mathcal{U}, \beta) = (\mathcal{U}_\sigma, \beta_\sigma)$, as required. □

We say that a metric D on X is a *minimal extension* of a distance d on X if $d \leq D$ and, for all metrics D' on X, $d \leq D' \leq D$ implies $D' = D$.

Corollary 8.4. *Let d be a subtree distance on X, $((G = (V, E), \omega), W)$ a subtree representation of d and ρ a geometric realization of (G, ω). Then a metric D on X is a minimal extension of d if and only if $D = d_{(G_\sigma,\omega_\sigma)}$ for some selection σ from W with respect to ρ.*

Proof. First assume that D is a minimal extension of d. Then, by Lemma 8.2, there exists a selection σ from W with respect to ρ such that $d \leq d_{(G_\sigma,\omega_\sigma)} \leq D$. By the minimality of D, this implies $D = d_{(G_\sigma,\omega_\sigma)}$, as required.

Next assume that σ is a selection from W with respect to ρ. Consider the metric $D = d_{(G_\sigma,\omega_\sigma)}$. Let D' be a metric on X with $d \leq D' \leq D$. Then, again by Lemma 8.2, there exists a selection σ' from W with respect to ρ such that $d \leq d_{(G_{\sigma'},\omega_{\sigma'})} \leq D'$. Hence we have

$$d \leq d_{(G_{\sigma'},\omega_{\sigma'})} = d_{(\mathcal{U}_{\sigma'},\beta_{\sigma'})} \leq D' \leq D = d_{(\mathcal{U}_\sigma,\beta_\sigma)} = d_{(G_\sigma,\omega_\sigma)}.$$

By Theorem 8.3, we must have $d_{(\mathcal{U}_{\sigma'},\beta_{\sigma'})} = d_{(\mathcal{U}_\sigma,\beta_\sigma)}$, implying $D' = D$, as required. □

9. Discussion

We conclude by giving some potential directions for future research. First, if $\mathcal{S}$ is a split system consisting of full splits of X then after finitely many iterations all vertices of the Buneman graph $B(\mathcal{S})$ are obtained (see e.g [14, Sec. 4.3]) by repeatedly forming the median starting with the vertices in $\{\phi_x : x \in X\}$, where ϕ_x is the $\mathcal{S}$-map defined in the introduction by putting $\phi_x(S) = S(x)$. Can the vertices of the Buneman graph $B(\mathcal{S})$ for a split system consisting of partial splits also be obtained in a similar way, possibly starting with the vertices in $\bigcup_{x\in X} W(x)$?

In another direction, it might be of interest to investigate how the Buneman graph of a split system is related to *split system closures*. There are various rules which can be applied to some collection of partial splits which aim to produce a new collection of full splits. Such rules go back to work of Meacham who introduced the "Z-rule" [26]. The repeated application of such rules to a split system eventually leads to a split system for which no new splits are generated after applying the rules. This final split system is known as the *closure* of the original split system, and such closures have been investigated in, for example, [21,29]. It would be interesting to investigate how the Buneman graph of a split system is related to the Buneman graph of different closures of the system. As a potential application, it might be worth looking into using the Buneman graph of a split system for constructing supernetworks from collections of partial phylogenetic trees (cf. [21] where Z-rule closures are used for this purpose).

As we have mentioned above, our definition of the Buneman graph for a set of partial splits was motivated by results in [18] on subtree distances. In this direction, to any metric on X one can associate the *Buneman complex* as well as the *tight span*, which are both polytopal complexes that in certain circumstances share key structural properties [15]. The tight span has been generalized to arbitrary distances on X (see [18]). In particular, in [18, Theorem 2.3] it is shown that a distance d has a subtree realization if and only if the 1-skeleton of the tight span of d is a tree. Can the Buneman complex also be generalized to arbitrary distances d and, if so, how is it related to the tight span of d?

As mentioned in the introduction, there is an extensive theory of median graphs. It would be interesting to investigate potential extensions of our results within the theory of these graphs. For example, in [4] a generalization of median graphs called *lopsided* sets is considered, and in [5] a generalization of median graphs to an infinite case is presented, and it could be worth considering how our results may have relevance within these settings. Other areas with potential connections or applications include cubical complexes and median algebras — see e.g. [3].

Finally, there is a rich theory concerned with embedding metrics into l_p-spaces (see e.g. [12]). For $p = 1$, a metric that can be embedded into some l_1-space is called an l_1-*metric*. It is well-known that a metric d on X is an l_1-metric if and only if there exists a weighted split system $(\mathcal{S}, \alpha)$ on X consisting of full splits with $d = d_{(\mathcal{S},\alpha)}$ (see e.g. [12, Chapter 11]). In contrast, every distance d on X can be written as

$$d = \sum_{\substack{a,b\in X\\ d(a,b)>0}} d(a,b)\cdot d_{\{a\}|\{b\}},$$

with the empty sum evaluating to 0. Thus, whilst weighted split systems that consist of full splits of X yield a proper subclass of all metrics on X (the class of l_1-metrics), weighted split systems that consist of partial splits of X without any restrictions yield the class of *all* distances on X. Even so, it might still be interesting to explore embeddings for distances corresponding to restricted classes of split systems (e.g. those arising in split decomposition [19]), or studying embeddings that represent elements in X by subsets rather than single points.

Acknowledgments

KTH, VM and DB thank The Royal Society, United Kingdom for its support through their International Exchanges Scheme. KTH and VM also thank the Department of Mathematics and Statistics, University of Otago, New Zealand for hosting them. All authors would like to thank two anonymous reviewers for their helpful comments on an earlier version of this paper.

Appendix. List of $\mathcal{S}$-maps in Example 2.3

The $\mathcal{S}$-maps $\phi_1, \phi_2, \ldots, \phi_{10}$ are listed below in this order from top to bottom. Each $\mathcal{S}$-map ϕ_i, $1 \le i \le 10$, is given as the 8-tuple $(\phi_i(S_1), \ldots, \phi_i(S_8))$.

$(\{x_1\}, X-\{x_2\}, X-\{x_3\}, X-\{x_4\}, X-\{x_5\}, \{x_1,x_2\}, \{x_1,x_2,x_5,x_6\}, \{x_1,x_2,x_3,x_6\})$

$(X-\{x_1\}, \{x_2\}, X-\{x_3\}, X-\{x_4\}, X-\{x_5\}, \{x_1,x_2\}, \{x_1,x_2,x_5,x_6\}, \{x_1,x_2,x_3,x_6\})$

$(X-\{x_1\}, X-\{x_2\}, X-\{x_3\}, X-\{x_4\}, X-\{x_5\}, \{x_1,x_2\}, \{x_1,x_2,x_5,x_6\}, \{x_1,x_2,x_3,x_6\})$

$(X-\{x_1\}, X-\{x_2\}, X-\{x_3\}, X-\{x_4\}, X-\{x_5\}, \{x_3,x_4,x_5,x_7\}, \{x_1,x_2,x_5,x_6\}, \{x_1,x_2,x_3,x_6\})$

$(X-\{x_1\}, X-\{x_2\}, X-\{x_3\}, X-\{x_4\}, X-\{x_5\}, \{x_3,x_4,x_5,x_7\}, \{x_3,x_4\}, \{x_1,x_2,x_3,x_6\})$

$(X-\{x_1\}, X-\{x_2\}, \{x_3\}, X-\{x_4\}, X-\{x_5\}, \{x_3,x_4,x_5,x_7\}, \{x_3,x_4\}, \{x_1,x_2,x_3,x_6\})$

$(X-\{x_1\}, X-\{x_2\}, X-\{x_3\}, X-\{x_4\}, X-\{x_5\}, \{x_3,x_4,x_5,x_7\}, \{x_1,x_2,x_5,x_6\}, \{x_4,x_5\})$

$(X-\{x_1\}, X-\{x_2\}, X-\{x_3\}, X-\{x_4\}, \{x_5\}, \{x_3,x_4,x_5,x_7\}, \{x_1,x_2,x_5,x_6\}, \{x_4,x_5\})$

$(X-\{x_1\}, X-\{x_2\}, X-\{x_3\}, X-\{x_4\}, X-\{x_5\}, \{x_3,x_4,x_5,x_7\}, \{x_3,x_4\}, \{x_4,x_5\})$

$(X-\{x_1\}, X-\{x_2\}, X-\{x_3\}, \{x_4\}, X-\{x_5\}, \{x_3,x_4,x_5,x_7\}, \{x_3,x_4\}, \{x_4,x_5\})$

Data availability

No data was used for the research described in the article.

References

[1] K. Ando, K. Sato, An algorithm for finding a representation of a subtree distance, J. Comb. Optim. 36 (2018) 742–762.
[2] C. Bachmaier, U. Brandes, B. Schlieper, Drawing phylogenetic trees, in: International Symposium on Algorithms and Computation, Springer, 2005, pp. 1110–1121.
[3] H.-J. Bandelt, V. Chepoi, Metric graph theory and geometry: a survey, Contemp. Math. 453 (2008) 49–86.
[4] H.-J. Bandelt, V. Chepoi, A. Dress, J. Koolen, Combinatorics of lopsided sets, European J. Combin. 27 (5) (2006) 669–689.
[5] H.-J. Bandelt, V. Chepoi, D. Eppstein, Combinatorics and geometry of finite and infinite squaregraphs, SIAM J. Discrete Math. 24 (4) (2010) 1399–1440.
[6] H.-J. Bandelt, P. Forster, B. Sykes, M. Richards, Mitochondrial portraits of human populations using median networks, Genetics 141 (2) (1995) 743–753.
[7] J.-P. Barthélemy, From copair hypergraphs to median graphs with latent vertices, Discrete Math. 76 (1) (1989) 9–28.
[8] L. Bénéteau, J. Chalopin, V. Chepoi, Y. Vaxès, Medians in median graphs and their cube complexes in linear time, J. Comput. System Sci. 126 (2022) 80–105.
[9] C. Berge, Hypergraphs: Combinatorics of Finite Sets, Elsevier, 1984.
[10] P. Buneman, The recovery of trees from measures of dissimilarity, in: F. Hodson, et al. (Eds.), Mathematics in the Archaeological and Historical Sciences, Edinburgh University Press, 1971, pp. 387–395.
[11] V. Chepoi, Separation of two convex sets in convexity structures, J. Geometry 50 (1994) 30–51.
[12] M. Deza, M. Laurent, Geometry of Cuts and Metrics, Springer, 1997.
[13] A. Dress, M. Hendy, K.T. Huber, V. Moulton, On the number of vertices and edges of the Buneman graph, Ann. Comb. 1 (1997) 329–337.
[14] A. Dress, K.T. Huber, J. Koolen, V. Moulton, A. Spillner, Basic Phylogenetic Combinatorics, Cambridge University Press, 2012.
[15] A. Dress, K.T. Huber, V. Moulton, An explicit computation of the injective hull of certain finite metric spaces in terms of their associated Buneman complex, Adv. Math. 168 (2002) 1–28.
[16] A. Dress, R. Scharlau, Gated sets in metric spaces, Aequationes Math. 34 (1987) 112–120.
[17] R. Hammack, W. Imrich, S. Klavžar, Handbook of Product Graphs, CRC Press, 2011.
[18] H. Hirai, Characterization of the distance between subtrees of a tree by the associated tight span, Ann. Comb. 10 (2006) 111–128.
[19] H. Hirai, A geometric study of the split decomposition, Discrete Comput. Geom. 36 (2006) 331–361.
[20] K.T. Huber, V. Moulton, C. Semple, T. Wu, Representing partitions on trees, SIAM J. Discrete Math. 28 (3) (2014) 1152–1172.
[21] D. Huson, T. Dezulian, T. Klopper, M. Steel, Phylogenetic super-networks from partial trees, IEEE/ACM Trans. Comput. Biol. Bioinformat. 1 (2004) 151–158.
[22] W. Imrich, S. Klavžar, A convexity lemma and expansion procedures for bipartite graphs, European J. Combin. 19 (1998) 677–685.
[23] S. Klavžar, H.M. Mulder, Partial cubes and crossing graphs, SIAM J. Discrete Math. 15 (2002) 235–251.
[24] M. Kovše, Complexity of phylogenetic networks: counting cubes in median graphs and related problems, in: Analysis of Complex Networks: From Biology To Linguistics, 2009, pp. 323–350.
[25] T. Maehara, K. Ando, Optimal algorithm for finding representation of subtree distance, IEICE Trans. Fundam. Electron. Commun. Comput. Sci. 105 (2022) 1203–1210.
[26] C. Meacham, Theoretical and computational considerations of the compatibility of qualitative taxonomic characters, Numerical Taxonomy (1983) 304–314.
[27] H.M. Mulder, The Interval Function of a Graph. Mathematical Centre Tracts, Mathematisch Centrum, Amsterdam, 1980, p. 132.
[28] H.M. Mulder, Median graphs: a structure theory, in: Advances in Interdisciplinary Applied Discrete Mathematics, World Scientific, 2011, pp. 93–125.
[29] C. Semple, M. Steel, Tree reconstruction via a closure operation on partial splits, in: Computational Biology: First International Conference on Biology, Informatics, and Mathematics, 2000, Selected Papers, Springer, 2001, pp. 126–134.
[30] C. Semple, M. Steel, Phylogenetics, Oxford University Press, 2003.
[31] M. van de Vel, Binary convexities and distributive lattices, Proc. Lond. Math. Soc. (3) 3 (1) (1984) 1–33.