Proceedings of the Edinburgh Mathematical Society (2026) **69**, 304–338
doi:10.1017/S0013091525100874

CAN YOU TELL A TREE FROM ITS BRANCHES?

CHENGYANG QIAN, YAOKUN WU AND MIN XIE

School of Mathematical Sciences and MOE-LSC, Shanghai Jiao Tong University, Shanghai, China

Corresponding author: Yaokun Wu, email: ykwu@sjtu.edu.cn

(Received 6 February 2024)

Abstract For each set X, an X-split is a partition of X into two parts. For each X-split $\mathbf{S}$ and each subset $Y \subseteq X$, the restriction of $\mathbf{S}$ on Y is the Y-split whose parts are obtained by intersecting the parts of $\mathbf{S}$ with Y. For a graph G with vertex set V, the G-coboundary size of a V-split $\mathbf{S}$ is the number of edges in G having non-empty intersections with both parts of $\mathbf{S}$. Let T be a tree without degree-two vertices, and let V and L denote its vertex set and leaf set, respectively. For each positive integer k, a k-split on T is an L-split that is the restriction of a V-split with T-coboundary size k, while a score-k split on T is a k-split on T that is not any k'-split for any integer $k' < k$. Buneman's split equivalence theorem states that the tree T is entirely encoded by its system of score-1 splits. We identify the unique exceptional case in which the tree T is not determined by its score-2 split system. To explore how our work can be extended to more general tree isomorphism problems, we propose several conjectures and open problems related to set systems and generalized Buneman graphs.

Keywords: 2-special configuration; (k, 2)-hypergraph; smallest 2-extension

2010 Mathematics subject classification: Primary: 05C05; Secondary: 05C70; 05C85; 92D15

1. Going beyond Buneman Theorem

Trees and associated tree-like objects are very important in mathematics and its applications. There are a variety of invariants which are shown or conjectured to encode a tree structure up to different kinds of equivalences [2, 3, 5, 13, 17, 23, 28, 48, 31, 55, 56, 58, 61]. Following Buneman [12], in this paper and its sequel [62–64], we will be considering how a tree is determined by its split system invariants. After cutting off some tree edges, the leaves of a tree may branch into different groups according to some natural rules. Can we determine the shape of a tree if we have on hand a 'good' set of such branchings? To provide a reasonable answer, we must describe the scenario more precisely below.

For any positive integer k, we write $[k]$ for the set $\{1, \ldots, k\}$. Let X be a set. A *partition* π of X is a family of disjoint subsets $X_1, \ldots, X_m$ of X whose union is X. We call the sets $X_1, \ldots, X_m$ the *parts* of π, and we often write π directly as $X_1 \mid \cdots \mid X_m$. Since there is no ordering of the parts in a partition, $Y \mid Z$ is the same as $Z \mid Y$. A partition of X into m non-empty parts will be termed an *m-partition* of X. For any partition π of X and any $W \subseteq X$, the *restriction* of π on W, denoted by $\pi|_W$, is a partition of W

© The Author(s), 2025. Published by Cambridge University Press on behalf of The Edinburgh Mathematical Society.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

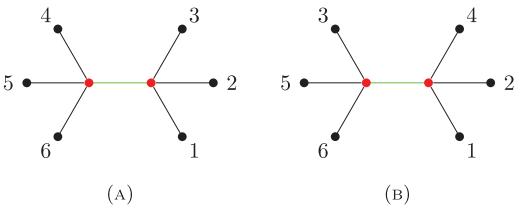


Figure 1. Phylogenetic [6]-trees T and T' are isomorphic but not [6]-isomorphic: (A) T and (B) T'

whose parts are the intersections of W with the parts of π. We also write $\pi' \subseteq \pi$ to mean that $\pi' = \pi|_W$ for some set W. An *X-split*, or a *split on X*, is simply a partition of X into two parts. A split of X is *proper* if its two parts are both non-empty; otherwise, it is *improper*. A family of X-splits is known as an *X-split system* or a *split system on X*. An X-split system $\mathcal{S}$ is *isomorphic* to an X'-split system $\mathcal{S}'$, denoted by $\mathcal{S} \cong \mathcal{S}'$, if there exists a bijection f from X to X' such that $\mathcal{S}' = \{f(Y) \mid f(Z) : Y \mid Z \in \mathcal{S}\}$.

A *graph* G is a pair of sets V_G and E_G with $\mathrm{E}_G \subseteq \binom{\mathrm{V}_G}{2}$; we call V_G its set of *vertices* and E_G its set of *edges*. Each edge $\{u, v\}$ of a graph is often written as uv. An *isomorphism* from a graph G to a graph G' is a bijection f from V_G to $\mathrm{V}_{G'}$ such that $uv \in \mathrm{E}_G$ exactly when $f(u)f(v) \in \mathrm{E}_{G'}$. We say that G and G' are *isomorphic*, designated by $G \cong G'$, when there is an isomorphism from G to G'. With a slight abuse of notation, we may sometimes write $G = G'$ to mean that $G \cong G'$ and write $G \in \mathcal{G}$ to mean that G is isomorphic to one element of $\mathcal{G}$.

Let G be a graph. For each $uv \in \mathrm{E}_G$, we call u and v the *endpoints* of uv and say that u and v are *adjacent* in G. For each $v \in \mathrm{V}_G$, the set of all vertices adjacent to v in G is denoted by $\mathrm{N}_G(v)$ and is named the *neighbors* of v in G; the size of $\mathrm{N}_G(v)$, denoted by $\deg_G(v)$, is the *degree* of v in G. For any non-negative integer ℓ, a *path* in G of *length* ℓ is a sequence of distinct vertices $v_0v_1 \cdots v_\ell$ such that $v_0v_1, \ldots, v_{\ell-1}v_\ell \in \mathrm{E}_G$; we do not distinguish this path from its reversal $v_\ell v_{\ell-1} \cdots v_0$, and we refer to it as a *path between* v_0 and v_ℓ. We call $v \in \mathrm{V}_G$ a *leaf* of G provided $\deg_G(v) = 1$, and we reserve the notation L_G for the set of all leaves of G.

A graph is called a *tree* if there exists a unique path between each pair of its vertices. For a tree T, any element of $\mathrm{V}_T \setminus \mathrm{L}_T$ is called an *inner vertex* of T. For any tree T and $u, v \in \mathrm{V}_T$, we write $T_{u,v}$ for the set of all edges $ww' \in \mathrm{E}_T$ such that w and w' are two adjacent vertices in the unique path between u and v in T. A tree with at least one edge and without any degree-two vertices is called a *phylogenetic tree*. A *binary tree* is a phylogenetic tree whose inner vertices are all of degree three. In this paper, we will always assume that a graph has a non-empty and finite vertex set. By this assumption, a graph is a tree if and only if we can remove leaves repeatedly to reach the one-vertex graph, and whenever a phylogenetic tree T is mentioned we should understand that L_T is a finite set of size at least two.

Let X be a finite set with $|X| \geqslant 2$. A tree T is an *X-tree* if $\{v \in \mathrm{V}_T : \deg_T(v) \leq 2\} = X$. Note that, if an X-tree T is phylogenetic, then $\mathrm{L}_T = X$. Two X-trees T and T' are *X-isomorphic* if there exists a graph isomorphism from T to T' whose restriction on X is the identity map; we write $T \cong_X T'$ to mean that they are X-isomorphic; see Fig. 1.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

Let T be a tree. For any proper L_T-split $Y \mid Z$, the *parsimony score* of $Y \mid Z$ on T, denoted by $\mathrm{PS}(T, Y \mid Z)$, is the minimum positive integer k such that there exists $F \in \binom{\mathrm{E}_T}{k}$ which intersects with $T_{y,z}$ for every $y \in Y$ and $z \in Z$ [36, 38, 66]. The parsimony score of $\mathrm{L}_T \mid \emptyset$ is defined to be 0. For any non-negative integer k, we write $\mathcal{P}_{T,k}$ for the set of all L_T-splits with parsimony score k on T, and name it the *score-k split system on T*. It is clear that $\mathcal{P}_{T,k}$, $k \geqslant 0$, are pairwise disjoint. Note that $\mathcal{P}_{T,0}$ only contains $\mathrm{L}_T \mid \emptyset$ and $\mathcal{P}_{T,1}$ is the usual split system considered in phylogenetic combinatorics, namely the one induced by deleting one edge of T.

A split system $\mathcal{S}$ is *compatible* provided $\prod_{i,j\in[2]} |Y_i \cap Z_j| = 0$ holds for all $Y_1 \mid Y_2, Z_1 \mid Z_2 \in \mathcal{S}$. A split is *trivial* if one of its parts is a singleton set. Here is the famous split equivalence theorem about score-1 split systems of phylogenetic trees. Note that what Buneman really proved are Theorem 1.1 (1) and (2); but it is not hard to derive Theorem 1.1 (3) as an easy consequence of Theorem 1.1 (2).

Theorem 1.1. (Buneman Theorem [*12*]). *Let X be a finite set.*

(1) *An X-split system $\mathcal{S}$ is the score-1 split system on a phylogenetic tree if and only if $\mathcal{S}$ is compatible, does not contain the improper X-split and contains all trivial X-splits.*
(2) *Two phylogenetic X-trees T and T' are X-isomorphic if and only if $\mathcal{P}_{T,1}$ and $\mathcal{P}_{T',1}$ are equal.*
(3) *Two phylogenetic trees T and T' are isomorphic if and only if $\mathcal{P}_{T,1}$ and $\mathcal{P}_{T',1}$ are isomorphic.*

Example 1.2. Let $\mathrm{K}_{1,3}$, $\mathtt{A}_i$ and $\widetilde{\mathtt{A}}_i$, $i \in \{-\infty, 0\}$, be the trees depicted in Fig. 2, which are pairwise non-isomorphic. Among these five trees, only $\mathtt{A}_{-\infty}$ is not a phylogenetic tree. Note that a tree T has $\mathcal{P}_{T,2} = \emptyset$ if and only if $T \in \{\mathtt{A}_{-\infty}, \widetilde{\mathtt{A}}_{-\infty}, \mathrm{K}_{1,3}\}$. One can also verify that $\mathcal{P}_{\mathtt{A}_0,2} = \mathcal{P}_{\widetilde{\mathtt{A}}_0,2} = \{Y \mid ([6] \setminus Y) : \ Y \in \binom{[6]}{2}\}$; see Lemma *5.1* if you wish to avoid performing the routine calculations yourself. A reader curious about the naming of the graphs in Fig. 2 may refer to § *6*.

There is a significant body of work aimed at deducing results in the same spirit as the Buneman Theorem (Theorem 1.1) [15, 36, 37, 40, 71, 44, 69, 70]. In this paper, we will develop additional variants of Theorem 1.1 (2) and (3); see Theorem 1.3. Corresponding possible variants of Theorem 1.1 (1) are certainly of interest, but so far we have no results to report in that direction.

Theorem 1.3. *Let X be a finite set with $|X| \geqslant 2$, and let T and T' be two phylogenetic X-trees.*

(1) *The tree T is isomorphic to $\mathtt{A}_0$ or $\widetilde{\mathtt{A}}_0$ if and only if $|X| = 6$ and $\mathcal{P}_{T,2} = \{Y \mid (X \setminus Y) : Y \in \binom{X}{2}\}$.*
(2) *Assume that $\{T, T'\} \setminus \{\mathtt{A}_0, \widetilde{\mathtt{A}}_0\} \neq \emptyset$. Then, $T \cong_X T'$ if and only if $\mathcal{P}_{T,2} = \mathcal{P}_{T',2}$.*
(3) *Assume that $\{T, T'\} \setminus \{\mathtt{A}_0, \widetilde{\mathtt{A}}_0\} \neq \emptyset$. Then, $T \cong T'$ if and only if $\mathcal{P}_{T,2} \cong \mathcal{P}_{T',2}$.*

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

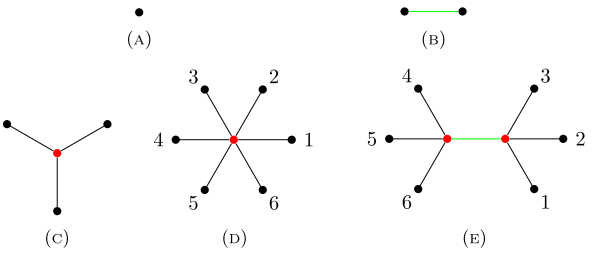


Figure 2. Two groups of trees with isomorphic score-2 split systems. A) $\mathtt{A}_{-\infty}$ B) $\widetilde{\mathtt{A}}_{-\infty}$ C) $\mathrm{K}_{1,3}$ D) $\mathtt{A}_0$ E) $\widetilde{\mathtt{A}}_0$

Clearly, if two phylogenetic X-trees are both isomorphic to $\mathtt{A}_0$, then they must be X-isomorphic. However, this does not hold for $\widetilde{\mathtt{A}}_0$, as can be seen from Fig. 1. This difference between $\widetilde{\mathtt{A}}_0$ and $\mathtt{A}_0$ is reflected in the following immediate consequence of Theorem 1.3 (2).

Corollary 1.4. *Suppose phylogenetic X-trees T and T' are isomorphic but not X-isomorphic. If $\mathcal{P}_{T,2} = \mathcal{P}_{T',2}$, then both T and T' are isomorphic to $\widetilde{\mathtt{A}}_0$.*

For any two binary X-trees T and T', Fischer [36, Proposition 1] observed that $\mathcal{P}_{T,2} = \mathcal{P}_{T',2}$ if and only if $T \cong_X T'$. To address general phylogenetic trees, our proof of Theorem 1.3 diverges significantly from her approach in [36].

Let G be a graph. For any $A \subseteq \mathrm{V}_G$, the *subgraph of G induced by A*, designated by $G[A]$, is the graph with vertex set A and edge set $\mathrm{E}_G \cap \binom{A}{2}$; we use the notation $G - A$ to stand for the subgraph of G induced by $\mathrm{V}_G \setminus A$. For any $F \subseteq \mathrm{E}_G$, $G - F$ represents the graph with vertex set V_G and edge set $\mathrm{E}_G \setminus F$. For a singleton set $\{a\}$ contained in either V_G or E_G, we write $G - a$ as shorthand for $G - \{a\}$. A tree T is a *subtree* of G if $\mathrm{V}_T \subseteq \mathrm{V}_G$ and $\mathrm{E}_T \subseteq \mathrm{E}_G$. For any $u, v \in \mathrm{V}_G$, we say that u and v are *reachable* in G provided there is a path between them in G. This *reachability relation* on V_G is an equivalence relation; a *connected component* of G is defined as an equivalence class in this equivalence relation. For any $v \in \mathrm{V}_G$, we use $\mathrm{C}_G[v]$ to represent the connected component of G containing v. Let T be a tree and take $v \in \mathrm{V}_T \setminus \mathrm{L}_T$ with $k \doteq \deg_T(v)$. The k-partition $X_1 \mid \cdots \mid X_k$ of L_T, where $\{X_1, \ldots, X_k\} = \{\mathrm{L}_T \cap \mathrm{C}_{T-v}[u] : u \in \mathrm{N}_T(v)\}$, is denoted by $\pi_{T,v}$ and called a *one-vertex partition* of T.

Our proof of Theorem 1.3 is largely algorithmic. For any given finite set X with $|X| \geqslant 2$ and any X-split system $\mathcal{S}$, our main algorithm, Algorithm 2, as presented in § 5, outputs all phylogenetic X-trees, up to X-isomorphism, whose score-2 split system is exactly $\mathcal{S}$. Our Algorithm 2 is based on Algorithm 1, which is introduced in § 3. For any given finite set X, integer $k \geqslant 4$, X-split system $\mathcal{S}$, and k-partition π of X, Algorithm 1 outputs all phylogenetic X-trees T having π as a one-vertex partition and satisfying $\mathcal{S} = \mathcal{P}_{T,2}$. In other words, when a potential phylogenetic tree T with $\mathcal{S} = \mathcal{P}_{T,2}$ contains a vertex of degree at least four, Algorithm 1 reconstructs this tree T; see Lemma 3.3. Algorithm 2 essentially consists of two parts: identifying one of three possible markers and executing the three corresponding reduction steps. The so-called markers are three special structures that need to be recognized from the given split system, namely a potential vertex of degree at least five, a potential cherry, and a potential claw; see § 2.1

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

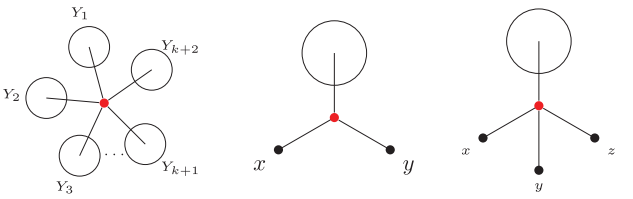


Figure 3. Three markers. A) $Y_1 \mid \cdots \mid Y_{k+2} = \pi_{T,v}$ for some $v \in \mathrm{V}_T$ with $\deg_T(v) \geqslant 5$. B) $\{x, y\}$ is a cherry. C) $\{x, y, z\}$ is a claw.

for precise definitions of cherry and claw, and see Fig. 3 for a depiction of the three markers. Assuming the existence of a phylogenetic tree T with $\mathcal{P}_{T,2} = \mathcal{S}$, a sketch of the corresponding reduction steps of Algorithm 2 is roughly as follows:

- If there exists a vertex of degree at least 5 in T, then reconstruct T using Algorithm 1.
- If there is a cherry in T, then we can perform a reduction by deleting cherries; see Lemma 3.1.
- If there is no vertex of degree at least 5 and at most one cherry in T, then T must contain a claw. Consequently, we can reconstruct T using Algorithm 1.

Note that detecting cherry information is straightforward (Corollary 2.7); however, the treatment of the other two types of markers is more involved. In § 4.1, we begin with Definitions 4.1 and 4.2, which introduce $(k, 2)$-sets with respect to a split system and 2-special configurations in a tree, respectively. The main result there, Lemma 4.4, provides a correspondence between $(k, 2)$-sets and 2-special configurations, helping us extract graph structures of the phylogenetic tree from split system information. In § 4.2, we demonstrate how to identify from $\mathcal{P}_{T,2}$ a potential vertex v in T with $\deg_T(v) \geqslant 5$ (Lemma 4.13), and, in the event that $|\mathrm{L}_T| \geqslant 7$ and $\max_{v \in \mathrm{V}_T} \deg_T(v) \leq 4$, how to detect a claw (Lemma 4.15).

We organize the proof details of Theorem 1.3 in the next four sections. Section 2 is dedicated to technical preparations. The main results of § 3 are Lemma 3.1 and Algorithm 1, where Lemma 3.1 describes the effect of the cherry-deletion operation on the score-2 split system, and Algorithm 1 provides a method to construct a tree T from its score-2 split system or its 2-split system (see § 2 for the definition) when $\pi_{T,v}$ is given as input for some $v \in \mathrm{V}_T$ with $\deg_T(v) \geqslant 4$. Note that Algorithm 1 will also be used in our subsequent work on recovering a tree from its 2-split system [62]. In § 4, we develop a method to recognize either a one-vertex partition into at least five non-empty parts or a claw in a tree from its score-2 split system. In § 5, we present Algorithm 2 and provide a concise proof of Theorem 1.3. We conclude this paper by highlighting some further results and open questions related to this line of research in § 6.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

2. Preliminaries

2.1. A bit more notations

For a set Y and a singleton set $Z = \{z\}$, we often write $Y \cup Z$ and $Y \setminus Z$ as $Y + z$ and $Y - z$, respectively.

Let G be a graph.

For distinct $x, y \in \mathrm{L}_G$ and $v \in \mathrm{V}_G$, we call $\{x, y\}$ a *cherry* and $\{x, y; v\}$ a *full cherry* of G if $\deg_G(v) = 3$ and $\{x, y\} \subseteq \mathrm{N}_G(v)$. We call $\{x, y, z\} \in \binom{\mathrm{L}_G}{3}$ a *claw* of G provided there exists $w \in \mathrm{V}_G$ such that $\mathrm{N}_G(w) \cap \mathrm{L}_G = \{x, y, z\}$ and $\deg_G(w) = 4$. For each integer $n \geqslant 3$, the *star graph* $\mathrm{K}_{1,n}$ is a tree on $n+1$ vertices in which every two edges share a common endpoint.

We say that $e \in \mathrm{E}_G$ is *incident* with $v \in \mathrm{V}_G$ in G if v is an endpoint of e, namely $v \in e$. The number of connected components of G is denoted by $\mathrm{c}(G)$. We refer to G as a *connected graph* whenever $\mathrm{c}(G) \leq 1$. For any $U, V \subseteq \mathrm{V}_G$, we say that U *and* V *are reachable to each other in* G provided there exist $u \in U$ and $v \in V$ such that $\mathrm{C}_G[u] = \mathrm{C}_G[v]$.

For any $u, v \in \mathrm{V}_G$, the shortest length of a path between them in G is called the *distance* between them in G, which is designated by $\mathrm{dist}_G(u, v)$. For any non-empty subsets $U, W \subseteq \mathrm{V}_G$, the notation $\mathrm{dist}_G(U, W)$ represents $\min_{u \in U, w \in W} \mathrm{dist}_G(u, w)$, whereas $[U, W]_G$ consists of all vertices $v \in \mathrm{V}_G$ such that $\mathrm{dist}_G(U, v) + \mathrm{dist}_G(v, W) = \mathrm{dist}_G(U, W)$.

For any $U \subseteq \mathrm{V}_G$, the *coboundary of* U *in* G, denoted by $\partial_G(U)$, is the set $\{uv \in \mathrm{E}_G : u \in U, v \in \mathrm{V}_G \setminus U\}$.

2.2. Parsimony score, ℓ-split and smallest 2-extension

In this subsection, we introduce two basic concepts for our work: ℓ-splits and smallest 2-extensions. In addition, we discuss the parsimony score of a 2-split and establish some basic results about smallest 2-extensions.

In mathematics, parity considerations are often used to associate an invariant to a graph [4, 24, 59]. Let us introduce a class of tree-like split systems from this parity perspective. Let T be a tree. For any $V \subseteq \mathrm{V}_T$ and $F \subseteq \mathrm{E}_T$, we define a split $\widehat{\mathbf{S}}_{T,F,V}$ on V by requiring $u, v \in V$ to fall into the same part of $\widehat{\mathbf{S}}_{T,F,V}$ if and only if $|T_{u,v} \cap F|$ is an even integer. Given any V_T-split $U \mid W$, it is clear that $\widehat{\mathbf{S}}_{T,F,\mathrm{V}_T} = U \mid W$ if and only if $F = \partial_T(U) = \partial_T(W)$. The coboundary map ∂_T [72, § 2] indeed induces a natural bijection from the set of all V_T-splits to the powerset of E_T, both having size $2^{|\mathrm{V}_T|-1} = 2^{|\mathrm{E}_T|}$. Take a non-negative integer ℓ. For any $V \subseteq \mathrm{V}_T$, we define the split system $\widehat{\mathcal{S}}_{T,\ell,V}$ on V by setting

$$\widehat{\mathcal{S}}_{T,\ell,V} \doteq \left\{ \widehat{\mathbf{S}}_{T,F,V} : \ F \in \binom{\mathrm{E}_T}{\ell} \right\}.$$

For any edge set $F \subseteq \mathrm{E}_T$, we exploit the shorthand $\mathbf{S}_{T,F}$, $\mathcal{S}_{T,\ell}$, $\widehat{\mathbf{S}}_{T,F}$ and $\widehat{\mathcal{S}}_{T,\ell}$ for $\widehat{\mathbf{S}}_{T,F,\mathrm{L}_T}$ $\widehat{\mathcal{S}}_{T,\ell,\mathrm{L}_T}$, $\widehat{\mathbf{S}}_{T,F,\mathrm{V}_T}$ and $\widehat{\mathcal{S}}_{T,\ell,\mathrm{V}_T}$, respectively. We call $\mathcal{S}_{T,\ell}$ the ℓ-*split system on* T. Clearly, $\mathcal{P}_{T,0} = \mathcal{S}_{T,0} = \{\mathrm{L}_T \mid \emptyset\}$, and $\mathrm{PS}(T, \mathbf{S}_{T,F}) \leq \ell$ when $F \in \binom{\mathrm{E}_T}{\ell}$.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

Example 2.1. Let T be a tree, and let $F \subseteq \mathrm{E}_G$. Then, $\mathbf{S}_{T,F} = \mathrm{L}_T \mid \emptyset$ if and only if $F = \partial_T(U)$ for some $U \subseteq \mathrm{V}_T \setminus \mathrm{L}_T$.

Lemma 2.2. *Let T be a tree. For every non-negative integer k, it holds that $\mathcal{P}_{T,k} = \mathcal{S}_{T,k} \setminus \left(\bigcup_{i=0}^{k-1} \mathcal{S}_{T,i}\right)$. Therefore, for every L_T-split $\mathbf{S}$, it holds that $\mathrm{PS}(T,\mathbf{S}) = \min\{k : \mathbf{S} \in \mathcal{S}_{T,k}\}$.*

Proof. It is clear that $\mathcal{P}_{T,0}, \mathcal{P}_{T,1}, \ldots$ are pairwise disjoint. Henceforth, to show that $\mathcal{P}_{T,k} = (\bigcup_{i=0}^{k} \mathcal{P}_{T,i}) \setminus (\bigcup_{i=0}^{k-1} \mathcal{P}_{T,i})$ is equal to $\mathcal{S}_{T,k} \setminus (\bigcup_{i=0}^{k-1} \mathcal{S}_{T,i}) = (\bigcup_{i=0}^{k} \mathcal{S}_{T,i}) \setminus (\bigcup_{i=0}^{k-1} \mathcal{S}_{T,i})$, we only need to verify that $\bigcup_{i=0}^{k} \mathcal{P}_{T,i} = \bigcup_{i=0}^{k} \mathcal{S}_{T,i}$ holds for each non-negative integer k. Since $\bigcup_{i=0}^{k} \mathcal{S}_{T,i} \subseteq \bigcup_{i=0}^{k} \mathcal{P}_{T,i}$ holds trivially, our task reduces to proving that $\mathcal{P}_{T,k} \subseteq \mathcal{S}_{T,k}$.

Consider an L_T-split $Y \mid Z \in \mathcal{P}_{T,k}$. By definition, there exists a set $F \in \binom{\mathrm{E}_T}{k}$ intersecting with $T_{y,z}$ for each $y \in Y$ and $z \in Z$. Since T is a tree, there are $k+1$ connected components $V_1, \ldots, V_{k+1}$ in $T - F$. Denote $V_i \cap \mathrm{L}_T$ by L_i for each $i \in [k+1]$. Clearly, $\mathrm{L}_i \mid \emptyset \subseteq Y \mid Z$ for each $i \in [k+1]$. For any $f \in F$, say $f = u_i u_j$ with $u_i \in V_i$ and $u_j \in V_j$, from $\mathrm{PS}(T, Y \mid Z) = |F|$ we derive the existence of $y \in Y$ and $z \in Z$ such that $\{f\} = F \cap T_{y,z}$. This then implies that both L_i and L_j are non-empty and

$$\mathrm{L}_i \mid \mathrm{L}_j \subseteq Y \mid Z. \tag{1}$$

We now construct a tree T' on $[k+1]$ such that ij is an edge if and only if $\mathrm{dist}_T(V_i, V_j) = 1$. Let $U \mid W$ be the unique partition of $\mathrm{V}_{T'}$ into its two independent sets. From Eq. (1), we infer that $\{Y, Z\} = \{\cup_{i \in U} \mathrm{L}_i, \cup_{j \in W} \mathrm{L}_j\}$, and thus $Y \mid Z = \mathbf{S}_{T,F} \in \mathcal{S}_{T,k}$, as desired. □

Lemma 2.3. *Let T be a tree and $e, e' \in \mathrm{E}_T$. Then $\mathbf{S}_{T,\{e,e'\}}$ is improper if and only if $e \neq e'$ and $\deg_T(w) = 2$ for all $w \in [e,e']_T$. In particular, if T is a phylogenetic tree, then $\mathrm{L}_T \mid \emptyset \notin \mathcal{S}_{T,1} \cup \mathcal{S}_{T,2}$.*

Proof. Assume that $e = uv$. Take $x \in \mathrm{L}_T \cap \mathrm{C}_{T-e}[u]$ and $y \in \mathrm{L}_T \cap \mathrm{C}_{T-e}[v]$. Clearly, x and y fall into different parts of $\mathbf{S}_{T,\{e\}}$, showing that $\mathbf{S}_{T,\{e\}}$ is proper.

If $e \neq e'$ and $\deg_T(w) = 2$ for all $w \in [e,e']_T$, then $\mathbf{S}_{T,\{e,e'\}}$ is surely improper.

It remains to show that $\mathbf{S}_{T,\{e,e'\}}$ is proper in the case that $e \neq e'$ and $[e,e']_T$ contains a vertex w with $\deg_T(w) \geqslant 3$. We first find $x, y \in \mathrm{L}_T$ such that $\{e, e'\} \subseteq T_{x,y}$. Thanks to $\deg_T(w) \geqslant 3$, we can then pick a vertex $w' \in \mathrm{N}_T(w)$ outside the path between x and y in T. Choose $z \in \mathrm{L}_T \cap \mathrm{C}_{T-w}[w']$. One can check that x and z fall into different parts of $\mathbf{S}_{T,\{e,e'\}}$, and thus $\mathbf{S}_{T,\{e,e'\}}$ is a proper split. □

Remark 2.4. Let T be a tree. By Lemma 2.2, $\mathcal{P}_{T,1} = \mathcal{S}_{T,1} \setminus \mathcal{S}_{T,0}$ and $\mathcal{P}_{T,2} = \mathcal{S}_{T,2} \setminus (\mathcal{S}_{T,1} \cup \mathcal{S}_{T,0})$. Recall that $\mathcal{S}_{T,0} = \{\mathrm{L}_T \mid \emptyset\}$. Therefore, a further application of Lemma 2.3 shows that $\mathcal{P}_{T,1} = \mathcal{S}_{T,1}$, and that, when T is even a phylogenetic tree, $\mathcal{P}_{T,2} = \mathcal{S}_{T,2} \setminus \mathcal{S}_{T,1}$.

Let T be a tree. According to Example 2.1, if $\mathbf{S}_{T,F_1} = \mathbf{S}_{T,F_2}$, then $F_1 \triangle F_2 = \partial_T(U)$ for some $U \subseteq \mathrm{V}_T \setminus \mathrm{L}_T$. If $|F_1| = 1$, $|F_2| = 2$, and T is phylogenetic, we can say more about the structure of F_1 and F_2, as shown in Lemma 2.5.

Lemma 2.5. *Let T be a phylogenetic tree. Take $\{e_1, e_2\} \in \binom{\mathrm{E}_T}{2}$ and $e_3 \in \mathrm{E}_T$. Then $\mathbf{S}_{T,\{e_1,e_2\}} = \mathbf{S}_{T,\{e_3\}}$ if and only if there exists $v \in \mathrm{V}_T$ with $\deg_T(v) = 3$ such that the three edges of T incident with v are e_1, e_2, and e_3.*

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

Proof. The backward direction is obvious. We only consider the forward direction.

Assume that $v_1 \in e_1$, $v_2 \in e_2$ and $\mathrm{dist}_T(e_1, e_2) = \mathrm{dist}_T(v_1, v_2)$. Note that v_1 and v_2 must be inner vertices of the phylogenetic tree T. If $v_1 = v_2$ and $\deg_T(v_1) = 3$, then it is direct from $\mathbf{S}_{T,\{e_1,e_2\}} = \mathbf{S}_{T,\{e_3\}}$ that e_1, e_2, and e_3 are the three edges incident with v_1 in T, as required. We thus assume that either $v_1 \neq v_2$ or $\deg_T(v_1) > 3$, and we aim to derive a contradiction. Our assumption now enables us find $w_1, w_2 \in \mathrm{V}_T$ such that $v_1w_1, v_2w_2 \in \mathrm{E}_T \setminus (\{e_1, e_2\} \cup T_{v_1,v_2})$ and $v_1w_1 \neq v_2w_2$. We can accordingly find $x_1, y_1, x_2, y_2 \in \mathrm{L}_T$ such that, for each $i \in [2]$, $e_{3-i} \notin T_{x_i,y_i} \supseteq \{e_i, v_iw_i\}$, thus implying that x_i and y_i belong to different parts of $\mathbf{S}_{T,\{e_1,e_2\}}$. Since T_{x_1,y_1} and T_{x_2,y_2} are disjoint, one of them cannot contain the edge e_3. This contradicts $\mathbf{S}_{T,\{e_1,e_2\}} = \mathbf{S}_{T,\{e_3\}}$. □

Corollary 2.6. *Let T be a phylogenetic tree and let $\{e_1, e_2\} \in \binom{\mathrm{E}_T}{2}$. Then, $\mathbf{S}_{T,\{e_1,e_2\}} \notin \mathcal{P}_{T,2}$ if and only if e_1 and e_2 share a common endpoint v with $\deg_T(v) = 3$.*

Proof. By Remark 2.4, $\mathbf{S}_{T,\{e_1,e_2\}} \notin \mathcal{P}_{T,2}$ if and only if there exists $e_3 \in \mathrm{E}_T$ such that $\mathbf{S}_{T,\{e_1,e_2\}} = \mathbf{S}_{T,\{e_3\}}$. It follows from Lemma 2.5 that the latter condition is further equivalent to the condition that e_1 and e_2 share a common endpoint v with $\deg_T(v) = 3$. □

Corollary 2.7. *Let T be a phylogenetic tree with at least three leaves and $\{x, y\} \in \binom{\mathrm{L}_T}{2}$. Then $\{x, y\} \mid (\mathrm{L}_T \setminus \{x, y\}) \notin \mathcal{P}_{T,2}$ if and only if $\{x, y\}$ is a cherry of T.*

Proof. Let e_1 and e_2 be the edges incident with x and y in T, respectively. Note that $\{x, y\} \mid (\mathrm{L}_T \setminus \{x, y\}) = \mathbf{S}_{T,\{e_1,e_2\}}$. The result now follows directly from Corollary 2.6. □

Let T be a tree and let F be a set. We adopt the convention that $\mathbf{S}_{T,F}$ denotes $\mathbf{S}_{T,F\cap\mathrm{E}_T}$.

Lemma 2.8. *Let T be a tree and let $\tilde{T}$ be a subtree of T. If $\mathrm{L}_{\tilde{T}} \subseteq \mathrm{L}_T$, then $\mathbf{S}_{\tilde{T},F} = \mathbf{S}_{T,F}|_{\mathrm{L}_{\tilde{T}}}$.*

Proof. Take $x, y \in \mathrm{L}_{\tilde{T}} \subseteq \mathrm{L}_T$. Clearly, we have $T_{x,y} = \tilde{T}_{x,y}$. Therefore, x and y belong to the same part of $\mathbf{S}_{\tilde{T},F}$ if and only if they belong to the same part of $\mathbf{S}_{T,F}$. □

Lemma 2.9. *Let T be a phylogenetic tree and let $w_0 \in \mathrm{V}_T$. Assume that $\deg_T(w_0) = k$ and $\mathrm{N}_T(w_0) \doteq \{w_1, \ldots, w_k\}$. For each $j \in [k]$, let $V_j \doteq \mathrm{C}_{T-w_0}[w_j]$ and $Y_j \doteq \mathrm{L}_T \cap V_j$. Assume that $Y \mid Z = \mathbf{S}_{T,\{e_1,e_2\}}$ for some $e_1, e_2 \in \mathrm{E}_T$ and $Y_i \subsetneq Y$ for some $i \in [k]$. If $|\{p \in [k] : Z \cap Y_p \neq \emptyset\}| \geqslant 3$, then $e_1 \neq e_2$ and $w_0w_i \in \{e_1, e_2\}$.*

Proof. Let us write T' for the tree $T[V_i + w_0]$. Take $y_i \in Y_i$ and $y \in Y \setminus Y_i$. Considering that $Z \cap Y_i \subseteq Z \cap Y = \emptyset$ and $|\{p \in [k] : Z \cap Y_p \neq \emptyset\}| \geqslant 3$, we may assume that $i \geqslant 4$ and $Z \cap Y_j \neq \emptyset$ for all $j \in [3]$. This implies the existence of $y_j \in Y_j$ for each $j \in [3]$ such that $|T_{y_j,y_i} \cap \{e_1, e_2\}| = 1$. Since T_{y_j,w_0}, $j \in [3]$, are three disjoint edge sets, the pigeonhole principle tells us that $\{e_1, e_2\} \cap \mathrm{E}_{T'} \neq \emptyset$. For each $z \in Z$, $\{e_1, e_2\} \cap T_{y,z}$ is a singleton set. As $T_{y,z} \cap \mathrm{E}_{T'} = \emptyset$, we see that $\{e_1, e_2\} \setminus \mathrm{E}_{T'} \neq \emptyset$. To sum up, we have $|\{e_1, e_2\} \cap \mathrm{E}_{T'}| = |\{e_1, e_2\} \setminus \mathrm{E}_{T'}| = 1$, which means that $e_1 \neq e_2$. Assume that $\{e_1\} = \{e_1, e_2\} \cap \mathrm{E}_{T'}$. If $V_i = \{w_i\}$, then clearly $e_1 = w_0w_i$. If $|V_i| \geqslant 2$, then, applying Lemma 2.8 for $F = \{e_1, e_2\}$ and $\tilde{T} = T[V_i]$, we find that $\mathbf{S}_{T[V_i],\{e_1\}} = Y_i \mid \emptyset$. In view of Lemma 2.3, we now conclude that $e_1 \notin \mathrm{E}_{T[V_i]}$, i.e., $e_1 = w_0w_i$. □

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

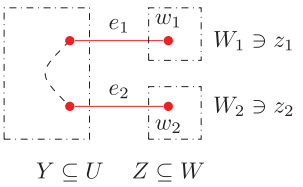


Figure 4. Edge boundary of size two in a tree.

Let T be a tree. For any L_T-split $\mathbf{S}$, the *full ℓ-split system of T induced by $\boldsymbol{S}$* is defined as $\{\widehat{\mathbf{S}} \in \widehat{\mathcal{S}}_{T,\ell} : \mathbf{S} = \widehat{\mathbf{S}}|_{\mathrm{L}_T}\}$.

Lemma 2.10. *Let T be a phylogenetic tree and let $\mathbf{S} \in \mathcal{S}_{T,2}$. Then the full 2-split system of T induced by $\mathbf{S}$ is compatible.*

Proof. Let $\mathbf{S} = Y \mid Z \in \mathcal{S}_{T,2}$. Take $\{e_1, e_2\}, \{e_1', e_2'\} \in \binom{\mathrm{E}_T}{2}$ such that $U \mid W = \widehat{\mathbf{S}}_{T,\{e_1,e_2\}}$ and $U' \mid W' = \widehat{\mathbf{S}}_{T,\{e_1',e_2'\}}$ are two splits satisfying $Y = U \cap \mathrm{L}_T = U' \cap \mathrm{L}_T$. We aim to show that either $U \subseteq U'$ or $W \subseteq W'$. Due to the symmetry between Y and Z as well as the symmetry between $U \mid W$ and $U' \mid W'$, we may assume that $\mathrm{c}(T[U]) = 1$, $\mathrm{c}(T[W]) = 2$ and W_1 and W_2 are the two connected components of $T[W]$. For each $i \in [2]$, take $z_i \in Z \cap W_i$ and let $\{w_i\} = W_i \cap e_i$. The scenario we now encounter is depicted in Fig. 4.

Since T is a phylogenetic tree, we have $\mathrm{L}_{T[W_1]} \cup \mathrm{L}_{T[W_2]} \subseteq \mathrm{L}_T$. Accordingly, as $(U \mid W)|_{\mathrm{L}_T} = (U' \mid W')|_{\mathrm{L}_T} = Y \mid Z$, Lemmas 2.3 and 2.8 tell us that, for each $i \in [2]$, $\mathrm{E}_{T[W_i]} \cap \{e_1', e_2'\}$ cannot be a singleton set. Without loss of generality, we distinguish the following two cases.

CASE 1. $\{e_1', e_2'\} \subseteq \mathrm{E}_{T[W_1]}$.
This forces $U' \supseteq U \cup W_2 \supseteq U$.

CASE 2. $\{e_1', e_2'\} \cap \mathrm{E}_{T[W]} = \emptyset$.
Since $\widehat{\mathbf{S}}_{T,\{e_1',e_2'\}}|_{\mathrm{L}_T} = (U' \mid W')|_{\mathrm{L}_T} = Y \mid Z$, we know that $T_{z_1,z_2} \cap \{e_1', e_2'\}$ contains zero or two elements. Given that $\{e_1', e_2'\} \cap \mathrm{E}_{T[W]} = \emptyset$, it follows that $T_{w_1,w_2} \cap \{e_1', e_2'\}$ has size zero or two. This demonstrates that $W' \supseteq W$, as required. □

Let T be a phylogenetic tree. Let Y be a subset of L_T such that $Y \mid (\mathrm{L}_T \setminus Y) \in \mathcal{S}_{T,2}$. Lemma 2.10 guarantees the existence of a unique inclusion-minimal element in $\{U \subseteq \mathrm{V}_T : U \mid (\mathrm{V}_T \setminus U) \in \widehat{\mathcal{S}}_{T,2}, Y = U \cap \mathrm{L}_T\}$. We denote this element by $\mathrm{Ext}_{T,Y}$ and call it the *smallest 2-extension of Y in T*. For a subset Z of L_T such that $Z \mid (\mathrm{L}_T \setminus Z) \in \mathcal{P}_{T,2}$, we know from Lemma 2.2 that $Z \mid (\mathrm{L}_T \setminus Z) \in \mathcal{S}_{T,2}$ and thus it makes sense to talk about $\mathrm{Ext}_{T,Z}$.

Lemma 2.11. *Let T be a phylogenetic tree. Consider $\{v_1u_1, v_2u_2\} \in \binom{\mathrm{E}_T}{2}$ such that $W \mid (\mathrm{V}_T \setminus W) = \widehat{\mathbf{S}}_{T,\{v_1u_1,v_2u_2\}}$ and $\{v_1, v_2\} \subseteq W$. If $Y = W \cap \mathrm{L}_T$ and*

$$\max\{\deg_T(v_1), \deg_T(v_2)\} + \mathrm{dist}_T(v_1, v_2) \geqslant 5, \tag{2}$$

then W coincides with $\mathrm{Ext}_{T,Y}$, the smallest 2-extension of Y in T.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

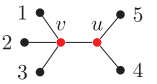


Figure 5. A tree with five leaves.

Proof. For each $i \in [2]$, let $T_i = T[\mathrm{C}_{T[W]}[v_i]]$. It is direct from definition that $Y \mid (\mathrm{L}_T \setminus Y) = \mathbf{S}_{T,\{v_1u_1,v_2u_2\}} \in \mathcal{S}_{T,2}$ and $\mathrm{Ext}_{T,Y} \subseteq W$. The latter implies that $\partial_T(\mathrm{Ext}_{T,Y}) \subseteq \mathrm{E}_{T[W]} \cup \{v_1u_1, v_2u_2\} = \mathrm{E}_{T_1} \cup \mathrm{E}_{T_2} \cup \{v_1u_1, v_2u_2\}$. Let $\partial_T(\mathrm{Ext}_{T,Y}) = \{e_1, e_2\} \in \binom{\mathrm{E}_T}{2}$. By symmetry, it suffices to verify that $E = \emptyset$, where $E \doteq \{e_1, e_2\} \cap \mathrm{E}_{T_1}$.

If $v_1 = v_2$, then $\deg_{T_1}(v_1) = \deg_T(v_1) - 2$, and hence, by Eq. (2),

$$\deg_{T_1}(v_1) \geqslant 5 - 2 = 3. \tag{3}$$

If $v_1 \neq v_2$, then $\deg_{T_1}(v_1) = \deg_T(v_1) - 1 \neq 1$. If $v_1 \neq v_2$ and $v_2 \in \mathrm{V}_{T_1}$, we have $\deg_{T_1}(v_2) = \deg_T(v_2) - 1 \neq 1$. These observations imply that $\mathrm{L}_{T_1} \subseteq Y$. Thus, we can apply Lemma 2.8 to get that $\mathbf{S}_{T_1,E} = \mathbf{S}_{T,\{e_1,e_2\}}|_{\mathrm{L}_{T_1}}$ is an improper split. Accordingly, Lemma 2.3 tells us that either $E = \emptyset$, or $|E| = 2$ and $\deg_{T_1}(w) = 2$ for all $w \in [e_1, e_2]_{T_1}$. Assume, for the sake of contradiction, that the second possibility happens. Since $\deg_{T_1}(w) \neq 2$ for all $w \in \mathrm{V}_{T_1} \setminus \{v_1, v_2\}$, we know that $\emptyset \neq [e_1, e_2]_{T_1} \subseteq \{v_1, v_2\}$, and hence $2 \in \{\deg_{T_1}(v_1), \deg_{T_1}(v_2)\}$. By virtue of Eq. (3), one of v_1, v_2 having degree two forces $v_1 \neq v_2$. We thus have two cases to consider: 1) $[e_1, e_2]_{T_1} = \{v_1, v_2\}$ is of size two; 2) $[e_1, e_2]_{T_1} = \{v_i\} \subsetneq \{v_1, v_2\}$ for some $i \in [2]$. For case 1), we find that $v_1v_2 \in \mathrm{E}_T$ and $\deg_T(v_1) = \deg_T(v_2) = 3$. It follows that $\max\{\deg_T(v_1), \deg_T(v_2)\} + \mathrm{dist}_T(v_1, v_2) = 4$, violating Eq. (2), as desired. For case 2), we have $\deg_T(v_i) = \deg_{T_1}(v_i) + 1 = 3$ and hence $\{e \in \mathrm{E}_T : \ v_i \in e\} = \{e_1, e_2, v_iu_i\}$. It follows that $\mathbf{S}_{T,\{v_1u_1,v_2u_2\}} = Y \mid (\mathrm{L}_T \setminus Y) = \mathbf{S}_{T,\{e_1,e_2\}} = \mathbf{S}_{T,\{v_iu_i\}}$, which contradicts Lemma 2.5. □

Example 2.12. Let T be the tree drawn in Fig. 5. Then, $\mathrm{Ext}_{T,\{1,2,3,5\}} = \{1, 2, 3, 5, v\}$ and $\mathrm{Ext}_{T,\{1,2,5\}} = \{1, 2, 5, v, u\}$. This shows that $Y \subseteq Y'$ does not necessarily imply $\mathrm{Ext}_{T,Y} \subseteq \mathrm{Ext}_{T,Y'}$.

3. Tree-reduction via cherry and tree-growing from non-cherry markers

Lemma 3.1 is a simple result used to handle cherry markers. Based on Lemma 3.1, our main algorithm, Algorithm 2, can reduce the problem of reconstructing a phylogenetic tree T from $\mathcal{P}_{T,2}$ to the case where T has at most one cherry. In Algorithm 2, after identifying from split system information either a vertex of degree at least five or a claw in the potential phylogenetic tree, we need to extract information from the given split system to grow the entire tree from these two non-cherry local markers. For this purpose, this section presents Algorithm 1 and demonstrates its usefulness in Lemma 3.3. To prove Lemma 3.3, it is necessary to first establish Lemma 3.2.

Lemma 3.1. *Let T be a phylogenetic tree with at least four leaves. If $\{x_1, x_2; v\}$ is a full cherry of T, then $\mathcal{P}_{T-\{x_1,x_2\},2} = \big\{(Y + v) \mid Z : \ (Y \cup \{x_1, x_2\}) \mid Z \in \mathcal{P}_{T,2}, Y \cup Z = \mathrm{L}_T \setminus \{x_1, x_2\}\big\}$.*

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

Proof. Let Y and Z be disjoint subsets of $\mathrm{L}_T \setminus \{x_1, x_2\}$ and let $T' = T - \{x_1, x_2\}$. Note that T' is a phylogenetic tree with $\mathrm{L}_{T'} = (\mathrm{L}_T \setminus \{x_1, x_2\}) + v$ and $\mathrm{E}_{T'} = \mathrm{E}_T \setminus \{x_1 v, x_2 v\}$.

Suppose $(Y + v) \mid Z \in \mathcal{P}_{T',2}$. By Lemma 2.2, there exists $\{e_1, e_2\} \in \binom{\mathrm{E}_{T'}}{2}$ such that $(Y + v) \mid Z = \mathbf{S}_{T',\{e_1,e_2\}} \in \mathcal{P}_{T',2}$. It follows from Corollary 2.6 that e_1 and e_2 do not share any common endpoint of degree three in T'. This indicates that $e_1, e_2 \in \mathrm{E}_T$ do not share a common endpoint of degree three in T, and hence, by Corollary 2.6 again, $\mathbf{S}_{T,\{e_1,e_2\}} \in \mathcal{P}_{T,2}$. Since $e_1, e_2 \notin \{x_1 v, x_2 v\}$, for every $x \in Y \cup Z$ and $i \in [2]$,

$$T_{x,x_i} \cap \{e_1, e_2\} = (T_{x,v} \cup \{x_i v\}) \cap \{e_1, e_2\} = T_{x,v} \cap \{e_1, e_2\} = (T')_{x,v} \cap \{e_1, e_2\}.$$

Thus, $(Y \cup \{x_1, x_2\}) \mid Z = \mathbf{S}_{T,\{e_1,e_2\}} \in \mathcal{P}_{T,2}$.

Suppose $(Y \cup \{x_1, x_2\}) \mid Z \in \mathcal{P}_{T,2}$. By Lemma 2.2, there exists a set $\{e, e'\} \in \binom{\mathrm{E}_T}{2}$ such that $(Y \cup \{x_1, x_2\}) \mid Z = \mathbf{S}_{T,\{e,e'\}} \in \mathcal{P}_{T,2}$. From Corollary 2.6, it follows that e and e' do not share any common endpoint of degree three in T. Because $|T_{x_1,x_2} \cap \{e, e'\}|$ is even and $x_1 v, x_2 v$ share a common endpoint of degree three in T, it holds that $\{e, e'\} \cap \{x_1 v, x_2 v\} = \emptyset$. In other words, $e, e' \in \mathrm{E}_{T'}$. Noting that $|T'_{x,v} \cap \{e, e'\}| = |T_{x,v} \cap \{e, e'\}| = |T_{x,x_i} \cap \{e, e'\}|$ for every $x \in Y \cup Z$ and $i \in [2]$, we conclude that $\mathbf{S}_{T',\{e,e'\}} = (Y + v) \mid Z$. Recall that e and e' do not share any common endpoint of degree three in T, and hence in T'. Consequently, it follows from Corollary 2.6 that $(Y + v) \mid Z \in \mathcal{P}_{T',2}$. □

Lemma 3.2. *Let T be a phylogenetic tree and take $v_0 \in \mathrm{V}_T$ with $\deg_T(v_0) = k \geqslant 4$. Suppose $\mathrm{N}_T(v_0) = \{v_1, \ldots, v_k\}$. For every $\ell \in [k]$, we denote by Y_ℓ the set $\mathrm{L}_T \cap \mathrm{C}_{T-v_0}[v_\ell]$ and by T_ℓ the tree $T[\mathrm{C}_{T-v_0}[v_\ell] + v_0]$. For every $\{i, j\} \in \binom{[k]}{2}$, it holds that $\mathcal{P}_{T_i,1} = \mathcal{S}_{i,j} = \mathcal{S}'_{i,j}$, where $\mathcal{S}_{i,j} \doteq \{Y' \mid ((Y_i \setminus Y') + v_0) : \emptyset \neq Y' \subseteq Y_i, (Y' \cup Y_j) \mid (\mathrm{L}_T \setminus (Y' \cup Y_j)) \in \mathcal{P}_{T,2}\}$ and $\mathcal{S}'_{i,j} \doteq \{Y' \mid ((Y_i \setminus Y') + v_0) : \emptyset \neq Y' \subseteq Y_i, (Y' \cup Y_j) \mid (\mathrm{L}_T \setminus (Y' \cup Y_j)) \in \mathcal{S}_{T,2}\}$.*

Proof. From Remark 2.4 we know that $\mathcal{P}_{T_i,1} = \mathcal{S}_{T_i,1}$ and $\mathcal{P}_{T,2} \subseteq \mathcal{S}_{T,2}$. Henceforth, we only need to prove that $\mathcal{S}_{T_i,1} \subseteq \mathcal{S}_{i,j}$ and $\mathcal{S}'_{i,j} \subseteq \mathcal{S}_{T_i,1}$.

Take an element from $\mathcal{S}_{T_i,1}$, say $Y' \mid ((Y_i \setminus Y') + v_0) = \mathbf{S}_{T_i,\{e\}}$ for some $e \in \mathrm{E}_{T_i}$. First, it follows from Lemma 2.3 that $\emptyset \neq Y' \subseteq Y_i$. We next observe that $\mathbf{S}_{T,\{e,v_0v_j\}} = (Y' \cup Y_j) \mid (\mathrm{L}_T \setminus (Y' \cup Y_j))$ and the only possible common endpoint of e and $v_0 v_j$ is v_0. Due to $\deg_T(v_0) \geqslant 4$, Corollary 2.6 claims that $\mathbf{S}_{T,\{e,v_0v_j\}} \in \mathcal{P}_{T,2}$. This completes the proof of $\mathcal{S}_{T_i,1} \subseteq \mathcal{S}_{i,j}$.

Take a set Y' such that $\mathbf{S} \doteq Y' \mid ((Y_i \setminus Y') + v_0) \in \mathcal{S}'_{i,j}$. Assume that $(Y' \cup Y_j) \mid (\mathrm{L}_T \setminus (Y' \cup Y_j)) = \mathbf{S}_{T,\{e_1,e_2\}}$ for some $\{e_1, e_2\} \in \binom{\mathrm{E}_T}{2}$. We need to verify that $Y' \mid ((Y_i \setminus Y') + v_0) \in \mathcal{S}_{T_i,1}$. If $Y' = Y_i$, then $\mathbf{S} = \mathbf{S}_{T_i,\{v_0v_i\}} \in \mathcal{S}_{T_i,1}$. Now assume that $\emptyset \neq Y' \subsetneq Y_i$. Since $k \geqslant 4$, there are at least three components of $T - v_0$ that intersect with $\mathrm{L}_T \setminus (Y' \cup Y_j)$. Applying Lemma 2.9 for $Y \doteq Y' \cup Y_j$ and $Z \doteq \mathrm{L}_T \setminus (Y' \cup Y_j)$ yields that $e_t = v_0 v_j$ for some $t \in [2]$. By Lemma 2.8, $\mathbf{S}_{T_i - v_0,\{e_{3-t}\}} = \mathbf{S}_{T,\{e_1,e_2\}}|_{Y_i} = Y' \mid (Y_i \setminus Y')$. Since $Y' \mid (Y_i \setminus Y')$ is proper, we deduce that $e_{3-t} \in \mathrm{E}_{T_i}$. Take a leaf $y \in \mathrm{L}_T \setminus (Y_i \cup Y_j)$. Then for every $x \in Y_i$, $(T_i)_{x,v_0} \cap \{e_{3-t}\} = T_{x,y} \cap \{e_1, e_2\}$. Consequently, $\mathbf{S} = \mathbf{S}_{T_i,\{e_{3-t}\}} \in \mathcal{S}_{T_i,1}$, as desired. □

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

Algorithm 1. Construct a phylogenetic X-tree from a split system and a k-partition of X

Input: A set X, a split system $\mathcal{S}$ on X, and a k-partition $Y_1 \mid \cdots \mid Y_k$ of X.
Output: A phylogenetic X-tree or $\emptyset$.
1: $v_0 \leftarrow$ an element not in X.
2: **for** every $i \in [k]$ **do**
3: $j \leftarrow$ the smallest integer in $[k] \setminus \{i\}$. ▷ Any $j \in [k] \setminus \{i\}$ is fine.
4: $\mathcal{S}_i \leftarrow \{Y' \mid (Y_i \setminus Y' + v_0) : \emptyset \neq Y' \subseteq Y_i, (Y' \cup Y_j) \mid (X \setminus (Y' \cup Y_j)) \in \mathcal{S}\}$.
5: **if** $\mathcal{S}_i$ is compatible and contains all trivial $(Y_i + v_0)$-splits **then**
6: $T_i \leftarrow$ the phylogenetic $(Y_i + v_0)$-tree reconstructed from $\mathcal{S}_i$ following any algorithmic proof of the Buneman Theorem (Theorem 1.1).
7: **else**
8: **return** $\emptyset$.
9: **end if**
10: **end for**
11: $T \leftarrow$ the phylogenetic X-tree with $\mathrm{V}_T = \bigcup_{i\in[k]} \mathrm{V}_{T_i}$ and $\mathrm{E}_T = \bigcup_{i\in[k]} \mathrm{E}_{T_i}$, where $\mathrm{V}_{T_i} \setminus \{v_0\}$, E_{T_i}, $i \in [k]$, are regarded as $2k$ pairwise disjoint sets.
12: **return** T.

Lemma 3.3. *Let $\mathcal{S}$ be a split system on a finite set X and let π be a k-partition of X, where $k \geqslant 4$. Up to X-isomorphism, there exists at most one phylogenetic X-tree T such that $\mathcal{S} \in \{\mathcal{S}_{T,2}, \mathcal{P}_{T,2}\}$ and π is a one-vertex partition of T. If such a phylogenetic X-tree T exists, then, given the inputs $X, \mathcal{S}$, and π, the output of Algorithm 1 is the X-tree T.*

Proof. Assume that T is a phylogenetic X-tree such that $\mathcal{S} \in \{\mathcal{S}_{T,2}, \mathcal{P}_{T,2}\}$ and π is a one-vertex partition of T. Let $\pi = Y_1 \mid \cdots \mid Y_k = \pi_{T,v_0}$ for some $v_0 \in \mathrm{V}_T \setminus \mathrm{L}_T$. We show that, given the inputs $\mathcal{S}$ and π, Algorithm 1 outputs T up to X-isomorphism. For every $i \in [k]$, let T_i be the unique subtree of T with $\mathrm{L}_{T_i} = Y_i + v_0$; and let $\mathcal{S}_i$ be the split system defined in Line 4 of Algorithm 1. By Lemma 3.2, we have $\mathcal{S}_i = \mathcal{P}_{T_i,1}$ for every $i \in [k]$. According to Theorem 1.1 (2), T_i can be uniquely recovered from $\mathcal{S}_i$ for every $i \in [k]$, as performed in Lines 5 to 9 of Algorithm 1. The construction of T from $T_1, \ldots, T_k$ is straightforward; see Line 11 of Algorithm 1.

Note that Algorithm 1 outputs at most one phylogenetic X-tree. Thus, up to X-isomorphism, there is at most one phylogenetic X-tree T satisfying the two required conditions, which completes the proof. □

4. Recognize a one-vertex partition into at least five parts or a claw

Let T be a phylogenetic tree. We have Corollary 2.7 as a simple tool to identify a cherry of T from $\mathcal{P}_{T,2}$. However, locating the other two types of markers in T from $\mathcal{P}_{T,2}$ is less straightforward. Section 4.1 establishes a correspondence between the so-called $(k, 2)$-set with respect to $\mathcal{P}_{T,2}$ and five types of 2-special configurations in the tree T; see

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

Definitions 4.1, 4.2 and Lemma 4.4. We continue in § 4.2 to use this correspondence to find the other two markers. Note that every vertex of degree at least five in T must show up in some 2-special configuration of type A. By encoding type-A 2-special configurations via split system information in Lemma 4.13, we can identify those large-degree vertex markers. Under some mild assumptions, Lemma 4.15 shows how to find a claw from the given split system input. In fact, as demonstrated in the proof of Lemma 4.15, each claw of T is embedded in some 2-special configuration of type E associated with T. Although we cannot directly identify all type-E 2-special configurations from $\mathcal{P}_{T,2}$, this connection with type E enables us to find all claw markers.

4.1. (k,2)-set and 2-special configuration

Definition 4.1 ((k,ℓ)-set). *Let X be a non-empty finite set, and let k, ℓ be positive integers such that $\ell \leq k \leq |X|$. If a set family $\mathcal{F}$ equals $\left\{\bigcup_{i\in I} X_i : I \in \binom{[k]}{\ell}\right\}$ for some k-partition $\pi = X_1 \mid \cdots \mid X_k$ of X, then $\mathcal{F}$ is called a (k,ℓ)-*hypergraph on X *and denoted by $\mathcal{H}_{\pi,\ell}$. Let $\mathcal{S}$ be an X-split system. For every $Y \subseteq X$, let $\Gamma_{\mathcal{S}}(Y)$ stand for the set of inclusion-maximal elements of $\{Z \subsetneq Y : Z \mid (X \setminus Z) \in \mathcal{S}\}$. We say that $Y \subseteq X$ is a (k,ℓ)-*set *with respect to $\mathcal{S}$ if $Y \mid (X \setminus Y) \in \mathcal{S}$ and $\Gamma_{\mathcal{S}}(Y)$ is a (k,ℓ)-hypergraph on* Y. *The reader may verify that there is no (k,k)-set with respect to $\mathcal{S}$.*

For any phylogenetic tree T with a vertex v satisfying $\deg_T(v) = k+2 \geqslant 5$ and $\mathrm{N}_T(v) = \{v_1, \ldots, v_{k+2}\}$, observe that $\mathrm{L}_T \cap (\bigcup_{i\in[k]} \mathrm{C}_{T-v}[v_i])$ is a $(k,2)$-set with respect to $\mathcal{P}_{T,2}$. Additional $(k,2)$-sets associated with $\mathcal{P}_{T,2}$ are described in Lemma 4.4. To introduce Lemma 4.4, the key lemma in our work, one more definition is required.

Definition 4.2 (2-special configuration). *Suppose T is a phylogenetic tree and $Y \mid (\mathrm{L}_T \setminus Y) \in \mathcal{S}_{T,2}$. Assume that $\partial_T(\mathrm{Ext}_{T,Y}) = \{v_1u_1, v_2u_2\} \in \binom{\mathrm{E}_T}{2}$, $\{v_1, v_2\} \subseteq \mathrm{Ext}_{T,Y}$, and $\deg_T(v_1) \geqslant \deg_T(v_2)$. Denote $\mathrm{c}(T[\mathrm{Ext}_{T,Y}] - (\{v_1, v_2\} \setminus \mathrm{L}_T))$ by $\phi_T(Y)$. We call (T, Y) a* 2-special configuration *whenever one of the following five conditions holds.*

- (A) *$v_1 = v_2$ and $\deg_T(v_1) \geqslant 5$. See Fig. 6(*A*).*
- (B) *$v_1v_2 \in \mathrm{E}_T$ and $\deg_T(v_1) = \deg_T(v_2) = 4$. See Fig. 6(*B*).*
- (C) *$v_1v_2 \in \mathrm{E}_T$ and $\deg_T(v_1) - 1 = \deg_T(v_2) = 3$. See Fig. 6(*C*).*
- (D) *$T[W]$ is connected, $\mathrm{dist}_T(v_1, v_2) > 1$ and $\deg_T(v_1) = \deg_T(v_2) = 3$. See Fig. 6(*D*).*
- (E) *$|Y| = \deg_T(v_1) = 4$, $\deg_T(v_2) = 1$, and either $u_1 \neq u_2$ or $\deg_T(u_1) \geqslant 4$. See Fig. 6(*E*).*

For any 2-special configuration (T, Y), we say that its type *is A, B, C, D, or E if the corresponding condition* (A), (B), (C), (D), *or* (E) *holds, respectively. When (T, Y) is a 2-special configuration of type* (A), (B), (C), (D), *or* (E), *as shown in Fig. 6, we have $\phi_T(Y) = \deg_T(v_1) - 2, 4, 3, 3,$ and 4, respectively.*

Remark 4.3. The conditions (A), (B), (C), (D), (E) listed in Definition 4.2 are mutually exclusive. Case (A) is the only case where $\mathrm{dist}_T(v_1, v_2) = 0$. In cases (B), (C) and (D), we have $\deg_T(v_i) \geqslant 3$ for each $i \in [2]$, while in case (E), we have $\deg_T(v_2) = 1$. Whether or not v_1 and v_2 are adjacent distinguishes (D) from (B) and (C). In case (B), it holds that $\deg_T(v_1) = \deg_T(v_2) = 4$, whereas in case (C), we have $\deg_T(v_2) = 3$.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

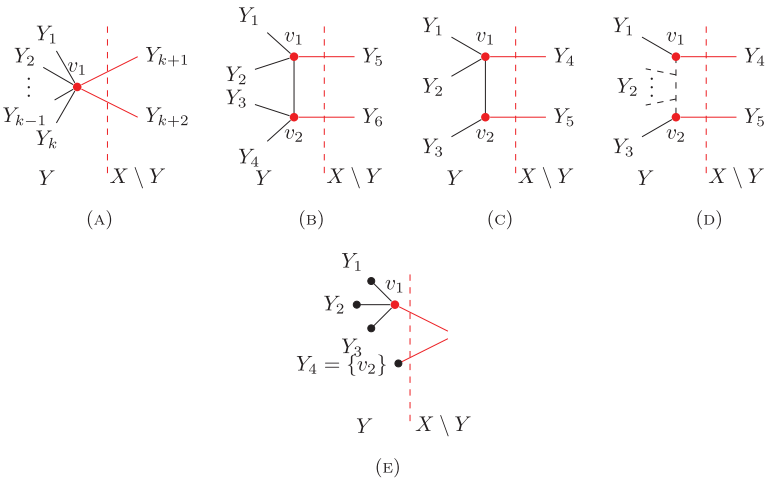


Figure 6. Five types of 2-special configurations (T, Y). Each Y_i represents a subset of $X = \mathrm{L}_T$ within a corresponding subtree. See Definition 4.2. A) $\phi_T(Y) = k$ B) $\phi_T(Y) = 4$ C) $\phi_T(Y) = 3$ D) $\phi_T(Y) = 3$ E) $\phi_T(Y) = 4$

Lemma 4.4. (Basic Lemma for score-2 split system) *Let $k \geqslant 3$ be an integer, let T be a phylogenetic tree, and let $Y \subseteq \mathrm{L}_T$. Then the following statements are equivalent:*

(1) *Y is a $(k, 2)$-set with respect to $\mathcal{P}_{T,2}$.*
(2) *$k = \phi_T(Y)$ and (T, Y) is a 2-special configuration.*

We refer to Lemma 4.4 as a basic lemma because it serves as a key step in connecting the graph structure with the split system structure. An analogous basic lemma for 2-split systems will be crucial for our work in [62]. To present a proof of Lemma 4.4, we require a convention and two additional technical lemmas.

Lemma 4.5. *Let T be a phylogenetic tree, and let $Y \mid (\mathrm{L}_T \setminus Y) \in \mathcal{P}_{T,2}$. Assume that $\{v_1u_1, v_2u_2\} = \partial_T(\mathrm{Ext}_{T,Y})$, where $v_1, v_2 \in \mathrm{Ext}_{T,Y}$. Then exactly one of the following holds:*

(a) *$v_1 = v_2$ and $\deg_T(v_1) \geqslant 5$.*
(b) *$v_1v_2 \in \mathrm{E}_T$ and $\max\{\deg_T(v_1), \deg_T(v_2)\} \geqslant 4$.*
(c) *$T[\mathrm{Ext}_{T,Y}]$ is connected and $\mathrm{dist}_T(v_1, v_2) > 1$.*
(d) *$T[\mathrm{Ext}_{T,Y}]$ is not connected, and either $u_1 \neq u_2$ or $\deg_T(u_1) \geqslant 4$.*

Proof. Surely, exactly one of the following possibilities must occur: (a') $\mathrm{dist}_T(v_1, v_2) = 0$; (b') $\mathrm{dist}_T(v_1, v_2) = 1$; (c') $\mathrm{c}(T[\mathrm{Ext}_{T,Y}]) = 1$ while $\mathrm{dist}_T(v_1, v_2) > 1$; (d') $\mathrm{c}(T[\mathrm{Ext}_{T,Y}]) > 1$, and hence $\mathrm{dist}_T(v_1, v_2) > 1$.

Case (c') corresponds precisely to case (c).

Let $\mathcal{U} = \{U \subseteq \mathrm{V}_T : U \mid W \in \widehat{\mathcal{S}}_{T,2}, Y = U \cap \mathrm{L}_T\}$. By definition, $\mathrm{Ext}_{T,Y}$ is the inclusion-minimal element in $\mathcal{U}$. This means that it cannot happen that $v_1v_2 \in \mathrm{E}_T$ and

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

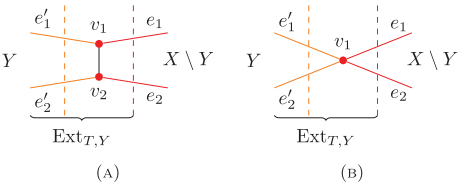


Figure 7. Excluding two possibilities in the proof of Lemma 4.5. A) $\widehat{\mathbf{S}}_{T,\{e_1,e_2\}} = U \mid W$, $\widehat{\mathbf{S}}_{T,\{e_1',e_2'\}} = U' \mid W'$, $Y \subseteq U' \subseteq U' \cup \{v_1, v_2\} = U$. B) $\widehat{\mathbf{S}}_{T,\{e_1,e_2\}} = U \mid W$, $\widehat{\mathbf{S}}_{T,\{e_1',e_2'\}} = U' \mid W'$, $Y \subseteq U' \subseteq U' \cup \{v_1\} = U$.

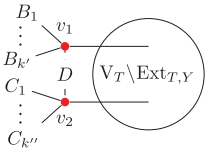


Figure 8. Distributions of B_i, C_i and D as described in Convention 4.6.

$\deg_T(v_1) = \deg_T(v_2) = 3$, as this would imply $\mathrm{Ext}_{T,Y} \setminus \{v_1, v_2\} \in \mathcal{U}$; see Fig. 7(A). We can also exclude the possibility of $v_1 = v_2$ and $\deg_T(v_1) = 4$, as this would give $\mathrm{Ext}_{T,Y} - v_1 \in \mathcal{U}$; see Fig. 7(B). Since $\mathbf{S}_{T,\{v_1u_1,v_2u_2\}} = Y \mid (\mathrm{L}_T \setminus Y) \in \mathcal{P}_{T,2}$, Corollary 2.6 asserts that neither $v_1 = v_2$ and $\deg_T(v_1) = 3$ nor $u_1 = u_2$ and $\deg_T(u_1) = 3$ is valid. Therefore, we now find that cases (a'), (b') and (d') are equivalent to (a), (b) and (d), respectively. □

Convention 4.6. *For the statements and proofs of Lemmas 4.4, 4.7 and Corollary 4.8, we introduce the following notations. Assume that T is a phylogenetic tree, $Y \mid (\mathrm{L}_T \setminus Y) \in \mathcal{P}_{T,2}$, and $|Y| \geqslant 3$. Let $\partial_T(\mathrm{Ext}_{T,Y}) = \{v_1u_1, v_2u_2\}$, where $\{v_1, v_2\} \subseteq \mathrm{Ext}_{T,Y}$ and $\deg_T(v_1) \geqslant \deg_T(v_2)$. Let $B_1, \ldots, B_{k'}$ be all distinct sets of the form $\mathrm{L}_T \cap \mathrm{C}_{T-v_1}[u]$, where $u \in \mathrm{N}_T(v_1) \setminus \{u_1, u_2\}$ and $v_2 \notin \mathrm{C}_{T-v_1}[u]$; let $C_1, \ldots, C_{k''}$ be all distinct sets of the form $\mathrm{L}_T \cap \mathrm{C}_{T-v_2}[u]$, where $u \in \mathrm{N}_T(v_2) \setminus \{u_1, u_2\}$ and $v_1 \notin \mathrm{C}_{T-v_2}[u]$. Let $B \doteq \bigcup_{i\in[k']} B_i$, $C \doteq \bigcup_{i\in[k'']} C_i$ and $D \doteq Y \setminus (B \cup C \cup \{v_1, v_2\})$. Note that $B \neq \emptyset$; $B = C$ if and only if $v_1 = v_2$; $C = \emptyset$ exactly when $\deg_T(v_2) = 1$; $D = \emptyset$ exactly when $T[\mathrm{Ext}_{T,Y}]$ is not connected or $\mathrm{dist}_T(v_1, v_2) \leq 1$. See Fig. 8 for an illustration.*

Lemma 4.7. *Let T be a phylogenetic tree, $Y \in \binom{\mathrm{L}_T}{\geqslant 3}$ and $Y \mid (\mathrm{L}_T \setminus Y) \in \mathcal{P}_{T,2}$. Then the following statements hold, where we use the convention in Convention 4.6.*

(1) *If $v_1 = v_2$ and $\deg_T(v_1) \geqslant 5$, then $\Gamma_{\mathcal{P}_{T,2}}(Y) = \{B_i \cup B_j : \{i, j\} \in \binom{[k']}{2}\}$.*
(2) *If $v_1v_2 \in \mathrm{E}_T$ and $\deg_T(v_2) \geqslant 4$, then $\Gamma_{\mathcal{P}_{T,2}}(Y) = \{B_i \cup C_j : i \in [k'], j \in [k'']\} \cup \{B, C\}$.*
(3) *If $v_1v_2 \in \mathrm{E}_T$ and $\deg_T(v_1) > \deg_T(v_2) = 3$, then $\Gamma_{\mathcal{P}_{T,2}}(Y) = \{B_i \cup C : i \in [k']\} \cup \{B\}$.*
(4) *If $T[\mathrm{Ext}_{T,Y}]$ is connected and $\mathrm{dist}_T(v_1, v_2) > 1$, then $\Gamma_{\mathcal{P}_{T,2}}(Y) = \{B_i \cup C_j : i \in [k'], j \in [k'']\} \cup \{B \cup D, C \cup D\}$.*

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

(5) *If* $\deg_T(v_2) = 1$, *then*

$$\Gamma_{\mathcal{P}_{T,2}}(Y) = \begin{cases} \{B_i + v_2 : \ i \in [k']\} \cup \{B - x : \ x \in B \setminus \mathrm{N}_T(v_1)\}, & \text{if } \deg_T(v_1) = 3, \\ \{B_i + v_2 : \ i \in [k']\} \cup \{B - x : \ x \in B\}, & \text{if } \deg_T(v_1) \geqslant 4. \end{cases}$$

(6) *If* $\mathrm{c}(T[\mathrm{Ext}_{T,Y}]) > 1$ *and* $\deg_T(v_2) \geqslant 3$, *then* $\Gamma_{\mathcal{P}_{T,2}}(Y) = \{B_i \cup C : \ i \in [k']\} \cup \{B \cup C_i : \ i \in [k'']\}$.

Proof. Let $V = \mathrm{V}_T$, $X = \mathrm{L}_T$ and $W = \mathrm{Ext}_{T,Y}$. Use the notation Δ for $\{Y' \subsetneq Y : Y' \mid (X \setminus Y') \in \mathcal{P}_{T,2}\}$. Note that $\Gamma_{\mathcal{P}_{T,2}}(Y)$ consists of all the inclusion-maximal elements of Δ. Let $Y' \in \Delta$ and $W' = \mathrm{Ext}_{T,Y'}$. Since T is a tree, deleting two edges from T results in exactly three connected components in the remaining graph. Therefore, we must have

$$\{\mathrm{c}(T[W']), \mathrm{c}(T - W')\} = [2]. \tag{4}$$

CASE 1. $T[W]$ is connected, and either $v_1 \neq v_2$ or $\deg_T(v_1) \geqslant 5$. This includes cases (1), (2), (3) and (4).

Since $T[W]$ is connected, both B and C are non-empty. This gives

$$B \neq \emptyset, C \neq \emptyset, Y = B \cup C \cup D, \tag{5}$$

and allows us to define $\mathcal{U}_1 \doteq \{B_i \cup C_j : \ i \in [k'], j \in [k''], B_i \neq C_j\} \subseteq 2^Y$. We claim that

$$\mathcal{U}_1 = \{Z \in \Gamma_{\mathcal{P}_{T,2}}(Y) : \ Z \cap B \neq \emptyset, Z \cap C \neq \emptyset\}. \tag{6}$$

Keeping in mind the validity of Eqs. (5) and (6), let us prove statements (1), (2), (3) and (4).

For statement (1), we are reduced to proving $D = \emptyset$ and $\{B_i : \ i \in [k']\} = \{C_j : \ j \in [k'']\}$. These follow directly from $v_1 = v_2$.

For statement (2), we only need to check that $D = \emptyset$, $k', k'' \geqslant 2$, and both B and C belong to Δ. The first and second conditions follow directly from $v_1v_2 \in \mathrm{E}_T$ and $\deg_T(v_1) \geqslant \deg_T(v_2) \geqslant 4$, respectively. Because $\deg_T(v_1) \geqslant \deg_T(v_2) \geqslant 4$, by Corollary 2.6, $B \mid (X \setminus B) = \mathbf{S}_{T,\{v_1v_2,v_1u_1\}} \in \mathcal{P}_{T,2}$ and $C \mid (X \setminus C) = \mathbf{S}_{T,\{v_1v_2,v_2u_2\}} \in \mathcal{P}_{T,2}$, as desired.

To prove (3), it suffices to verify that $D = \emptyset$, $k' > k'' = 1$ and $B \in \Delta$. The first two are direct from $v_1v_2 \in \mathrm{E}_T$ and $\deg_T(v_1) > \deg_T(v_2) = 3$. By Corollary 2.6, the last one follows from $\deg_T(v_1) \geqslant 4$ and $B \mid (X \setminus B) = \mathbf{S}_{T,\{v_1v_2,v_1u_1\}}$.

As for statement (4), it remains to show that $D \neq \emptyset$ and that both $B \cup D$ and $C \cup D$ belong to Δ. The former follows directly from $\mathrm{c}(T[W]) = 1$ and $\mathrm{dist}_T(v_1, v_2) > 1$. To show the latter, by symmetry, it suffices to prove that $(B \cup D) \mid (X \setminus (B \cup D)) \in \mathcal{P}_{T,2}$. According to Convention 4.6, there exists $w_2' \in \mathrm{N}_T(v_2)$ such that $D = X \cap \mathrm{C}_{T-v_1-v_2}[w_2']$. Note that v_1, v_2, u_1, and w_2' are pairwise distinct. Then, by Corollary 2.6, $(B \cup D) \mid (X \setminus (B \cup D)) = \mathbf{S}_{T,\{v_1u_1,v_2w_2'\}} \in \mathcal{P}_{T,2}$, as desired.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

To complete the proof in this case, we need to verify Eq. (6). To obtain Eq. (6), we will deduce that $\mathcal{U}_1 \subseteq \Delta$, and that

$$Y' \subseteq B_i \cup C_j \tag{7}$$

whenever $i \in [k']$ and $j \in [k'']$ are two elements satisfying $B_i \neq C_j$, $Y' \cap B_i \neq \emptyset$ and $Y' \cap C_j \neq \emptyset$.

Pick $B_i \cup C_j \in \mathcal{U}_1$. On the one hand, $B_i \cup C_j \neq Y$. Otherwise, either $v_1 = v_2$ and $\deg_T(v_1) = 4$ or $v_1v_2 \in \mathrm{E}_T$ and $\deg_T(v_1) = \deg_T(v_2) = 3$, violating Lemma 4.5. On the other hand, by Convention 4.6, there are distinct $w_i \in \mathrm{N}_T(v_1) - v_2$ and $w_j \in \mathrm{N}_T(v_2) - v_1$ such that $B_i = X \cap \mathrm{C}_{T-v_1}[w_i]$ and $C_j = X \cap \mathrm{C}_{T-v_2}[w_j]$. As either $v_1 \neq v_2$ or $\deg_T(v_1) \geqslant 5$, by Corollary 2.6, $(B_i \cup C_j) \mid (X \setminus (B_i \cup C_j)) = \mathbf{S}_{T,\{v_1w_i,v_2w_j\}}$ belongs to $\mathcal{P}_{T,2}$, as desired. This proves $\mathcal{U}_1 \subseteq \Delta$.

If $v_1, v_2 \notin W'$, then, by Eq. (4), $\mathrm{c}(T[W'] - \{v_1, v_2\}) = \mathrm{c}(T[W']) \leq 2$. However, $Y' \setminus (B_i \cup C_j)$, $Y' \cap B_i$ and $Y' \cap C_j$ are pairwise non-reachable in $T[W'] - \{v_1, v_2\}$, while $Y' \cap B_i \neq \emptyset$ and $Y' \cap C_j \neq \emptyset$. This forces Eq. (7) to hold. Henceforth, it remains to show that $v_1, v_2 \notin W'$. Let us pick $x_0 \in Y \setminus Y'$, $x_1 \in X \cap \mathrm{C}_{T-v_1}[u_1]$ and $x_2 \in X \cap \mathrm{C}_{T-v_2}[u_2]$. Then x_0, x_1, and x_2 are pairwise non-reachable leaves in $T - \{v_1, v_2\}$ and thus in $T - (W' \cup \{v_1, v_2\})$, which means $\mathrm{c}(T - (W' \cup \{v_1, v_2\})) \geqslant 3$. We thus derive from Eq. (4) that at least one of v_1 and v_2 does not belong to W', say v_t. As a result, $Y' \cap B_i \neq \emptyset$ and $Y' \cap C_j \neq \emptyset$ are non-reachable in $T[W']$. It must happen that $\mathrm{c}(T[W']) \geqslant 2$. This, together with Eq. (4), implies that $\mathrm{c}(T - W') = 1$. As x_0 and x_{3-t} are leaves of $T - W'$ that are non-reachable in $T - v_{3-t}$, we conclude that $v_{3-t} \notin W'$, thus completing our proof in Case 1.

CASE 2. $T[W]$ is not connected. This includes cases (5) and (6).

Before verifying statements (5) and (6), we establish two relevant facts. The first fact is that

$$Y' \subseteq B_i \cup (Y \setminus B) \tag{8}$$

holds for some $i \in [k']$ whenever none of $B \setminus Y'$, $Y' \cap B$ and $Y' \setminus B$ is empty. Observe that, for distinct $i, j \in [k']$, $Y' \cap B_i$, $Y' \cap B_j$ and $Y' \setminus B$ are pairwise non-reachable in $T[W' - v_1]$. Henceforth, if $v_1 \notin W'$, we can infer from $\mathrm{c}(T[W' - v_1]) = \mathrm{c}(T[W']) \leq 2$ and $Y' \setminus B \neq \emptyset$ that Eq. (8) holds, as desired. Our task is thus to deduce $v_1 \notin W'$. Choose $x_0 \in B \setminus Y'$ arbitrarily. Since $\{x_0\} \cup (X \setminus Y) \subseteq X \setminus Y' = \mathrm{L}_{T-W'}$ and $v_1 \in [\{x_0\}, X \setminus Y]_T$, we only need to check that $\mathrm{c}(T - W') = 1$. By Lemma 4.5, it holds that either $u_1 \neq u_2$ or $\deg_T(u_1) \geqslant 4$. Hence there exists $\{w_1, w_2\} \in \binom{\mathrm{N}_T(u_1) \setminus \{v_1, v_2\}}{2}$. Pick $x_1 \in X \cap \mathrm{C}_{T-u_1}[w_1]$ and $x_2 \in X \cap \mathrm{C}_{T-u_1}[w_2]$. Observe that x_0, x_1, and x_2 are pairwise non-reachable in $T - u_1$ and thus in $T - W' - u_1$. Consequently, $\mathrm{c}(T - W' - u_1) \geqslant 3$. It then follows from $\mathrm{c}(T - W') \leq 2$ that $u_1 \notin W'$. This, along with $Y' \cap B \neq \emptyset$ and $Y' \setminus B \neq \emptyset$, implies that $\mathrm{c}(T[W']) \geqslant 2$. We can thus apply Eq. (4) to obtain $\mathrm{c}(T - W') = 1$, as desired.

We now want to deduce

$$\mathcal{U}_2 = \{Z \in \Gamma_{\mathcal{P}_{T,2}}(Y) : \ Z \nsubseteq B, B \nsubseteq Z, Z \cap B \neq \emptyset\}, \tag{9}$$

where $\mathcal{U}_2 \doteq \{B_i \cup (Y \setminus B) : i \in [k']\}$. If $\mathcal{U}_2 \subseteq \Delta$, then it is clear that Eq. (8) implies Eq. (9). So, what remains to prove is $\mathcal{U}_2 \subseteq \Delta$. As $T[W]$ is not connected, it holds that

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

$k' = \deg_T(v_1) - 1 \geqslant 2$. Hence $Y \notin \mathcal{U}_2$. Let $Z = B_i \cup (Y \setminus B)$ for some $i \in [k']$. To finish the proof of $\mathcal{U}_2 \subseteq \Delta$, we aim to show that $Z \mid (X \setminus Z) \in \mathcal{P}_{T,2}$. Following Convention 4.6, there is a $w_i \in \mathrm{N}_T(v_1) - u_2$ such that $B_i = X \cap \mathrm{C}_{T-v_1}[w_i]$. By Corollary 2.6, $Z \mid (X \setminus Z) = \mathbf{S}_{T,\{v_1w_i,v_2u_2\}} \in \mathcal{P}_{T,2}$, as desired.

In the sequel, making use of the conditions $k' \geqslant 2$ and Eq. (9) implicitly, we complete the proof of cases (5) and (6).

Let us start with statement (5). From $\deg_T(v_2) = 1$ it follows that $Y \setminus B = \{v_2\}$. Thus, for any set Z, $B \subsetneq Z \subseteq Y$ implies $Z = Y$; either $B = Z$ or $Z \subseteq Y \setminus B$ implies $Z \mid (X \setminus Z) \notin \mathcal{P}_{T,2}$. These observations indicate that $\{Z \subseteq Y : \ B \subseteq Z \text{or } Z \cap B = \emptyset\}$ does not intersect with $\Gamma_{\mathcal{P}_{T,2}}(Y)$. Therefore, we only need to check that

$$\{Z \in \Gamma_{\mathcal{P}_{T,2}}(Y) : \ Z \subsetneq B\} = \begin{cases} \{B - x : \ x \in B \setminus \mathrm{N}_T(v_1)\}, & \text{if } \deg_T(v_1) = 3, \\ \{B - x : \ x \in B\}, & \text{if } \deg_T(v_1) \geqslant 4. \end{cases} \tag{10}$$

The proof of Eq. (10) is divided into two cases.

(1) Either $\deg_T(v_1) \geqslant 4$ or $B \cap \mathrm{N}_T(v_1) = \emptyset$.
It is sufficient to verify that $B - x \in \Delta$ for all $x \in B$. Consider an arbitrary $x \in B$, and let e be the edge incident to x in T. Whether $\deg_T(v_1) \geqslant 4$ or $B \cap \mathrm{N}_T(v_1) = \emptyset$, the edges v_1u_1 and e cannot share a common endpoint of degree three in T. Consequently, by Corollary 2.6, $(B - x) \mid (X \setminus (B - x)) = \mathbf{S}_{T,\{v_1u_1,e\}} \in \mathcal{P}_{T,2}$, namely $B - x \in \Delta$.

(2) $\deg_T(v_1) = 3$ and $B \cap \mathrm{N}_T(v_1) \neq \emptyset$.
From $\deg_T(v_1) = 3$ we have $|B \cap \mathrm{N}_T(v_1)| \leq 2$. If $|B \cap \mathrm{N}_T(v_1)| = 2$, then $|Y| = |B| + 1 = 3$, and hence Eq. (10) follows from Corollary 2.7. Now, assume that $B \cap \mathrm{N}_T(v_1)$ contains only one leaf x_0. Let $\mathrm{N}_T(v_1) = \{u_1, x_0, w\}$. Then, $(B - x_0) \mid (X \setminus (B - x_0)) = \mathbf{S}_{T,\{v_1w\}} \notin \mathcal{P}_{T,2}$. Thus, a proper subset of B belonging to Δ must be a subset of $B - x$ for some $x \in B - x_0$. To show Eq. (10), it remains to verify that $B - x \in \Delta$ for all $x \in B - x_0$, which follows from similar arguments in the previous case.

It is time to prove statement (6). From $\deg_T(v_2) \geqslant 3$ and $\mathrm{c}(T[W]) > 1$, it follows that $k'' \geqslant 2$ and $C = Y \setminus B \neq \emptyset$. Since $C = Y \setminus B$, we obtain $\mathcal{U}_2 = \{B_i \cup C : \ i \in [k']\}$. Let $\mathcal{U}_2' \doteq \{C_j \cup B : \ j \in [k'']\}$. Our goal is to show that $\Gamma_{\mathcal{P}_{T,2}}(Y) = \mathcal{U}_2 \cup \mathcal{U}_2'$. By switching the roles of B and C and replacing $\{B_i : \ i \in [k']\}$ with $\{C_j : \ j \in [k'']\}$, we know from Eq. (9) that $\mathcal{U}_2' = \{Z \in \Gamma_{\mathcal{P}_{T,2}}(Y) : \ Z \nsubseteq C, C \nsubseteq Z, Z \cap C \neq \emptyset\}$. Take $Z \in \Gamma_{\mathcal{P}_{T,2}}(Y) \setminus \mathcal{U}_2'$. Then $C \subsetneq Z \subsetneq Y$. This implies $Z \nsubseteq B$, $B \nsubseteq Z$, and $Z \cap B \neq \emptyset$. Thus , $Z \in \mathcal{U}_2$ follows, as was to be shown. □

For each family of sets $\mathcal{F}$, designate by $\varrho_2(\mathcal{F})$ the set family $\{F \cap F' : \ \{F, F'\} \in \binom{\mathcal{F}}{2}\} \setminus \{\emptyset\}$. For any k-partition π with $k \geqslant 3$, it is noteworthy that $\varrho_2(\mathcal{H}_{\pi,2}) = \pi$.

Proof of Lemma 4.4 Let $V = \mathrm{V}_T$ and $X = \mathrm{L}_T$. When (1) holds, we necessarily have $|Y| \geqslant 3$ and $Y \mid (X \setminus Y) \in \mathcal{P}_{T,2}$. If (2) holds, the definition of a 2-special configuration implies $|Y| \geqslant 3$, and, in view of Corollary 2.6, $Y \mid (X \setminus Y) \in \mathcal{P}_{T,2}$. Henceforth, hereafter

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

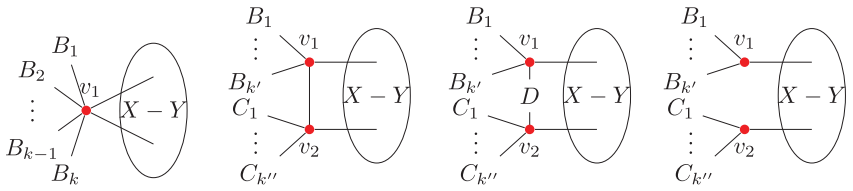


Figure 9. Different cases in the proof of Lemma 4.4. A) Case 1 B) Case 2 C) Case 3 D) Case 4

we shall assume that $|Y| \geqslant 3$, $Y \mid (X \setminus Y) \in \mathcal{P}_{T,2}$, $W \doteq \mathrm{Ext}_{T,Y}$, and adopt the convention from Convention 4.6 regarding $k', k'', v_1, v_2, B, C, D$, etc.

Let $\Pi \doteq \varrho_2(\Gamma_{\mathcal{P}_{T,2}}(Y))$. It is clear that condition (1) is equivalent to the following condition (1'):

(1') Π is a k-partition of Y, and $\Gamma_{\mathcal{P}_{T,2}}(Y) = \mathcal{H}_{\Pi,2}$.

According to Lemma 4.5, there are four cases to consider. We are going to establish the equivalence between (1') ((1)) and (2) in each of these cases.

CASE 1. $v_1 = v_2$ and $\deg_T(v_1) \geqslant 5$, as given in Lemma 4.5 (a) and shown in Fig. 9(A).

In this case, (T, Y) is a 2-special configuration of type A and $\phi_T(Y) = k'$. Hence, (2) amounts to requiring $k = k'$.

By Lemma 4.7 (1), $\Gamma_{\mathcal{P}_{T,2}}(Y) = \left\{B_i \cup B_j : \ \{i,j\} \in \binom{[k']}{2}\right\}$, which states that Y is a $(k', 2)$-set with respect to $\mathcal{P}_{T,2}$. Therefore, (1) also holds if and only if $k = k'$.

CASE 2. $v_1v_2 \in \mathrm{E}_T$ and $\deg_T(v_1) = \max\{\deg_T(v_1), \deg_T(v_2)\} \geqslant 4$. This corresponds to Lemma 4.5 (b), as shown in Fig. 9(B).

In this case, $D = \emptyset$, $k' = \deg_T(v_1) - 2 \geqslant 2$, $k'' = \deg_T(v_2) - 2$, and $\phi_T(Y) = k' + k''$.

CASE 2.1. $\deg_T(v_2) \geqslant 4$, which means that $k'' \geqslant 2$.

In this case, (T, Y) is a 2-special configuration if and only if its type is B. Thus, (2) is equivalent to $k' = k'' = 2$ and $k = 4$.

It follows from Lemma 4.7 (2) that

$$\Pi = \{B_i : \ i \in [k']\} \cup \{C_j : \ j \in [k'']\},$$

which is a $(k' + k'')$-partition of Y. Having a look into Lemma 4.7 (2) again, we find that $\Gamma_{\mathcal{P}_{T,2}}(Y) = \mathcal{H}_{\Pi,2}$ if and only if $k' = k'' = 2$. Combining these two facts yields that (1') is also equivalent to $k' = k'' = 2$ and $k = 4$.

CASE 2.2. $\deg_T(v_2) = 3$, equivalently $k'' = 1$. In this case, (T, Y) is a 2-special configuration if and only if its type is C. Therefore, (2) is equivalent to $k = k' + 1 = 3$.

In view of Lemma 4.7 (3), we have

$$\Pi = \{B_i : \ i \in [k']\} \cup \{C\},$$

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

which is a $(k'+1)$-partition of Y. Comparing it with Lemma 4.7 (3), we know that $\Gamma_{\mathcal{P}_{T,2}}(Y) = \mathcal{H}_{\Pi,2}$ if and only if $k' = 2$. Thus, (1') is also equivalent to $k = k'+1 = 3$.

CASE 3. $T[W]$ is connected and $\mathrm{dist}_T(v_1, v_2) > 1$, as given in Lemma 4.5 (c) and shown in Fig. 9(C).

In this case, $k' = \deg_T(v_1) - 2 \geqslant 1$, $k'' = \deg_T(v_2) - 2 \geqslant 1$ and $\phi_T(Y) = k' + k'' + 1$. Additionally, (T, Y) is a 2-special configuration if and only if it is of type D, which is equivalent to $\deg_T(v_1) = \deg_T(v_2) = 3$. Hence, (2) holds if and only if $k' = k'' = 1$ and $k = 3$.

By Lemma 4.7 (4), it happens $\Gamma_{\mathcal{P}_{T,2}}(Y) = \{B_i \cup C_j : i \in [k'], j \in [k'']\} \cup \{B \cup D, C \cup D\}$. Accordingly,

$$\Pi = \{B_i : i \in [k']\} \cup \{C_j : j \in [k'']\} \cup \{D\},$$

which is a $(k' + k'' + 1)$-partition of Y. We see that $\Gamma_{\mathcal{P}_{T,2}}(Y) = \mathcal{H}_{\Pi,2}$ if and only if $k' = k'' = 1$. Thus, (1') is also equivalent to $k' = k'' = 1$ and $k = k' + k'' + 1 = 3$.

CASE 4. $T[W]$ is not connected, and either $u_1 \neq u_2$ or $\deg_T(u_1) \geqslant 4$. This corresponds to the case of Lemma 4.5 (d), as illustrated in Fig. 9(D).

CASE 4.1. $v_2 \in \mathrm{L}_T$, namely $\deg_T(v_2) = 1$.

Note that $Y = B + v_2$. We thus derive from $|Y| \geqslant 3$ that $\phi_T(Y) = \deg_T(v_1) = k' + 1 \geqslant 3$.

CASE 4.1.1. Suppose $\deg_T(v_1) = 3$, and hence (T, Y) is not a 2-special configuration and $k' = 2$. In this case, statement (2) cannot hold. By virtue of $k' = 2$, Lemma 4.7 (5) yields that $\{B_i + v_2 : i \in [2]\} \subseteq \Gamma_{\mathcal{P}_{T,2}}(Y)$ and $\{v_2\} \in \Pi$. Thus, if $\Gamma_{\mathcal{P}_{T,2}}(Y) = \mathcal{H}_{\Pi,2}$, then $\Pi = \{\{v_2\}, B_1, B_2\}$ and hence $B = B_1 \cup B_2 \in \mathcal{H}_{\Pi,2} \setminus \Gamma_{\mathcal{P}_{T,2}}(Y)$. Therefore, statement (1') cannot hold either.

CASE 4.1.2. Suppose $\deg_T(v_1) \geqslant 4$. Whenever (T, Y) is a 2-special configuration, its type must be E. By further taking into account that $|Y| = |B| + 1 \geqslant k' + 1 = \phi_T(Y) = \deg_T(v_1) \geqslant 4$, (2) holds if and only if $k' = |B| = 3$.

From Lemma 4.7 (5) and $|B| \geqslant 3$, it follows that

$$\Pi = \binom{B}{|B|-2} \cup \left(\bigcup_{i:|B_i| \geqslant 2, i \in [k']} \binom{B_i}{|B_i|-1} \right) \cup \{B_i : i \in [k']\} \cup \{\{v_2\}\}.$$

As $\binom{B}{|B|-2} \subseteq \Pi$, Π is a k-partition for $k \geqslant 3$ only if $|B| = 3$. If $|B| = 3$, then $k' = 3$ and by Lemma 4.7 (5) we see that (1') holds. Thus (1') is also equivalent to $k' = |B| = 3$.

CASE 4.2. $\deg_T(v_2) \geqslant 3$.

In this case, $k' = \deg_T(v_1) - 1 \geqslant 2$, $k'' = \deg_T(v_2) - 1 \geqslant 2$ and (2) cannot hold. By Lemma 4.7 (6), $B_i \cup C_j \in \Pi$ for each $i \in [k']$ and $j \in [k'']$. Thus, Π is not a partition, implying that (1') cannot hold either. □

Corollary 4.8. *Let T be a phylogenetic tree and let $Y \subseteq \mathrm{L}_T$. Suppose (T, Y) is a 2-special configuration and $\varrho_2(\Gamma_{\mathcal{P}_{T,2}}(Y)) = \{Y_1, \ldots, Y_k\} \in \binom{2^Y}{k}$. Then the connected*

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

components of $T[\mathrm{Ext}_{T,Y}] - (\{v_1, v_2\} \setminus \mathrm{L}_T)$ *can be enumerated as* $V_1, \ldots, V_k$ *such that* $Y_i = V_i \cap \mathrm{L}_T$ *for all* $i \in [k]$.

Proof. As (T, Y) is a 2-special configuration and $\varrho_2(\Gamma_{\mathcal{P}_{T,2}}(Y)) = \{Y_1, \ldots, Y_k\}$, from Lemma 4.4 it follows that Y is a $(k, 2)$-set with respect to $\mathcal{P}_{T,2}$ and $k = \phi_T(Y) = \mathrm{c}(T[\mathrm{Ext}_{T,Y}] - (\{v_1, v_2\} \setminus \mathrm{L}_T))$. Thus we can enumerate the connected components of $T[\mathrm{Ext}_{T,Y}] - (\{v_1, v_2\} \setminus \mathrm{L}_T)$ as $V_1, \ldots, V_k$.

Adopt notations from Convention 4.6. We see from Fig. 6 that

$$\{V_i \cap \mathrm{L}_T : i \in [k]\} = \begin{cases} \{B_i : i \in [k']\} \cup \{C_j : \ j \in [k'']\}, & \text{if } (T,Y) \text{ is of type } \mathtt{A}, \mathtt{B} \text{ or } \mathtt{C}, \\ \{B, C, D\}, & \text{if } (T,Y) \text{ is of type } \mathtt{D}, \\ \big\{\{y\} : \ y \in Y\big\}, & \text{if } (T,Y) \text{ is of type } \mathtt{E}. \end{cases}$$

According to Lemma 4.7, we have

$$\varrho_2(\Gamma_{\mathcal{P}_{T,2}}(Y)) = \begin{cases} \{B_i : i \in [k']\} \cup \{C_j : \ j \in [k'']\}, & \text{if } (T,Y) \text{ is of type } \mathtt{A}, \mathtt{B} \text{ or } \mathtt{C}, \\ \{B, C, D\}, & \text{if } (T,Y) \text{ is of type } \mathtt{D}, \\ \big\{\{y\} : \ y \in Y\big\}, & \text{if } (T,Y) \text{ is of type } \mathtt{E}. \end{cases}$$

Therefore, it follows that $\{V_i \cap \mathrm{L}_T : \ i \in [k]\} = \varrho_2(\Gamma_{\mathcal{P}_{T,2}}(Y)) = \{Y_1, \ldots, Y_k\}$. □

Remark 4.9. Let T be a phylogenetic tree and let $Y \subseteq \mathrm{L}_T$ be a $(k, 2)$-set with respect to $\mathcal{P}_{T,2}$. Suppose $\varrho_2(\Gamma_{\mathcal{P}_{T,2}}(Y)) = \{Y_1, \ldots, Y_k\}$, $\partial_T(\mathrm{Ext}_{T,Y}) = \{v_1u_1, v_2u_2\}$, $\{v_1, v_2\} \subseteq \mathrm{Ext}_{T,Y}$ and $\deg_T(v_1) \geqslant \deg_T(v_2)$. By Lemma 4.4, (T, Y) is a 2-special configuration.

(1) If (T, Y) is of type B or C, then by Corollary 4.8, we can let $\{Y_1, Y_2\} = \{\mathrm{L}_T \cap \mathrm{C}_{T-\{v_1,v_2\}}[v] : \ v \in \mathrm{N}_T(v_1) - u_1\}$ and $\{Y_3, \ldots, Y_k\} = \{\mathrm{L}_T \cap \mathrm{C}_{T-\{v_1,v_2\}}[v] : \ v \in \mathrm{N}_T(v_2) - u_2\}$. See Figs. 6(B) and 6(C).
(2) If (T, Y) is of type D, then by Corollary 4.8, we can let $\{Y_1, Y_2\} = \{\mathrm{L}_T \cap \mathrm{C}_{T-\{v_1,v_2\}}[v] : \ v \in \mathrm{N}_T(v_1) - u_1\}$ and $\{Y_2, Y_3\} = \{\mathrm{L}_T \cap \mathrm{C}_{T-\{v_1,v_2\}}[v] : \ v \in \mathrm{N}_T(v_2) - u_2\}$. See Fig. 6(D).
(3) If (T, Y) is of type E, then we can let $Y_1 \cup Y_2 \cup Y_3 = \mathrm{L}_T \cap \mathrm{N}_T(v_1)$ and $Y_4 = \{v_2\}$. For distinct $i, j \in [4]$, let $Y_{ij} = \mathrm{L}_T \setminus (Y_i \cup Y_j)$. See Fig. 6(E).
(4) If the type of (T, Y) is one of A, B, C and D, then let $Y_{k+1} = \mathrm{L}_T \cap \mathrm{C}_{T-v_1}[u_1]$ and $Y_{k+2} = \mathrm{L}_T \cap \mathrm{C}_{T-v_2}[u_2]$. For distinct $i, j \in [k+2]$, let $Y_{ij} = \mathrm{L}_T \setminus (Y_i \cup Y_j)$. See Figs. 6(A), 6(B), 6(C) and 6(D).

Corollary 4.10. *Let* T *be a phylogenetic tree, and let* $Y \subseteq \mathrm{L}_T$. *Suppose* (T, Y) *is a 2-special configuration whose type is not* E, *and adopt the notations from Remark 4.9. Then the following two statements hold for all* $\{i, j\} \in \binom{[k+2]}{2}$.

(1) *If* $v_1, v_2 \in [Y_i, Y_j]_T$, *then* Y_{ij} *is a* $(k, 2)$*-set with respect to* $\mathcal{P}_{T,2}$.
(2) *If the type of* (T, Y) *is* A, *then* $\varrho_2(\Gamma_{\mathcal{P}_{T,2}}(Y_{ij})) = \{Y_p : \ p \in [k+2] \setminus \{i, j\}\}$.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

Proof. (1). By Corollary 4.8 and Remark 4.9, there exist $w_1, w_2 \in \mathrm{N}_T(v_1) \cup \mathrm{N}_T(v_2)$ such that $Y_i = \mathrm{L}_T \cap \mathrm{C}_{T-\{v_1,v_2\}}[w_1]$ and $Y_j = \mathrm{L}_T \cap \mathrm{C}_{T-\{v_1,v_2\}}[w_2]$. Since $v_1, v_2 \in [Y_i, Y_j]_T$, it follows that $v_1, v_2 \in [w_1, w_2]_T$. Thus, by symmetry we may assume that $w_1 \in \mathrm{N}_T(v_1)$ and $w_2 \in \mathrm{N}_T(v_2)$. Denote $\mathrm{C}_{T-\{w_1,w_2\}}[v_1]$ by W. Then $Y_{ij} = W \cap \mathrm{L}_T$, $\partial_T(W) = \{v_1w_1, v_2w_2\}$ and $\{v_1, v_2\} \subseteq W$. As v_1 and v_2 satisfy Eq. (2), by Lemma 2.11, W is exactly $\mathrm{Ext}_{T,Y_{ij}}$. Because $T[W]$ is connected, by the definitions of types `A`, `B`, `C` and `D`, (T, Y_{ij}) is a 2-special configuration of the same type as (T, Y). Then, by Lemma 4.4, Y_{ij} is a $(k, 2)$-set with respect to $\mathcal{P}_{T,2}$, as desired.

(2). From the above proof of (1), we know that (T, Y_{ij}) is a 2-special configuration of type `A`. Since $v_1 = v_2$ and $Y_{ij} \mid (Y_i \cup Y_j) = \mathbf{S}_{T,\{v_1w_1,v_2w_2\}}$, the claim follows directly from Corollary 4.8. □

4.2. Type `A` and type `E`

We begin with a definition of certain structures associated with a split system. It will soon become evident that these structures are designed to encode 2-special configurations of type `A` and type `E`.

Definition 4.11. *Let $\mathcal{S}$ be a split system on a finite set X and let $k \geqslant 3$ be an integer. We use $\mathcal{Y}_k(\mathcal{S})$ to denote the set family consisting of all $(k, 2)$-sets with respect to $\mathcal{S}$. Let $\mathcal{Y}_k^{\mathtt{A}}(\mathcal{S})$ consist of all $Y \in \mathcal{Y}_k(\mathcal{S})$ such that $X \setminus (Y_i \cup Y_j) \in \mathcal{Y}_k(\mathcal{S})$ for all distinct $i, j \in [k]$, where $\{Y_1, \ldots, Y_k\} = \varrho_2(\Gamma_{\mathcal{S}}(Y))$. Let $\mathcal{Y}^{\mathtt{A}}(\mathcal{S}) = \bigcup_{i \geqslant 3} \mathcal{Y}_i^{\mathtt{A}}(\mathcal{S})$. Let $\mathcal{Y}^{claw}(\mathcal{S})$ consist of all $\{x, y, z\} \in \binom{X}{3}$ for which there exists $\{w_1, w_2, w_3\} \in \binom{X\setminus\{x,y,z\}}{3}$ such that $\{x, y, z, w_i\} \in \mathcal{Y}_4(\mathcal{S})$ for all $i \in [3]$.*

Let T be a phylogenetic tree, and let $k \geqslant 5$. It follows directly from Lemma 4.4 that $Y \in \mathcal{Y}_k^{\mathtt{A}}(\mathcal{P}_{T,2})$ only if (T, Y) is a 2-special configuration of type `A`. However, a stronger version of this statement is required, which we establish in Lemma 4.13.

Lemma 4.12. *Let T be a phylogenetic tree and let $Y \subseteq \mathrm{L}_T$ be a $(k, 2)$-set with respect to $\mathcal{P}_{T,2}$. Adopt the convention in Remark 4.9.*

(1) *Suppose the type of (T, Y) is `B`. If both Y_{12} and Y_{34} are $(4, 2)$-sets with respect to $\mathcal{P}_{T,2}$, then $T \cong \widetilde{\mathtt{A}}_0$.*
(2) *If the type of (T, Y) is `C` or `D`, then Y_{12} is not a $(3, 2)$-set with respect to $\mathcal{P}_{T,2}$.*
(3) *Suppose the type of (T, Y) is `E`. For any $\{i, j\} \in \binom{[3]}{2}$, Y_{ij} is a $(4, 2)$-set with respect to $\mathcal{P}_{T,2}$ if and only if $T \cong \widetilde{\mathtt{A}}_0$.*

Proof. We start with (1). By Lemma 4.4, Y_{12} being a $(4, 2)$-set with respect to $\mathcal{P}_{T,2}$ implies that (T, Y) is a 2-special configuration. Clearly, $Y_{12} = \mathbf{S}_{T,\{v_1u_1,v_1v_2\}}$ while u_1 and v_2 satisfy Eq. (2). Thus, by Lemma 2.11, $\mathrm{Ext}_{T,Y_{12}} = \mathrm{C}_{T-v_1}[u_1] \cup \mathrm{C}_{T-v_1}[v_2]$. Since $T[\mathrm{Ext}_{T,Y_{12}}]$ is not connected, the only possible type of (T, Y_{12}) is `E`. The definition of type `E` requires that each Y_i, $i \in \{3, 4, 5, 6\}$, consists of a single element. Similarly, if Y_{34} is a $(4, 2)$-set with respect to $\mathcal{P}_{T,2}$, then all Y_i, $i \in \{1, 2, 5, 6\}$, have size one. In conclusion, if both Y_{12} and Y_{34} are $(4, 2)$-sets with respect to $\mathcal{P}_{T,2}$, then all Y_i, $i \in [6]$, are singleton sets, which implies that $T \cong \widetilde{\mathtt{A}}_0$.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

Now, we consider (2). Suppose the type of (T, Y) is C or D. Let w_2 be the unique vertex in $\mathrm{N}_T(v_2) \cap [v_1, v_2]_T$. Then $Y_{12} = \mathbf{S}_{T,\{v_1u_1, w_2v_2\}}$ while u_1 and v_2 satisfy Eq. (2). Hence, by Lemma 2.11, $\mathrm{Ext}_{T,Y_{12}} = \mathrm{C}_{T-v_1}[u_1] \cup \mathrm{C}_{T-w_2}[v_2]$. Note that $T[\mathrm{Ext}_{T,Y_{12}}]$ is not connected and $\deg_T(v_2) = 3$. By definition, (T, Y) is not a 2-special configuration. Consequently, by Lemma 4.4, Y_{12} is not a $(3, 2)$-set with respect to $\mathcal{P}_{T,2}$.

At last, we examine (3). By symmetry, suppose $\{i, j\} = [2]$. Denote the only leaf in Y_i by y_i for each $i \in [3]$. It is clear that $Y_{12} \mid \{y_1, y_2\} = \mathbf{S}_{T,\{v_1y_3, v_1u_1\}}$. Let $U \doteq \mathrm{V}_T \setminus \{y_1, y_2, v_1\}$. From $\widehat{\mathbf{S}}_{T,\{v_1y_3,v_1u_1\}} = U \mid \{y_1, y_2, v_1\}$, it follows that $\{y_3, v_2\} \subseteq \mathrm{Ext}_{T,Y_{12}} \subseteq U$. Since y_3 and v_2 are not reachable in $T[U]$, $T[\mathrm{Ext}_{T,Y_{12}}]$ is not connected. Therefore, by Lemma 4.4, Y_{12} is a $(4, 2)$-set with respect to $\mathcal{P}_{T,2}$ if and only if (T, Y_{12}) is a 2-special configuration of type E. The latter condition is equivalent to $\mathrm{L}_T \setminus \{y_1, y_2, y_3\}$ being a claw of T, i.e., $T \cong \widetilde{\mathtt{A}}_0$. □

Lemma 4.13. *Let T be a phylogenetic tree with $T \ncong \widetilde{\mathtt{A}}_0$. For any $Y \subseteq \mathrm{L}_T$, we have $Y \in \mathcal{Y}^{\mathtt{A}}(\mathcal{P}_{T,2})$ if and only if (T, Y) is a 2-special configuration of type A.*

Proof. The sufficiency follows from Corollary 4.10 (1), while the necessity follows from Lemma 4.12. □

We proceed to investigate how to encode the possible claw structure of a tree based on its score-2 split system.

Lemma 4.14. *Let T be a phylogenetic tree, and let $Y \subseteq \mathrm{L}_T$ be a $(4, 2)$-set with respect to $\mathcal{P}_{T,2}$. Suppose the type of (T, Y) is B and adopt the notations in Remark 4.9. Let $j \in \{5, 6\}$. If there exists a set $I \in \binom{[4]}{3}$ and a vertex $w \in Y_j$ such that $((\bigcup_{i\in I} Y_i)+w) \mid ((\bigcup_{i\in[6]\setminus I} Y_i) - w) \in \mathcal{P}_{T,2}$, then $Y_j = \{w\}$.*

Proof. By symmetry, it suffices to verify the result for $I = [3]$ and $j \in \{5, 6\}$. We only consider the case $I = [3]$ and $j = 5$, as the case $I = [3]$ and $j = 6$ can be verified with the same argument.

There exist $e_1, e_2 \in \mathrm{E}_T$ such that $((\bigcup_{i\in[3]} Y_i) + w) \mid ((\bigcup_{i\in\{4,5,6\}} Y_i) - w) = \mathbf{S}_{T,\{e_1,e_2\}}$. Let $T' = T - V_5$ and $w_4 \in V_4 \cap \mathrm{N}_T(v_2)$, where $V_i \doteq \mathrm{C}_{T-\{v_1,v_2\}}[Y_i]$ for $i = 4, 5$. Because v_2w_4 and v_2u_2 are not incident with a common vertex of degree 3 in T', combining Lemma 2.8 and Corollary 2.6 yields $\mathbf{S}_{T',\{e_1,e_2\}\cap \mathrm{E}_{T'}} = \mathbf{S}_{T,\{e_1,e_2\}}|_{\mathrm{L}_{T'}} = (Y_1 \cup Y_2 \cup Y_3) \mid (Y_4 \cup Y_6) = \mathbf{S}_{T',\{v_2w_4,v_2u_2\}} \in \mathcal{P}_{T',2}$. This implies $e_1, e_2 \in \mathrm{E}_{T'}$. Consequently, $\mathbf{S}_{T,\{e_1,e_2\}}|_{Y_5} = \{w\} \mid (Y_5 - w)$ is improper, which means $Y_5 = \{w\}$, as desired. □

Lemma 4.15. *Let T be a phylogenetic tree with at least seven leaves. Suppose $\max_{v\in\mathrm{V}_T} \deg_T(v) \leq 4$. For any $\{x, y, z\} \in \binom{\mathrm{L}_T}{3}$, $\{x, y, z\} \in \mathcal{Y}^{claw}(\mathcal{P}_{T,2})$ if and only if $\{x, y, z\}$ is a claw of T.*

Proof. We assume that $\{x, y, z\}$ is a claw of T. Let u_0 be the common neighbor of x, y, and z in T, and let v_0 be the only vertex in $\mathrm{N}_T(u_0) \setminus \{x, y, z\}$. Let

$$W = \begin{cases} \mathrm{L}_T \setminus (\mathrm{N}_T(v_0) \cup \{x, y, z\}), & \text{if } \deg_T(v_0) = 3, \\ \mathrm{L}_T \setminus \{x, y, z\}, & \text{otherwise.} \end{cases}$$

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

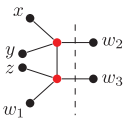


Figure 10. A possible tree appeared in the proof of Lemma 4.15.

It follows from $|\mathrm{L}_T| \geqslant 7$ that $|W| \geqslant 3$. Let $w \in W$. It is easy to verify that $\mathrm{Ext}_{T,\{x,y,z,w\}} = \{x, y, z, w, u_0\}$, and thus $(T, \{x, y, z, w\})$ is a 2-special configuration of type $\mathtt{E}$. Then, by Lemma 4.4, $\{x, y, z, w\}$ is a $(4, 2)$-set with respect to $\mathcal{P}_{T,2}$. Consequently, $\{x, y, z\} \in \mathcal{Y}^{\mathrm{claw}}(\mathcal{P}_{T,2})$.

Conversely, suppose that $\{x, y, z\} \in \mathcal{Y}^{\mathrm{claw}}(\mathcal{P}_{T,2})$. That is, there exists $\{w_1, w_2, w_3\} \in \binom{\mathrm{L}_T \setminus \{x,y,z\}}{3}$ such that $\{x, y, z, w_i\}$ is a $(4, 2)$-set with respect to $\mathcal{P}_{T,2}$ for all $i \in [3]$. This, together with Lemma 4.4 and the condition $\max_{v \in \mathrm{V}_T} \deg_T(v) \leq 4$, reveals that $(T, \{x, y, z, w_i\})$ is a 2-special configuration of type $\mathtt{B}$ or $\mathtt{E}$ for each $i \in [3]$. Since a leaf is contained in at most one claw of T, to prove that $\{x, y, z\}$ is a claw of T, it suffices to show that the type of $(T, \{x, y, z, w_i\})$ is $\mathtt{E}$ for each $i \in [3]$, which would mean that $\{x, y, z, w_i\}$ contains a claw of T. If $(T, \{x, y, z, w_i\})$ is of type $\mathtt{B}$, then applying Lemma 4.14 with $Y = \{x, y, z, w_i\}$ yields $\mathrm{L}_T \setminus Y = \{w_1, w_2, w_3\} - w_i$, which violates the assumption $|\mathrm{L}_T| \geqslant 7$; see Fig. 10. □

5. Tree reconstruction algorithm

Let us present Algorithm 2, our algorithm for recovering a phylogenetic tree from its score-2 split system. After demonstrating in Lemma 5.2 how Algorithm 2 accomplishes the three reduction steps mentioned in § 1, we will prove Theorem 1.3.

Lemma 5.1. *Let T be a phylogenetic X-tree. Then, $T \in \{\mathtt{A}_0, \widetilde{\mathtt{A}}_0\}$ if and only if $|X| = 6$ and $\mathcal{P}_{T,2} = \{Y \mid (X \setminus Y) : Y \in \binom{X}{2}\}$.*

Proof. Suppose $T \in \{\mathtt{A}_0, \widetilde{\mathtt{A}}_0\}$. It is straightforward to verify that $|X| = 6$ and $\mathcal{S}_{T,2} = \{Y \mid (X \setminus Y) : Y \in \binom{X}{2}\}$. Since T has no vertex of degree 3, it follows from Remark 2.4 and Corollary 2.6 that $\mathcal{P}_{T,2} = \mathcal{S}_{T,2}$.

Conversely, suppose $|X| = 6$ and $\mathcal{P}_{T,2} = \{Y \mid (X \setminus Y) : Y \subseteq X, |Y| = 2\}$. Let us pick $u, v \in \mathrm{V}_T \setminus X$ such that $\mathrm{dist}_T(u, v) = \max\{\mathrm{dist}_T(w, w') : w, w' \in \mathrm{V}_T \setminus X\}$. Then we must have $|\mathrm{N}_T(u) \setminus X| \leq 1$ and $|\mathrm{N}_T(v) \setminus X| \leq 1$. We know from Corollary 2.7 that T contains no cherry. Consequently, $|\mathrm{N}_T(u) \cap X| \geqslant 3$ and $|\mathrm{N}_T(v) \cap X| \geqslant 3$. Notice that $|X| = 6$. We thus conclude that $T \cong \mathtt{A}_0$ if $u = v$, and $T \cong \widetilde{\mathtt{A}}_0$ if $u \neq v$. □

Lemma 5.2. *Let X be a finite set with $|X| \geqslant 2$, and let $\mathcal{S}$ be a split system on X.*

(1) *If $\mathcal{S} \ncong \mathcal{P}_{\mathtt{A}_0,2}$, then Algorithm 2 outputs either a singleton set or the empty set.*
(2) *The output of Algorithm 2 is the set of all phylogenetic X-trees T with $\mathcal{P}_{T,2} = \mathcal{S}$ up to X-isomorphism.*

Proof. According to Lemma 5.1, $\mathcal{S}$ is isomorphic to $\mathcal{P}_{\mathtt{A}_0,2}$ if and only if $|X| = 6$ and $\mathcal{S} = \{Y \mid (X \setminus Y) : Y \subseteq X, |Y| = 2\}$. Except for its Line 7, Algorithm 2 always returns either a singleton set or the empty set. We thus verify (1).

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

Algorithm 2. Reconstruct a phylogenetic tree T from $\mathcal{P}_{T,2}$

Input: A split system $\mathcal{S}$ on a finite set X, where $|X| \geqslant 2$.
Output: A set of phylogenetic X-trees.
if $|X| = 2$ & $\mathcal{S} = \emptyset$ **then**
  return The tree with vertex set X and edge set $\{X\}$.
else if $|X| = 2$ **then**
  return $\emptyset$.
end if ▷ Now $|X| \geqslant 3$.
if $|X| = 6$ & $\mathcal{S} = \{Y \mid (X \setminus Y) :\ Y \subseteq X, |Y| = 2\}$ **then**
  return The set of all phylogenetic X-trees isomorphic to $\mathtt{A}_0$ or $\widetilde{\mathtt{A}}_0$.
end if
if $\mathcal{Y}^{\mathtt{A}}(\mathcal{S}) \neq \emptyset$ **then** ▷ (T, Y) is of type $\mathtt{A}$.
  $Y \leftarrow$ an arbitrary element of $\mathcal{Y}^{\mathtt{A}}(\mathcal{S})$.
  $\{Y_1, \ldots, Y_k\} \leftarrow \varrho_2(\Gamma_{\mathcal{S}}(Y))$.
  $\{Y_{k+1}, Y_{k+2}\} \leftarrow \varrho_2(\Gamma_{\mathcal{S}}(X \setminus (Y_1 \cup Y_2))) \setminus \{Y_1, \ldots, Y_k\}$.
  $T \leftarrow$ the output of Algorithm 1 using the split system $\mathcal{S}$ and the $(k+2)$-partition $Y_1 \mid \cdots \mid Y_{k+2}$ of X.
  if $\mathcal{P}_{T,2} = \mathcal{S}$ **then**
    return $\{T\}$.
  else
    return $\emptyset$.
  end if
end if
$n \leftarrow |X|$.
$\mathcal{S}' \leftarrow \mathcal{S}$. ▷ We will modify $\mathcal{S}'$ in the following steps.
$\mathcal{P}^{\mathrm{cherry}} \leftarrow \left\{\{x, y\} \in \binom{X}{2} : \{x, y\} \mid (X \setminus \{x, y\}) \notin \mathcal{S}\right\}$.
while $|X| \geqslant 4$ & $\mathcal{P}^{\mathrm{cherry}} \neq \emptyset$ & ($|X| \neq 7$ **or** $|\mathcal{P}^{\mathrm{cherry}}| \geqslant 2$) **do**
  $\{x, y\} \leftarrow$ an element of $\mathcal{P}^{\mathrm{cherry}}$.
  $X \leftarrow (X \setminus \{x, y\}) + v_{\{x,y\}}$, where $v_{\{x,y\}}$ is a new element outside of X.
  $\mathcal{S}' \leftarrow \{((X \setminus S) + v_{\{x,y\}}) \mid S :\ ((X \setminus S) \cup \{x, y\}) \mid S \in \mathcal{S}'\}$.
  $\mathcal{P}^{\mathrm{cherry}} \leftarrow \{\{x, y\} \in \binom{X}{2} : \{x, y\} \mid (X \setminus \{x, y\}) \notin \mathcal{S}'\}$.
end while
if $|X| \leq 6$ **then** ▷ Note that $|X| \geqslant 3$.
  $T \leftarrow$ an X-tree $K_{1,|X|}$.
else if $\mathcal{Y}^{\mathrm{claw}}(\mathcal{S}') = \emptyset$ **then**
  return $\emptyset$.

(*Continued*)

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

33: **else** ▷ Now $\max_{v\in V_T} \deg_T(v) \leq 4$, and T has at least one claw.
34: $\{x, y, z\} \leftarrow$ an arbitrary element of $\mathcal{Y}^{\text{claw}}(\mathcal{S}')$. ▷ $\{x, y, z\}$ is a claw of T.
35: $T \leftarrow$ the output of Algorithm 1 with the split system $\mathcal{S}'$ and the 4-partition $\{x\} \mid \{y\} \mid \{z\} \mid (X \setminus \{x, y, z\})$ of X.
36: **if** $\mathcal{P}_{T,2} \neq \mathcal{S}'$ **then**
37: **return** $\emptyset$.
38: **end if**
39: **end if**
40: **while** $|X| < n$ **do**
41: $\{x, y\} \leftarrow$ a set $\{x, y\}$ such that $v_{\{x,y\}} \in X$.
42: $X \leftarrow (X \cup \{x, y\}) \setminus \{v_{\{x,y\}}\}$.
43: $T \leftarrow$ the tree T with vertex set $V_T \cup \{x, y\}$ and edge set $E_T \cup \{v_{\{x,y\}}x, v_{\{x,y\}}y\}$.
44: **end while**
45: **if** $\mathcal{P}_{T,2} = \mathcal{S}$ **then**
46: **return** $\{T\}$.
47: **else**
48: **return** $\emptyset$.
49: **end if**

Suppose there is a phylogenetic X-tree T such that $\mathcal{S} = \mathcal{P}_{T,2}$. To prove (2), we need to show that, up to X-isomorphism, T is a member of the output of Algorithm 2.

If $|X| = 2$, the only phylogenetic X-tree is the tree with one edge, whose score-2 split system is $\emptyset$. Lines 1 to 5 of Algorithm 2 handle this case. If $\mathcal{S} \cong \mathcal{P}_{\mathtt{A}_0,2}$, then, by Lemma 5.1, the phylogenetic X-trees output in Line 7 are exactly all phylogenetic X-trees whose score-2 split systems are $\mathcal{S}$. In the remainder of this proof, assume $|X| \geqslant 3$ and $\mathcal{S} \ncong \mathcal{P}_{\mathtt{A}_0,2}$. Hence, T is not isomorphic to $\widetilde{\mathtt{A}}_0$.

CASE 1. $\mathcal{Y}^{\mathtt{A}}(\mathcal{S}) \neq \emptyset$. Lines 9 to 19 of Algorithm 2 address this case.

Select a Y from $\mathcal{Y}^{\mathtt{A}}(\mathcal{S})$ as described in Line 10. By Lemma 4.13, (T, Y) is a 2-special configuration of type $\mathtt{A}$. This implies the existence of a $v_1 \in \text{Ext}_{T,Y}$ that serves as an endpoint of both edges in $\partial_T(\text{Ext}_{T,Y})$. By Corollary 4.8 and Corollary 4.10 (2), the set $\{Y_1, \ldots, Y_{k+2}\}$ defined in Lines 11 and 12 corresponds precisely to $\{X \cap C_{T-v_1}[u] : u \in N_T(v_1)\}$. Subsequently, by Lemma 3.3, Algorithm 1 offers a method to reconstruct T from $\mathcal{P}_{T,2}$ and $Y_1 \mid \cdots \mid Y_{k+2}$. In this case, Algorithm 2 outputs $\{T\}$, as demonstrated in its Lines 13 and 15.

CASE 2. $\mathcal{Y}^{\mathtt{A}}(\mathcal{S}) = \emptyset$.

Execute the while loop in Lines 23 to 28. Suppose the while loop terminates after m iterations. For each $k \in [m]$, the values of variables $\{x, y\}$, X, $\mathcal{S}'$ and $\mathcal{P}^{\text{cherry}}$ in Lines 24 to 27 at the end of the k-th iteration are represented as $\{x_k, y_k\}$, X_k, $\mathcal{S}_k$ and $\mathcal{P}_k^{\text{cherry}}$, respectively. Let $T_0 = T$, $X_0 = X$, $\mathcal{S}_0 = \mathcal{S} = \mathcal{P}_{T,2}$ and $\mathcal{P}_0^{\text{cherry}}$ be the set defined in Line 22.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

We claim that, for each $k \in [m]$, $\mathcal{S}_k$ is the score-2 split system on a phylogenetic X_k-tree $T_k \doteq T_{k-1} - \{x_k, y_k\}$, where $\{x_k, y_k\}$ is a cherry of T_{k-1}. Since $\mathcal{S}_0 = \mathcal{P}_{T_0,2}$, we only need to prove the claim inductively for each $k \in [m]$ under the hypothesis that $\mathcal{S}_{k-1} = \mathcal{P}_{T_{k-1},2}$ for a phylogenetic X_{k-1}-tree T_{k-1}. By virtue of Corollary 2.7, $\{x_k, y_k\} \in \mathcal{P}_{k-1}^{\text{cherry}}$ is a cherry of T_{k-1}; and then by Lemma 3.1, the split system $\mathcal{S}_k$ is exactly the score-2 split system on the phylogenetic X_k-tree $T_k = T_{k-1} - \{x_k, y_k\}$. Our claim thus follows. We shall repeatedly use this claim implicitly in our remaining proof.

For each $k \in [m]$, T_{k-1} is obtained from T_k by adding two edges $x_k v_{\{x_k,y_k\}}$ and $y_k v_{\{x_k,y_k\}}$, where $v_{\{x_k,y_k\}}$ is defined in Line 25. So, we can construct T from T_m step by step, as demonstrated in Lines 40 to 44. Our task now reduces to finding T_m from the X_m-split system $\mathcal{S}_m$.

Observe that $\mathcal{S}_m \not\cong \mathcal{P}_{\mathtt{A}_0,2}$, which also means $T_m \not\cong \widetilde{\mathtt{A}}_0$. Otherwise, due to $\mathcal{S} \not\cong \mathcal{P}_{\mathtt{A}_0,2}$, the integer m must be positive and $T_{m-1} - \{x_m, y_m\} = T_m \cong \widetilde{\mathtt{A}}_0$. The latter means $|X_{m-1}| = 7$ and T_{m-1} contains exactly one cherry. From Corollary 2.7 and $\mathcal{S}_{m-1} = \mathcal{P}_{T_{m-1},2}$, we have $|\mathcal{P}_{m-1}^{\text{cherry}}| = 1$. Thus, the while loop terminates after $m-1$ iterations, contradicting our supposition.

It follows from Lemma 4.13 and $\mathcal{Y}^{\mathtt{A}}(\mathcal{S}) = \emptyset$ that $\max_{v \in \mathrm{V}_T} \deg_T(v) \leq 4$. As T_m is a subtree of T, it also holds that $\max_{v \in \mathrm{V}_{T_m}} \deg_{T_m}(v) \leq 4$. Because the while loop in Lines 23 to 28 terminates, there are three cases: $|X_m| = 3$; $\mathcal{P}_m^{\text{cherry}} = \emptyset$; $|X_m| = 7$ and $|\mathcal{P}_m^{\text{cherry}}| \leq 1$.

CASE 2.1. $3 \leq |X_m| \leq 6$. Lines 29 to 30 of Algorithm 2 deal with this case.

If $|X_m| = 3$, we must have $T_m \cong \mathrm{K}_{1,3}$.

When $4 \leq |X_m| \leq 6$, due to $T_m \not\cong \widetilde{\mathtt{A}}_0$, the tree T_m has at most one claw. Moreover, in this case, T_m has no cherry and $\max_{v \in \mathrm{V}_{T_m}} \deg_{T_m}(v) \leq 4$. We thus deduce that T_m has at most one inner vertex. This shows that $|X_m| = 4$ and $T_m \cong \mathrm{K}_{1,4}$.

CASE 2.2. $|X_m| \geqslant 7$. Lines 31 to 39 of Algorithm 2 deal with this case.

We first claim that T_m has a claw. Pick $w_1, w_2 \in \mathrm{V}_{T_m}$ such that $\mathrm{dist}_{T_m}(w_1, w_2) = \max\{\mathrm{dist}_{T_m}(u, v) : \ u, v \in \mathrm{V}_{T_m} \setminus X_m\}$. Then, for each $i \in [2]$, the number of leaves of T_m contained by $\mathrm{N}_{T_m}(w_i)$ equals $\deg_{T_m}(w_i) - 1$, which is at most 3. Furthermore, by Corollary 2.7 and $|\mathcal{P}_m^{\text{cherry}}| \leq 1$, T_m has at most one cherry. So, $\mathrm{N}_{T_m}(w_1) \cap X_m$ or $\mathrm{N}_{T_m}(w_2) \cap X_m$ is a claw of T_m, as desired. Thanks to Lemma 4.15, the set $\{x, y, z\}$ defined in Line 34 is a claw of T_m. Therefore, by Lemma 3.3, Line 35 of Algorithm 2 obtains the unique phylogenetic tree T_m up to X_m-isomorphism such that $\mathcal{P}_{T_m,2} = \mathcal{S}_m$.

In conclusion, Algorithm 2 outputs all phylogenetic X-trees T up to X-isomorphism such that $\mathcal{P}_{T,2} = \mathcal{S}$. □

Proof of Theorem 1.3 (1). This is Lemma 5.1.

(2). Clearly, the score-2 split systems of two X-isomorphic phylogenetic X-trees are equal. Conversely, since one of the phylogenetic trees T and T' is isomorphic to neither $\mathtt{A}_0$ nor $\widetilde{\mathtt{A}}_0$, by (1), the split system $\mathcal{P}_{T,2} = \mathcal{P}_{T',2}$ is not isomorphic to $\mathcal{P}_{\mathtt{A}_0,2}$. By Lemma 5.2, statement (2) now follows.

(3). Without loss of generality, assume that T' is isomorphic to neither $\mathtt{A}_0$ nor $\widetilde{\mathtt{A}}_0$. It suffices to show that $\mathcal{P}_{T,2} \cong \mathcal{P}_{T',2}$ implies $T \cong T'$.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

The condition $\mathcal{P}_{T,2} \cong \mathcal{P}_{T',2}$ means that there exists a bijection f on X such that

$$\mathcal{P}_{T',2} = \{f(A) \mid f(B) : \ A \mid B \in \mathcal{P}_{T,2}\}.$$

Let T'' be the graph with vertex set $\tilde{f}(\mathrm{V}_T)$ and edge set $\{\tilde{f}(u)\tilde{f}(w) : \ uw \in \mathrm{E}_T\}$, where

$$\tilde{f}(v) = \begin{cases} f(v), & \text{if } v \in X; \\ v, & \text{if } v \in \mathrm{V}_T \setminus X. \end{cases}$$

Clearly, T'' is a phylogenetic X-tree isomorphic to T and $\mathcal{P}_{T'',2} = \mathcal{P}_{T',2}$. Thus, by (2), $T' \cong_X T''$. Hence $T \cong T'' \cong T'$. □

6. Further research

As mentioned in the proof of Lemma 5.1, $\mathcal{S}_{\mathtt{A}_0,2} = \mathcal{P}_{\mathtt{A}_0,2} = \mathcal{P}_{\widetilde{\mathtt{A}}_0,2} = \mathcal{S}_{\widetilde{\mathtt{A}}_0,2}$. Let $\{\mathtt{A}_i, \widetilde{\mathtt{A}}_i\}$, $i \in \{2,3,4\}$, be three additional pairs of phylogenetic trees as shown in Fig. 11. We can verify that $\mathcal{S}_{\mathtt{A}_i,2} = \mathcal{S}_{\widetilde{\mathtt{A}}_i,2}$ for all $i \in \{2,3,4\}$. In Figs. 2 and 11 we highlight some edges in green to indicate the differences between $\mathtt{A}_i$ and $\widetilde{\mathtt{A}}_i$. Similar to the pair $\{\mathtt{A}_{-\infty}, \widetilde{\mathtt{A}}_{-\infty}\}$ displayed in Fig. 2, these four pairs of phylogenetic trees are distinguished in Conjecture 6.1 (1), which serves as a 2-split system counterpart to Theorem 1.3 (3). Let $\mathtt{B}_1, \mathtt{B}_2, \mathtt{B}_3$, be the phylogenetic trees in Fig. 12. Conjecture 6.1 (2) is a 2-split system counterpart of Corollary 1.4.

Conjecture 6.1. Let X be a finite set, and let T and T' be two phylogenetic X-trees.

(1) Assume that $\{T, T'\} \notin \left\{\{\mathtt{A}_0, \widetilde{\mathtt{A}}_0\}, \{\mathtt{A}_2, \widetilde{\mathtt{A}}_2\}, \{\mathtt{A}_3, \widetilde{\mathtt{A}}_3\}, \{\mathtt{A}_4, \widetilde{\mathtt{A}}_4\}\right\}$. Then, $\mathcal{S}_{T,2} \cong \mathcal{S}_{T',2}$ if and only if $T \cong T'$.
(2) Assume that $\mathcal{S}_{T,2} = \mathcal{S}_{T',2}$ and $T \ncong_X T'$. Then, T and T' are isomorphic if and only if they are both isomorphic to a common member of $\widetilde{\mathtt{A}}_0, \widetilde{\mathtt{A}}_3, \widetilde{\mathtt{A}}_4, \mathtt{B}_1, \mathtt{B}_2, \mathtt{B}_3$.

We will prove Conjecture 6.1 in [62]. This paper is concerned with score-2 split systems while [62] is devoted to 2-split systems. It is thus not surprising that in [62] we also need Algorithm 1 and we have a counterpart of Lemma 4.4 as a basic lemma there.

Let $\ell \geqslant 2$ be an integer. Wilde and Fischer [69, Corollary 3] presented two phylogenetic $[2\ell+2]$-trees T_1 and T_1' satisfying $\mathcal{P}_{T_1,\ell} = \mathcal{P}_{T_1',\ell}$ which are isomorphic but not $[2\ell+2]$-isomorphic. In Conjecture 6.2 (5), we further claim that there exist non-isomorphic phylogenetic $[2\ell+2]$-trees T_1 and T_1' such that $\mathcal{P}_{T_1,\ell} = \mathcal{P}_{T_1',\ell}$. Wilde and Fischer [70, Theorem 5] proved that, for every $|X| \geqslant 2\ell+3$ and every two binary phylogenetic X-trees T_1 and T_1', $\mathcal{P}_{T_1,\ell} = \mathcal{P}_{T_1',\ell}$ implies that T_1 and T_1' are X-isomorphic. Conjecture 6.2 (4) is a generalization of this result to non-binary case. We report analogous results in Conjecture 6.2 (1), (2) and (3).

Conjecture 6.2. Let ℓ be a positive integer, and let X be a finite set. When $\ell \geqslant 2$, let $X_\ell \doteq \{x_1, \ldots, x_{\ell+2}, y_1, \ldots, y_{\ell+2}\}$ be a set of size $2\ell+4$, and, for every $j \in [\ell]$, let

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

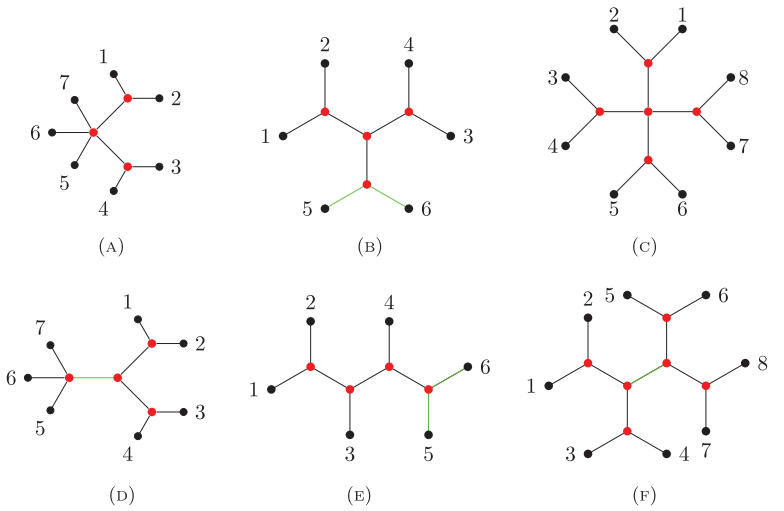


Figure 11. Three pairs of trees with isomorphic 2-split systems. (A) $\mathtt{A}_2$, (B) $\mathtt{A}_3$, (C) $\mathtt{A}_4$, (D) $\widetilde{\mathtt{A}}_2$, (E) $\widetilde{\mathtt{A}}_3$, (F) $\widetilde{\mathtt{A}}_4$

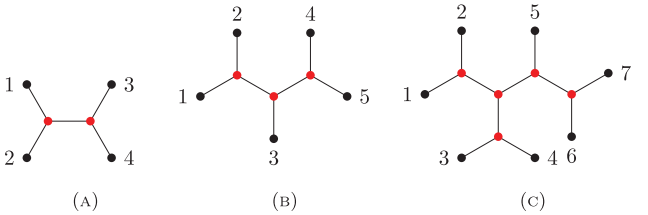


Figure 12. Three binary trees. (A) $\mathtt{B}_1$, (B) $\mathtt{B}_2$, (C) $\mathtt{B}_3$

$T_{\ell,j}$ be the X_ℓ-tree shown in Fig. 13. Note that $T_{\ell,j}$ and $T_{\ell,j'}$ are non-isomorphic for all $\{j, j'\} \in \binom{X_\ell}{2}$.

(1) If $|X| \geqslant 2\ell + 5$, then two phylogenetic X-trees T and T' are X-isomorphic if and only if $\mathcal{S}_{T,\ell} = \mathcal{S}_{T',\ell}$.
(2) If $|X| \geqslant 2\ell + 5$, then two phylogenetic X-trees T and T' are X-isomorphic if and only if $\bigcup_{0\leq i\leq \ell} \mathcal{P}_{T,i} = \bigcup_{0\leq i\leq \ell} \mathcal{P}_{T',i}$.
(3) If $\ell \geqslant 2$, then $|X_\ell| = 2\ell + 4$, the phylogenetic X_ℓ-trees $T_{\ell,j}$, $j \in [\ell]$, are pairwise non-isomorphic, but $\bigcup_{0\leq i\leq \ell} \mathcal{P}_{T_{\ell,j},i}$ is a split system independent of $j \in [\ell]$.
(4) If $|X| \geqslant 2\ell + 3$, then two phylogenetic X-trees T and T' are X-isomorphic if and only if $\mathcal{P}_{T,\ell} = \mathcal{P}_{T',\ell}$.
(5) If $\ell \geqslant 3$, then $|X_{\ell-1}| = 2\ell + 2$ and the phylogenetic $X_{\ell-1}$-trees $T_{\ell-1,j}$, $j \in [\ell - 1]$, are pairwise non-isomorphic but share a common score-ℓ split system.

We have verified Conjecture 6.2 (1) and we plan to further study if the bound of $|X| \geqslant 2\ell + 5$ there is sharp before completing the corresponding paper. We will prove Conjecture 6.2 (2) and (3) in [63], and Conjecture 6.2 (4) and (5) in [64].

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

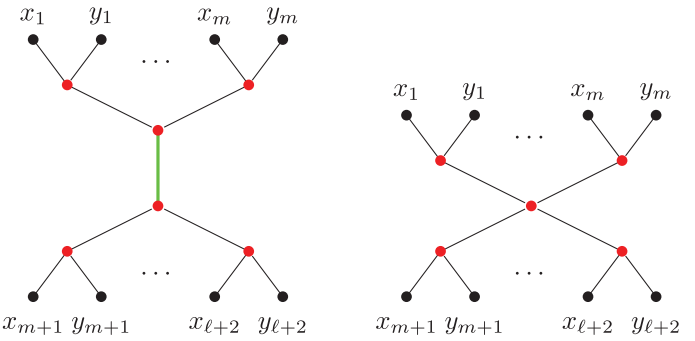


Figure 13. ℓ trees T sharing the same $\mathcal{P}_{T,\leq \ell}$ and the same $\mathcal{P}_{T,\ell+1}$, where $\ell \geqslant 2$. (A) Trees $T_{\ell,m}$, where $m-1 \in [\ell-1]$. (B) The tree $T_{\ell,1}$.

In order to tell a tree from its score-2 split system, one important step in our work is to tell a vertex of large degree; see Case 1 of the proof of Lemma 5.2 (2). To extend this step for the recognition of large degree vertex from general score-ℓ split systems, we propose Conjectures 6.3 and 6.4. Note that both conjectures hold true when $\ell = 2$, as guaranteed by Corollary 4.8, Corollary 4.10 (2) and Lemma 4.13.

Conjecture 6.3. Let k and ℓ be two integers such that $k \geqslant \ell + 3 \geqslant 5$. Let T be a phylogenetic tree, and take $Y \mid (\mathrm{L}_T \setminus Y) \in \mathcal{P}_{T,\ell}$. Assume that $\pi = Y_1 \mid \cdots \mid Y_k$ is a k-partition of Y such that $\Gamma_{\mathcal{P}_{T,\ell}}(Y) = \mathcal{H}_{\pi,\ell}$, namely, Y is a (k,ℓ)-set with respect to $\mathcal{P}_{T,\ell}$. Then, there exists a unique $v \in \mathrm{V}_T$ such that the connected components of $T - v$ can be enumerated as $V_1, \ldots, V_{k+\ell}$ with $V_i \cap \mathrm{L}_T = Y_i$ for all $i \in [k]$.

Conjecture 6.4. Let k and ℓ be two integers such that $k \geqslant \ell + 1 \geqslant 3$, and let T be a phylogenetic tree. Assume that $T \ncong \widetilde{\mathtt{A}}_0$. Let $\pi = Y_1 \mid \cdots \mid Y_{k+\ell}$ be a $(k+\ell)$-partition of L_T such that $(\bigcup_{i\in I} Y_i) \mid (\bigcup_{i\in[k+\ell]\setminus I} Y_i) \in \mathcal{P}_{T,\ell}$ and $\Gamma_{\mathcal{P}_{T,\ell}}(\bigcup_{i\in[k+\ell]\setminus I} Y_i) = \{\bigcup_{i\in I'} Y_i : I' \in \binom{[k+\ell]\setminus I}{\ell}\}$ for every $I \in \binom{[k+\ell]}{\ell}$. Then, there exists a unique $v \in \mathrm{V}_T$ that allows us to enumerate the connected components of $T - v$ as $V_1, \ldots, V_{k+\ell}$ such that $V_i \cap \mathrm{L}_T = Y_i$ for all $i \in [k+\ell]$.

Buneman graphs and Buneman complexes have been useful tools for visualizing split systems [8, 12, 18, 25, 27–29, 57]. The Buneman Theorem concerning 1-split systems, namely Theorem 1.1, can be proved using the Buneman graph. It is reasonable to expect that a similar approach could be applied to directly reconstruct a tree from a suitable k-split system or score-k split system. Below, we outline one such possible procedure. Let X be a finite set, let $\mathcal{S}$ be a split system over X, and let ℓ be a positive integer. Let $\mathrm{V}_\ell(\mathcal{S})$ be the set of all functions $f : \mathcal{S} \to 2^X$ satisfying the following two conditions:

(1) The condition $f(\mathbf{S}) \in \mathbf{S}$ holds for all $\mathbf{S} \in \mathcal{S}$.
(2) The intersection $\bigcap_{\mathbf{S}\in\mathcal{S}'} f(\mathbf{S}) \neq \emptyset$ holds for every $\mathcal{S}' \in \binom{\mathcal{S}}{\ell}$.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

We define the *ℓth generalized Buneman graph*, denoted by $\mathrm{B}_\ell(\mathcal{S})$, as the graph with vertex set $\mathrm{V}_\ell(\mathcal{S})$ and with $uv \in \binom{\mathrm{V}_\ell(\mathcal{S})}{2}$ being an edge of $\mathrm{B}_\ell(\mathcal{S})$ exactly when the Hamming distance between u and v is the smallest among all pairs of distinct elements from $\mathrm{V}_\ell(\mathcal{S})$. Note that what is known as the Buneman graph [12, p. 388] [28, § 4.1] in the literature corresponds to our $\mathrm{B}_2(\mathcal{S})$. For every positive integer ℓ, let δ_ℓ denote the smallest positive integer d such that $\mathrm{B}_{\ell+1}(\mathcal{S}_{T,\ell}) \cong T$ for every tree T with at least three vertices whose inner vertices have degrees no smaller than d; let δ'_ℓ denote the smallest positive integer d such that $\mathrm{B}_{\ell+1}(\mathcal{P}_{T,\ell}) \cong T$ for every tree T with at least three vertices whose inner vertices have degrees no smaller than d. When Buneman [12, p. 389] used the Buneman graph to establish Theorem 1.1, what he actually proved is just $\delta_1 = \delta'_1 = 3$.

Conjecture 6.5. (1) For every positive integer ℓ, we have $\infty > \delta_\ell = \delta'_\ell > 2\ell$. (2) $\delta_2 = \delta'_2 = 7$.

Question 6.6. *Let* T *be a phylogenetic tree, let ℓ be a positive integer, let $\mathbf{S} \in \mathcal{S}_{T,\ell}$ and let $\mathcal{S}$ be the full ℓ-split system of* T *induced by $\boldsymbol{S}$. When is $\mathcal{S}$ compatible? When $\ell = 1$, $\mathcal{S}$ is a singleton set and hence compatible. When $\ell = 2$, the compatibility of $\mathcal{S}$ is ensured by Lemma 2.10. Note that compatible split systems arise in Buneman theorem (Theorem 1.1 (1)) and Apresjan's Lemma [28, Lemma 7.5].*

Though a split can be considered an unoriented/affine object, it has a close relationship with its oriented/projective counterparts [20, 30, 51]. Split systems, as well as the similar concept of separation systems – particularly those with tree-like behaviours – have been an important research subject in geometric group theory [1, 33, 68], cluster algebra [19, 41], electrical networks [39], random processes [35, 67], structural graph theory and matroid theory [11, 22, 34, 42, 43, 52, 59, 65], extremal set theory [32, 53], and cluster analysis and phylogenetic combinatorics [12, 26, 28, 45–47, 49]. It seems promising to further expand our current work to the setting of other mathematical contexts, such as any of the above-mentioned ones.

Splits are very special partitions. What are the partition counterparts for Theorem 1.3, conjectures 6.1 and 6.2? Note that, in biology, a character is a partition of a subset of species. We refer to [3, 9, 10, 50, 58, 67] for some interesting study of partitions from tree structures. For a tree T, the splits of L_T discussed in this paper are related to unions of one or two convex sets in T. It would be intriguing to explore how to advance this research in an extended study of abstract convexity [9, 7, 14, 16, 21, 60].

Acknowledgements. The thorough review from our referees had a profound impact on improving the quality of the presentation of this paper. This work was supported by the Fundamental Research Funds for the Central Universities. We dedicate this paper to the memory of Andreas Dress (26 August, 1938–23 February, 2024), who worked from 2005 to 2010 in the Life Science Building at Yueyang Road 320, Shanghai. From 1946 to 1948, Shiing-Shen Chern (26 October, 1911–3 December, 2004) also worked in the same building. Wenjun Wu, a distinguished alumnus of Jiao Tong University, followed Chern to study algebraic topology at Yueyang Road 320 during that time; according to his recollections, Chern's office was located on the second floor. We have vivid memories of the third floor, particularly room 305 (Dress's office), the seminar room, and the common room. It was through interactions with individuals who frequented the third

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

floor, including Dress and his colleagues, that we gained a foundational understanding of the intrinsic beauty of trees.

References

(1) S. Adam-Day, On the continuous gradability of the cut-point orders of $\mathbb{R}$-trees, *Topology Appl.* **306** (2022), 107937, pp. 21. doi:10.1016/j.topol.2021.107937.

(2) M. Aivaliotis, G. Gordon and W. Graveman, When bad things happen to good trees, *J. Graph Theory* **37**(2): (2001), 79–99. doi:10.1002/jgt.1004.

(3) J. Aliste-Prieto, J. L. Martin, J. D. Wagner and J. Zamora, Chromatic symmetric functions and polynomial invariants of trees, *Bull. Lond. Math. Soc.* **56**(11): (2024), 3452–3476. doi:10.1112/blms.13144.

(4) J. Backelin and S. Linusson, Parity splits by triple point distances in X-trees, *Ann. Comb.* **10**(1): (2006), 1–18. doi:10.1007/s00026-006-0270-1.

(5) J. C. Baez and N. Otter, Operads and phylogenetic trees, *Theory Appl. Categ.* **32** (2017), 1397–1453, www.tac.mta.ca/tac/volumes/32/40/32-40abs.html.

(6) H.-J. Bandelt and A. Röhl, Quasi-median hulls in Hamming space are Steiner hulls, *Discrete Appl. Math.* **157**(2): (2009), 227–233. doi:10.1016/j.dam.2006.09.015.

(7) P. Bertrand and J. Diatta, An interval convexity-based framework for multilevel clustering with applications to single-linkage clustering, *Discrete Appl. Math.* **342** (2024), 38–63. doi:10.1016/j.dam.2023.08.003.

(8) S. Besenbacher, T. Mailund, L. Westh-Nielsen and C. N. S. Pedersen, RBT – A tool for building refined Buneman trees, *Bioinformatics* **21**(8): (2004), 1711–1712. doi:10.1093/bioinformatics/bti195.

(9) M. Bordewich, I. M. Deutschmann, M. Fischer, E. Kasbohm, C. Semple and M. Steel, On the information content of discrete phylogenetic characters, *J. Math. Biol.* **77**(3): (2018), 527–544. doi:10.1007/s00285-017-1198-2.

(10) M. Bordewich, C. Semple and M. Steel, Identifying X-trees with few characters, *Electron. J. Combin.* **13**(1): (2006), R83, pp. 11. doi:10.37236/1109.

(11) N. Bowler and J. Kneip, Separations of sets, *Order* **37**(2): (2020), 411–425. doi:10.1007/s11083-019-09512-4.

(12) P. Buneman, The recovery of trees from measures of dissimilarity. *Mathematics in the Archeological and Historical Sciences*, In F. Hodson, D. Kendall P. Tautu, (Eds.) pp. 387–395 (Edinburgh University Press, United Kingdom, 1971). https://homepages.inf.ed.ac.uk/opb/homepagefiles/phylogeny-scans/manuscripts.pdf.

(13) F. Burghart, Polynomial invariants for rooted trees related to their random destruction, *Electron. J. Combin.* **31**(4): (2024), P4.37, pp. 19. doi:10.37236/11975.

(14) D. Cantone, J. -P. Doignon, A. Giarlotta and S. Watson, Resolutions of convex geometries, *Electron. J. Combin.* **28**(4): (2021), P4.26, pp. 39. doi:10.37236/10328.

(15) M. Casanellas and J. Fernández-Sánchez, Relevant phylogenetic invariants of evolutionary models, *J. Math. Pures Appl. (9)* **96**(3): (2011), 207–229. doi:10.1016/j.matpur.2010.11.002.

(16) J. Chalopin, V. Chepoi and U. Giocanti, Graphs with convex balls, *Geom. Dedicata* **217**(4): (2023), 67, pp. 52. doi:10.1007/s10711-023-00803-0.

(17) S. Chaudhary and G. Gordon, Tutte polynomials for trees, *J. Graph Theory* **15**(3): (1991), 317–331. doi:10.1002/jgt.3190150308.

(18) Y.-B. Choe, K. T. Huber, J. H. Koolen, Y. S. Kwon and V. Moulton, Counting vertices and cubes in median graphs of circular split systems, *European J. Combin.* **29**(2): (2008), 443–456. doi:10.1016/j.ejc.2007.02.003.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

(19) V. I. Danilov, A. V. Karzanov and G. A. Koshevoy, On interrelations between strongly, weakly and chord separated set-systems (a geometric approach), *J. Algebraic Combin.* **54**(4): (2021), 1299–1327. doi:10.1007/s10801-021-01047-5.

(20) G. Dasarathy, E. Mossel, R. Nowak and S. Roch, A stochastic Farris transform for genetic data under the multispecies coalescent with applications to data requirements, *J. Math. Biol.* **84**(5): (2022), 36, pp. 37. doi:10.1007/s00285-022-01731-5.

(21) M. Develin and B. Sturmfels, Tropical convexity, *Doc. Math.* **9** (2004), 1–27. doi:10.4171/DM/154.

(22) R. Diestel, F. Hundertmark and S. Lemanczyk, Profiles of separations: in graphs, matroids, and beyond, *Combinatorica* **39**(1): (2019), 37–75. doi:10.1007/s00493-017-3595-y.

(23) O. Dovgoshey and A. Kostikov, Locally finite ultrametric spaces and labeled trees, *J. Math. Sci. New York* **276**(5): (2023), 614–637. doi:10.1007/s10958-023-06786-3.

(24) A. W. M. Dress, K. T. Huber and V. Moulton. Some variations on a theme by Buneman, *Ann. Comb.* **1**(4): (1997), 339–352. doi:10.1007/BF02558485.

(25) A. W. M. Dress, K. T. Huber and V. Moulton. A comparison between two distinct continuous models in projective cluster theory: The median and the tight-span construction, *Ann. Comb.* **2**(4): (1998), 299–311. doi:10.1007/BF01608527.

(26) A. W. M. Dress, M. D. Hendy, K. T. Huber and V. Moulton, On the number of vertices and edges of the Buneman graph, *Ann. Comb.* **1**(4): (1997), 329–337. doi:10.1007/BF02558484.

(27) A. W. M. Dress, K. T. Huber, J. Koolen, V. Moulton and A. Spillner, *Basic Phylogenetic Combinatorics.*, (Cambridge University Press, Cambridge, 2012) doi:10.1017/CBO9781139019767.

(28) A. W. M. Dress, K. T. Huber and V. Moulton, An explicit computation of the injective hull of certain finite metric spaces in terms of their associated Buneman complex, *Adv. Math.* **168**(1): (2002), 1–28. doi:10.1006/aima.2001.2039.

(29) A. W. M. Dress, K. T. Huber and V. L. Moulton, Some uses of the Farris transform in mathematics and phylogenetics – a review, *Ann. Comb.* **11**(1): (2007), 1–37. doi:10.1007/s00026-007-0302-5.

(30) A. W. M. Dress, M. Klucznik, J. H. Koolen and V. L. Moulton, $2kn - \binom{2k+1}{2}$: A note on extremal combinatorics of cyclic split systems, *Sém. Lothar. Combin.* **47** (2001), B47, pp. 17. https://eudml.org/doc/121892.

(31) A. W. M. Dress and Y. Wu, Phylogenetic analysis and mod 2 homology, *Forum Math.* **24**(3): (2012), 495–505. doi:10.1515/form.2011.070.

(32) M. J. Dunwoody and B. Krön, Vertex cuts, *J. Graph Theory* **80**(2): (2015), 136–171. doi:10.1002/jgt.21844.

(33) C. Elbracht, J. Kneip and M. Teegen, Trees of tangles in infinite separation systems, *Math. Proc. Cambridge Philos. Soc.* **173**(2): (2022), 297–327. doi:10.1017/S0305004121000530.

(34) S. N. Evans, *Probability and Real Trees, volume 1920 of Lecture Notes in Mathematics*. (Springer, Berlin, 2005) doi:10.1007/978-3-540-74798-7.

(35) M. Fischer, On the uniqueness of the maximum parsimony tree for data with up to two substitutions: An extension of the classic Buneman theorem in phylogenetics, *Mol. Phylogenet. Evol* **137** (2019), 127–137. doi:10.1016/j.ympev.2019.03.013.

(36) M. Fischer, Defining binary phylogenetic trees using parsimony, *Ann. Comb.* **27**(3): (2023), 457–467. doi:10.1007/s00026-022-00627-x.

(37) M. Fischer, L. van Iersel, S. Kelk and C. Scornavacca, On computing the maximum parsimony score of a phylogenetic network, *SIAM J. Discrete Math.* **29**(1): (2015), 559–585. doi:10.1137/140959948.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

(38) S. Forcey, Circular planar electrical networks, split systems, and phylogenetic networks, *SIAM J. Appl. Algebra Geom.* **7**(1): (2023), 49–76. doi:10.1137/22M1473844.

(39) A. Francis, K. T. Huber, V. Moulton and T. Wu, Encoding and ordering $*X$*-cactuses, *Adv. in Appl. Math.* **142** (2023), 102414, pp. 31. doi:10.1016/j.aam.2022.102414. 31 pp.

(40) P. Galashin and A. Postnikov, Purity and separation for oriented matroids, *Mem. Amer. Math. Soc.* **289**(1439): (2023) pp. v+79. doi:10.1090/memo/1439.

(41) P. Gambette, K. T. Huber and G. E. Scholz, Uprooted phylogenetic networks, *Bull. Math. Biol.* **79**(9): (2017), 2022–2048. doi:10.1007/s11538-017-0318-x.

(42) E. Gioan and C. Paul, Split decomposition and graph-labelled trees: characterizations and fully dynamic algorithms for totally decomposable graphs, *Discrete Appl. Math.* **160**(6): (2012), 708–733. doi:10.1016/j.dam.2011.05.007.

(43) P. A. Goloboff and M. Wilkinson, On defining a unique phylogenetic tree with homoplastic characters, *Mol. Phylogenet. Evol* **122** (2018), 95–101. doi:10.1016/j.ympev.2018.01.020.

(44) M. Hellmuth, K. T. Huber, V. Moulton, G. E. Scholz and P. F. Stadler, Injective split systems, *Graphs Combin.* **39**(4): (2023), 65, pp. 22. doi:10.1007/s00373-023-02660-w.

(45) M. Hellmuth, D. Schaller and P. F. Stadler, Compatibility of partitions with trees, hierarchies, and split systems, *Discrete Appl. Math.* **314** (2022), 265–283. doi:10.1016/j.dam.2022.03.014.

(46) M. Hellmuth, D. Schaller and P. F. Stadler, Clustering systems of phylogenetic networks, *Theory Biosci.* **142**(4): (2023), 301–358. doi:10.1007/s12064-023-00398-w.

(47) P. A. Helminck, Invariants for trees of non-archimedean polynomials and skeleta of superelliptic curves, *Math. Z.* **301**(2): (2022), 1259–1297. doi:10.1007/s00209-021-02959-5.

(48) M. Hendriksen and N. Kapust, On the comparison of incompatibility of split systems across different numbers of taxa, *Bull. Math. Biol.* **83**(7): (2021), 78, pp. 15. doi:10.1007/s11538-021-00911-6.

(49) K. T. Huber, V. Moulton, C. Semple and T. Wu, Representing partitions on trees, *SIAM J. Discrete Math.* **28**(3): (2014), 1152–1172. doi:10.1137/130906192.

(50) A. Kleinman, M. Harel and L. Pachter, Affine and projective tree metric theorems, *Ann. Comb.* **17**(1): (2013), 205–228. doi:10.1007/s00026-012-0173-2.

(51) M. Koloschin, T. Krill and M. Pitz, End spaces and tree-decompositions, *J. Combin. Theory Ser. B* **161** (2023), 147–179. doi:10.1016/j.jctb.2023.02.007.

(52) A. Kupavskii, J. Pach and I. Tomon, On the size of K-cross-free families, *Combinatorica* **39**(1): (2019), 153–164. doi:10.1007/s00493-017-3792-8.

(53) S-D. Li, Y-Z. Fan, T. Ito, M. Karimi and J. Xu, The isomorphism problem of trees from the viewpoint of Terwilliger algebras, *Id/No* **177**: (2021), 105328, pp. 11. doi:10.1016/j.jcta.2020.105328.

(54) P. Liu, A tree distinguishing polynomial, *Discrete Appl. Math.* **288** (2021), 1–8. doi:10.1016/j.dam.2020.08.019.

(55) P. Liu, P. Biller, M. Gould and C. Colijn, Analyzing phylogenetic trees with a tree lattice coordinate system and a graph polynomial, *Systematic Biology* **71**(6): (2022), 1378–1390. doi:10.1093/sysbio/syac008.

(56) N. A. Loehr and G. S. Warrington, A rooted variant of Stanley's chromatic symmetric function, *Discrete Math.* **347**(3): (2024), 113805, pp. 18. doi:10.1016/j.disc.2023.113805.

(57) N. Misra, G. Blelloch, R. Ravi and R. Schwartz, Generalized Buneman pruning for inferring the most parsimonious multi-state phylogeny, *J. Comput. Biol.* **18**(3): (2011), 445–457. doi:10.1089/cmb.2010.0254.

(58) R. Orellana and G. Scott, Graphs with equal chromatic symmetric functions, *Discrete Math.* **320** (2014), 1–14. doi:10.1016/j.disc.2013.12.006.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press

(59) S. Oum, Rank-width and vertex-minors, *J. Combin. Theory Ser. B* **95**(1): (2005), 79–100. doi:10.1016/j.jctb.2005.03.003.

(60) E. Petrov, Weak similarities of finite ultrametric and semimetric spaces, *p-Adic Numbers Ultrametric Anal. Appl.* **10**(2): (2018), 108–117. doi:10.1134/S2070046618020048.

(61) J. C. Pons, T. M. Coronado, M. Hendriksen and A. Francis, A polynomial invariant for a new class of phylogenetic networks, *PLoS One* **17**(5): (2022), e0268181, pp. 22. doi:10.1371/journal.pone.0268181.

(62) C. Qian, Y. Wu and M. Xie. How to tell a phylogenetic tree from its 2-split system?, 2025.

(63) C. Qian, Y. Wu and M. Xie. A phylogenetic tree with at least $2\ell+5$ leaves is determined by its score-$\leq \ell$ split system, 2025. URL: https://math.sjtu.edu.cn/faculty/ykwu/data/Paper/split202507.pdf.

(64) C. Qian, Y. Wu and M. Xie. Reconstruct a phylogenetic tree with at least $2\ell+3$ leaves from its score-ℓ split system, 2025.

(65) B. Schröter, Multi-splits and tropical linear spaces from nested matroids, *Discrete Comput. Geom.* **61**(3): (2019), 661–685. doi:10.1007/s00454-018-0021-1.

(66) C. Semple and M. Steel, *Phylogenetics. Oxford Lecture Series in Mathematics and its Applications.*, Volume 24, (Oxford University Press, Oxford, 2003).

(67) M. Steel. *Phylogeny – Discrete and Random Processes in Evolution*, *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), 2016, Volume 89 Philadelphia, PA, doi:10.1137/1.9781611974485.ch1.

(68) A. Tserunyan, A descriptive construction of trees and Stallings' theorem. In *Trends in Set Theory*, volume of *Contemp. Math.* Volume 752, pp. 191–207 (Amer. Math. Soc. Providence, RI, 2020) doi:10.1090/conm/752/15137.

(69) M. Wilde and M. Fischer, Defining binary phylogenetic trees using parsimony: new bounds, *Discrete Appl. Math.* **341** (2023), 257–269. doi:10.1016/j.dam.2023.08.005.

(70) M. Wilde and M. Fischer, A complete characterization of pairs of binary phylogenetic trees with identical A_k-alignments. 2024, arXiv:2408.07011.

(71) G. R. Wilkes. On the structure of vertex cuts separating the ends of a graph, *Pacific J. Math.* **278**(2): (2015), 479–510. doi:10.2140/pjm.2015.278.479.

(72) Y. Wu, Lit-only sigma game on a line graph, *European J. Combin.* **30**(1): (2009), 84–95. doi:10.1016/j.ejc.2008.02.003.

https://doi.org/10.1017/S0013091525100874 Published online by Cambridge University Press