

# Combinatorial properties of triplet covers for binary trees

Stefan Grünewald<sup>a</sup>, Katharina T. Huber<sup>\*b</sup>, Vincent Moulton<sup>b</sup>, Mike Steel<sup>c</sup>

<sup>a</sup>*CAS-MPG Partner Institute for Computational Biology, Chinese Academy of Sciences  
Key Laboratories of Computational Biology, Shanghai, China*

<sup>b</sup>*School of Computing Sciences, University of East Anglia, Norwich, UK*

<sup>c</sup>*Biomathematics Research Centre, University of Canterbury, Christchurch, NZ*

---

## Abstract

It is a classical result that an unrooted tree  $T$  having positive real-valued edge lengths and no vertices of degree two can be reconstructed from the induced distance between each pair of leaves. Moreover, if each non-leaf vertex of  $T$  has degree 3 then the number of distance values required is linear in the number of leaves. A canonical candidate for such a set of pairs of leaves in  $T$  is the following: for each non-leaf vertex  $v$ , choose a leaf in each of the three components of  $T - v$ , group these three leaves into three pairs, and take the union of this set over all choices of  $v$ . This forms a so-called ‘triplet cover’ for  $T$ . In the first part of this paper we answer an open question (from 2012) by showing that the induced leaf-to-leaf distances for any triplet cover for  $T$  uniquely determine  $T$  and its edge lengths. We then investigate the finer combinatorial properties of triplet covers. In particular, we describe the structure of triplet covers that satisfy one or more of the following properties of being minimal, ‘sparse’, and ‘shellable’.

*Keywords:* Phylogenetic tree, triplet cover, tree-distances, Hall’s theorem, ample patchwork, shellability

---

\*Corresponding author

---

*Email addresses:* stefan@picb.ac.cn (Stefan Grünewald), K.Huber@uea.ac.uk (Katharina T. Huber\*), V.Moulton@uea.ac.uk (Vincent Moulton), Mike.Steel@canterbury.ac.nz (Mike Steel)

## 1. Introduction

Trees with a label set  $X$  of leaves play a central role in many areas of classification, such as systematic biology and linguistics. In these settings, it is usually assumed that the non-leaf vertices of the tree have degree at least three, and that there is an assignment of a positive real-valued length to each edge of  $T$ . A classical and important result from the 1960s and 1970s asserts that any such (unrooted) tree  $T$  with edge lengths is uniquely determined from the induced leaf-to-leaf distances between each pair of elements of  $X$ . This result is the basis of widely-used methods for inferring trees from distance data, such as the popular ‘Neighbor-Joining’ algorithm [1].

When the unrooted tree  $T$  is binary (each non-leaf vertex has degree 3) then we do not require distance values for all of the  $\binom{n}{2}$  pairs from leaf set  $X$  (where  $n = |X| \geq 3$ ), since just  $2n - 3$  carefully selected pairs of leaves suffice to determine  $T$  and its edge lengths (see [2]; more recent results appear in [3], motivated by the irregular distribution of genes across species in biological data). This value of  $2n - 3$  cannot be made any smaller, since a binary unrooted tree with  $n$  leaves has  $2n - 3$  edges, and the inter-leaf distances are linear combinations of the corresponding  $2n - 3$  edge lengths (so, by linear algebra, these values cannot be uniquely determined by fewer than  $2n - 3$  equations).

There is a particularly natural way to select a subset of  $\binom{X}{2}$  for  $T$  when  $T$  is binary. Since each non-leaf vertex is incident with three subtrees of  $T$ , let us (i) select a leaf from each subtree, (ii) consider the three pairs of leaves we can form from this triple, and then (iii) take the union of these sets of pairs over all non-leaf vertices of  $T$ . This process produces a ‘triplet cover’ of  $T$  (defined more precisely below), as illustrated in Fig. 1.

A triplet cover need not be of this minimum size (i.e. of size  $2n - 3$ ) and in an earlier paper we characterized when it is [4]. That paper also established that in this case the resulting triplet cover is ‘shellable’, complementing other recent work into phylogenetic ‘lasso’ sets [3, 5], as well as a Hall-type characterization of the median function on trees in [6].

In this paper, we present three main new results. Our first result (Theorem 2.1, which is a special case of Theorem 2.3) answers in the affirmative a question that has been open since 2012, namely do the distances between leaves induced by a triplet cover on a binary tree with positive edge lengths determine the tree? Our second main result (Theorem 4.1) describes the structure of ‘sparse’ triplet covers in terms of a 2-tree decomposition of a certain graph. Our third main result (Theorem 5.4) provides a sufficient condition for a triplet cover to be ‘shellable’. Along the way, a number of other properties of triplet covers are derived. We begin with some definitions.

### 1.1. Definitions

Let  $X$  be a finite set  $|X| \geq 3$ . Given a set  $\mathcal{C}$  of subsets of a set  $Y$ , we let  $\bigcup \mathcal{C} = \bigcup_{t \in \mathcal{C}} t$ , and we denote elements in  $\binom{X}{2}$  and  $\binom{X}{3}$  also by  $xy$  and  $xyz$ , respectively, where  $x, y, z \in X$ .

Given a graph  $G = (V, E)$ , we let  $V = V(G)$  denote its vertex set and  $E = E(G)$  its edge-set.

- A *phylogenetic  $X$ -tree* is a tree  $T = (V, E)$  that has leaf-set  $X$  and for which the non-leaf vertices have degree at least 3. If all non-leaf vertices of  $T$  have degree exactly 3 then we say that  $T$  is a *binary phylogenetic  $X$ -tree* (or simply a binary phylogenetic tree when the leaf set is clear or not important).
- We let  $\mathring{V} = \mathring{V}(T) \subseteq V(T)$  denote the set of interior vertices of  $T$ . If  $T$  is binary then  $|\mathring{V}| = |X| - 2$ .
- A *cherry* of a binary phylogenetic  $X$ -tree  $T$  is a pair of leaves of  $T$  that are adjacent to a common vertex.
- Two phylogenetic  $X$ -trees  $T = (V, E)$  and  $T' = (V', E')$  are *isomorphic*, denoted  $T \cong T'$ , precisely if there is a graph isomorphism  $\varphi$  from  $T$  to  $T'$  that sends leaf  $x$  of  $T$  to leaf  $x$  of  $T'$  for each  $x \in X$ .

Suppose that  $T$  is a phylogenetic  $X$ -tree as above.

- Given an element  $x \in X$ , where  $|X| \geq 4$ , we let  $T - x$  denote the phylogenetic  $X$ -tree which is obtained by removing the leaf  $x$  and the edge that contains it from  $T$  and suppressing the remaining degree 2 vertex.
- Suppose that  $\mathcal{T}$  is a subset of  $\binom{X}{2}$ . We say that a triple  $abc \in \binom{X}{3}$  *supports* a vertex  $v \in \mathring{V}$  (relative to  $\mathcal{T}$ ) if we can select leaves  $a, b, c \in X$ , one from each component of  $T - v$ , such that  $ab, ac, bc \in \mathcal{T}$ .
- We call a subset  $\mathcal{T} \subseteq \binom{X}{2}$  a *triplet cover* for  $T$  if, for each element  $v \in \mathring{V}$ , there is some element in  $\binom{X}{3}$  that supports  $v$  (relative to  $\mathcal{T}$ ). An example is shown in Fig. 1. We call each element in  $\mathcal{T}$  a *cord*.
- Given a subset  $\mathcal{T} \subseteq \binom{X}{2}$  and  $x \in X$ , we let  $\mathcal{T}^{-x} = \mathcal{T} - \{xa : a \in X - \{x\} \text{ and } xa \in \mathcal{T}\}$ . In other words,  $\mathcal{T}^{-x}$  is the subset of  $\mathcal{T}$  obtained by removing from  $\mathcal{T}$  precisely those cords which contain  $x$ .

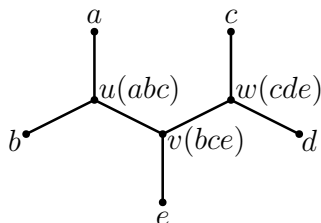


Figure 1: A binary phylogenetic  $X$ -tree  $T$  for  $X = \{a, b, c, d, e\}$ , and, in parentheses, triples of elements from  $X$  whose medians correspond to the indicated interior vertices. The corresponding triplet cover is  $\mathcal{T} = \{ab, ac, bc, be, ce, cd, de\}$ .

## 2. Tree distances from any triplet cover determines the underlying binary tree

The *triplet cover question* posed in [3] and discussed further in [5], asks the following. Suppose that  $\mathcal{T}$  is a triplet cover for a binary phylogenetic tree  $X$ -tree  $T$  having strictly positive edge lengths. Then is  $T$  the only phylogenetic

$X$ -tree (up to isomorphism) with strictly positive edge lengths that can induce the same distance values on pairs of leaves chosen from  $\mathcal{T}$ ?

It is fairly straightforward to show that the triplet cover question has an affirmative answer if we impose the additional restriction on each possible alternative tree  $T'$  that  $\mathcal{T}$  is also a triplet cover for  $T'$  (Proposition 1 of [3]). However, this additional restriction cannot be assumed *a priori* for the general question. The following theorem provides an affirmative answer to the triplet cover question in general.

To state this theorem, we require one further definition. Given a phylogenetic  $X$ -tree  $T$  and an assignment  $\ell$  of strictly positive lengths to the edges of  $T$ , let  $d_{(T,\ell)}(x,y)$  denote the total length of the path between  $x$  and  $y$  in  $T$  (i.e. the sum of the lengths of the edges in the unique path in  $T$  that connects  $x$  to  $y$ ).

**Theorem 2.1.** *Suppose that  $\mathcal{T}$  is a triplet cover for a binary phylogenetic  $X$ -tree  $T$ , where  $|X| \geq 3$ , and that  $\ell$  is an assignment of strictly positive edge lengths for  $T$ . Then for any phylogenetic  $X$ -tree  $T'$  and any assignment  $\ell'$  of strictly positive edge lengths for  $T'$ , if  $d_{(T,\ell)}(x,y) = d_{(T',\ell')}(x,y)$  for all  $xy \in \mathcal{T}$  we have  $T \cong T'$  and  $\ell = \ell'$ .*

*Remarks:* It suffices to show that  $T \cong T'$ , since  $\ell = \ell'$  then follows easily (for example, by Proposition 1 of [3]). Note also that Theorem 2.1 is not true when  $|X| = 2$ , as we may take  $T'$  to have a single vertex, and  $T$  to have two vertices joined by an edge of arbitrary length (in this case  $\mathcal{T} = \emptyset$ ).

We prove the other (main) part of the theorem by establishing a slightly more general result (Theorem 2.3) which allows a more streamlined proof-by-contradiction argument based on the assumption that a (minimal) counterexample exists.

The proof of the following result (Proposition 2.2) does not require or use Theorem 2.1 (nor is that theorem implied by Proposition 2.2). However, the two results are complementary since Theorem 2.1 ensures that the reconstructed tree (namely  $T$ ) produced by the algorithm described in Proposition 2.2 is the only tree that can realize  $d$  for all elements  $xy \in \mathcal{T}$ .

**Proposition 2.2.** *There is a polynomial-time algorithm that will reconstruct any binary phylogenetic  $X$ -tree  $T$  together with a strictly positive edge length assignment  $\ell$  given the values of  $d = d_{(T,\ell)}$  on elements  $xy \in \mathcal{T}$  where  $\mathcal{T}$  is a triplet cover for  $T$ .*

*Proof:* To this end, for any leaf  $x$  of  $X$  let

$$\lambda(x) = \frac{1}{2} \min\{d(x, z) + d(x, z') - d(z, z') : xz, xz' \in \mathcal{T}, z \neq z'\}.$$

Notice that the set in the definition of  $\lambda(x)$  is non-empty since  $\mathcal{T}$  is a triplet cover of  $T$  (consider the vertex of  $T$  adjacent to leaf  $x$ ) and so  $\lambda(x)$  can be determined (in polynomial-time) from the values of  $d$  on  $\mathcal{T}$ . Moreover,  $\lambda(x)$  is the length of the pendant edge of  $T$  incident with leaf  $x$ .

CLAIM:  $x$  and  $y$  form a cherry of  $T$  if and only if  $xy \in \mathcal{T}$  and  $d(x, y) = \lambda(x) + \lambda(y)$ .

*Proof of claim:* First suppose that  $x$  and  $y$  form a cherry of  $T$ . We have  $xy \in \mathcal{T}$  since  $\mathcal{T}$  is a triplet cover for  $T$  (consider the interior vertex of  $T$  that is adjacent to  $x$  and  $y$ ). Moreover,  $d(x, y) = \lambda(x) + \lambda(y)$ . Conversely if  $x$  and  $y$  do not form a cherry of  $T$  and if  $xy \in \mathcal{T}$  then  $d(x, y) > \lambda(x) + \lambda(y)$  since there is at least one interior edge, having strictly positive length, in the path between  $x$  and  $y$ . This establishes the claim.

It follows from the claim that from  $\mathcal{T}$  and  $d|_{\mathcal{T}}$  we can identify a cherry  $x, y$  of  $T$ , as well as the length of the pendant edges incident with  $x$  and  $y$  (namely,  $\lambda(x)$  and  $\lambda(y)$ ). Let us now remove  $x$  from  $X$  to give a reduced label set  $X' = X - x$ , remove  $xy$  from  $\mathcal{T}$  and replace each pair  $xz \in \mathcal{T}$  by a pair  $yz$  to obtain a modified set  $\mathcal{T}' \subseteq \binom{X'}{2}$ . For any pair  $xz \in \mathcal{T}$  that has been replaced by  $yz \in \mathcal{T}'$  set  $d'(y, z) = d(x, z) + \lambda(y) - \lambda(x)$ , and let  $d'$  coincide with  $d$  for all other elements of  $\mathcal{T}'$ . Then  $\mathcal{T}'$  is now a triplet cover for the binary tree  $T - x$  with its induced (strictly positive) edge lengths. By induction on  $|X|$  we can continue this cherry identification and deletion process until we obtain a binary phylogenetic tree on just three leaves. By reversing this process the original tree  $T$  and its edge lengths  $\ell$  can then be reconstructed (in polynomial time). ■

### 2.1. Proof of Theorem 2.1

As mentioned, we need only prove that a binary phylogenetic tree  $T$  is uniquely determined (up to isomorphism) by the values of  $d(x, y) = d_{(T, l)}(x, y)$  (for some strictly positive assignment  $l$  of edge lengths to  $T$ ) for pairs  $x, y$  in a triplet cover of  $T$  (the uniqueness of the edge lengths was shown in [3] and is straightforward). Also, it is convenient for the proof to work with a slightly more general class of trees than binary phylogenetic trees, namely ‘binary  $X$ -trees’. To define this class, recall that an  $X$ -tree  $T$  is a tree for which each vertex of degree at most 2 is labelled by at least one element of  $X$ , and each element of  $X$  labels exactly one vertex of the tree. Thus  $T$  may have unlabelled vertices, but these must have degree 3 or more, and  $T$  may have vertices labelled by more than one element of  $X$ . We refer to  $X$  as the *label set* of  $T$ . Thus, a phylogenetic  $X$ -tree is an  $X$ -tree in which only the leaves are labelled, and each leaf is labelled by only a single element of  $X$ .

Deleting an edge  $e$  of a  $X$ -tree and considering the two connected components of  $T - e$  partitions  $X$  into two nonempty sets. Any such bipartition of  $X$  consisting of two blocks, say  $A$  and  $B$ , is called an  $X$ -split and we denote it by writing  $A|B$ . Moreover, the set of splits arising from an  $X$ -tree  $T$  in this way (by cutting edges of  $T$ ) is *pairwise compatible*, which means that, for any two splits  $A|B$  and  $A'|B'$  in the set, at least one of the four intersections  $A \cap A', A \cap B', B \cap A', B \cap B'$  is empty. It is a classical result (due to Peter Buneman) that  $X$ -trees (up to a natural notion of isomorphism) are in bijective correspondence with the sets of pairwise compatible  $X$ -splits, with each  $X$ -split in the set corresponding to a unique edge of the associated tree. Thus two  $X$ -trees  $T$  and  $T'$  are isomorphic, written  $T \cong T'$ , precisely if they have the same set of  $X$ -splits. This definition agrees with the earlier notion of isomorphism when restricted to binary phylogenetic  $X$ -trees.

We will say that an  $X$ -tree  $T$  is *binary* if (i) each vertex of degree at most 1 is labelled by one or two elements of  $X$ , (ii) each vertex of degree 2 is labelled by exactly one element of  $X$ , and (iii) all remaining vertices are unlabelled and have degree 3. Note that if we take any binary phylogenetic  $X$ -tree and collapse

any (possibly empty) subset of pendant edges of this tree, we obtain a binary  $X$ -tree; conversely, each binary  $X$ -tree  $T$  is obtained from a unique binary phylogenetic  $X$ -tree  $T_B$  by collapsing a unique subset of pendant edges of  $T$ . Note that  $T_B$  is the  $X$ -tree whose set of  $X$ -splits consists of the  $X$ -splits of  $T$  together with any *trivial  $X$ -splits* (i.e.  $x|X - x, x \in X$ ) not already present in  $T$ . This is illustrated in Fig. 2.

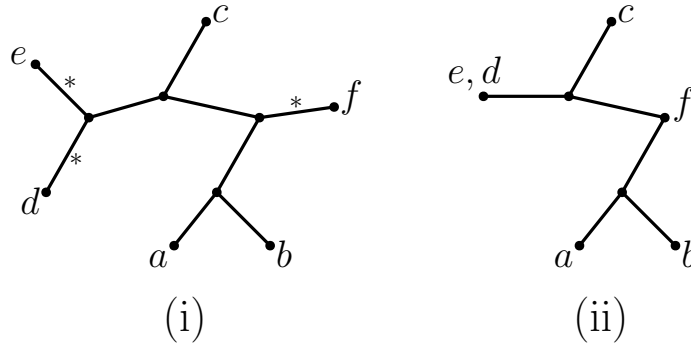


Figure 2: (i) A binary phylogenetic  $X$ -tree for  $X = \{a, b, c, d, e, f\}$ . If we collapse the three edges indicated by \* we obtain the binary  $X$ -tree shown in (ii).

It is straightforward to extend the notion of triplet cover from a binary phylogenetic  $X$ -tree to a binary  $X$ -tree: We say that a subset  $\mathcal{T}$  of  $\binom{X}{2}$  is a *triplet cover* of a binary  $X$ -tree  $T$ , if  $\mathcal{T}$  is a triplet cover (in the usual sense) of the associated binary phylogenetic  $X$ -tree  $T_B$ .

Given an  $X$ -tree  $T$  and a function  $\ell$  that assigns strictly positive lengths to each edge of  $T$ , let  $d_{(T,\ell)} : X \times X \rightarrow \mathbb{R}^{\geq 0}$  be the induced distance function on  $X$  in which  $d_{(T,\ell)}(x, y)$  is the sum of the lengths of the edges on the (unique) path in  $T$  connecting the vertices of  $T$  labelled by  $x$  and  $y$ . Notice that  $d_{(T,\ell)}$  takes the value 0 if and only if  $x$  and  $y$  label the same vertex of  $T$ . The function  $d_{(T,\ell)}$  is non-negative, symmetric and satisfies the triangle inequality, and so  $d_{(T,\ell)}$  is a pseudometric on  $X$ .

**Theorem 2.3.** *Suppose that  $\mathcal{T}$  is a triplet cover for a binary  $X$ -tree  $T$ , where  $|X| \geq 3$ , and that  $\ell$  is an assignment of strictly positive edge lengths for  $T$ .*



Then for any  $X$ -tree  $T'$  and any assignment  $\ell'$  of strictly positive edge lengths for  $T'$ , if  $d_{(T,\ell)}(x,y) = d_{(T',\ell')}(x,y)$  for all  $xy \in \mathcal{T}$  we have  $T \cong T'$ .

*Proof:* The theorem is readily verified for  $|X| = 3$ . Suppose there is a counterexample to Theorem 2.3 when  $|X| \geq 4$ . In that case, we can select a counterexample — say  $\mathcal{T}, T, T'$  — for which (i)  $|X| \geq 4$  is minimal (call this minimal value  $n$ ) and (ii) within all counterexamples with  $|X| = n$  the sum of the number of edges in  $T$  and in  $T'$  is minimal. We will show that we can then always construct another counterexample that either (i) has a label set of size  $n - 1 \geq 3$ , or has a smaller total number of edges across the two trees, contradicting the minimality assumptions, or (ii) has a label set of size 3, for which the result holds. Such a contradiction implies that no counterexample  $\mathcal{T}, T, T'$  to Theorem 2.3 can exist.

We first establish the following claims:

- (i) There is no trivial  $X$ -split present in both  $T$  and  $T'$ .
- (ii)  $T'$  does not contain any trivial  $X$ -split.

To establish Claim (i), suppose that some trivial  $X$ -split, say  $x|X - x$  (for some  $x \in X$ ) is present in both  $T$  and  $T'$ . Let  $e_x$  and  $e'_x$  denote the pendant edges of  $T$  and  $T'$  that are incident with  $x$ , let  $\ell(e_x), \ell'(e'_x)$  denote their lengths, and let  $\ell_x = \min\{\ell(e_x), \ell'(e'_x)\}$ . If  $\ell(e_x) = \ell'(e'_x)$  then collapse the pendant edges  $e_x$  and  $e'_x$ , while if  $\ell(e_x) \neq \ell'(e'_x)$  then collapse the shorter of the two pendant edges  $e_x$  and  $e'_x$ , and reduce the length of the other pendant edge by  $\ell_x$ . The resulting pair of modified trees still consists of a binary  $X$ -tree and an  $X$ -tree, both with strictly positive edge lengths, and the distance between  $x$  and any other element of  $X$  in both  $T$  and in  $T'$  has been reduced by  $\ell_x$  while all other distances remain the same. Thus, the sum of the number of edges in  $T$  and  $T'$  has been reduced by at least 1. However the two modified trees, along with the original triplet cover  $\mathcal{T}$  provide a smaller counterexample (in terms of the sum of the number of edges in  $T$  and  $T'$ ) violating the minimality assumption. This establishes Claim (i).

To establish Claim (ii), assume that there is a trivial  $X$ -split  $x|X - x$  (for

some  $x \in X$ ) present in  $T'$ . By Claim (i),  $x|X - x$  is not present in  $T$ , and so either  $x$  labels an interior vertex of  $T$ , or else  $x$  together with another element of  $X$ , say  $y$ , labels a leaf of  $T$ . In either case, since  $\mathcal{T}$  is a triplet cover for  $T$  there is a triplet  $wxy$  with  $\{wx, wy, xy\} \subseteq \mathcal{T}$  and with  $d_{(T,\ell)}(w, x) + d_{(T,\ell)}(x, y) = d_{(T,\ell)}(w, y)$ . But then (since  $d_{(T,\ell)}(x', x'') = d_{(T',\ell')}(x', x'')$  for all  $x', x'' \in \mathcal{T}$ ) it follows that  $T'$  does not have  $x$  as the sole label of one of its leaves (i.e.  $T'$  does not contain  $x|X - x$ ). Hence,  $T'$  does not contain any trivial  $X$ -split. This establishes Claim (ii).

Now, let  $v$  be any leaf of  $T'$ , and let  $Y \subset X$  denote its label set (note that  $T'$  cannot be a single-vertex tree since otherwise  $T$  would be also, and so  $|X| \leq 2$ , which violates our assumption that  $n > 3$ ). By Claim (ii),  $|Y| > 1$ . Thus one of the following two cases must apply:

**Case A:**  $T$  contains all trivial  $X$ -splits  $y|X - y$  for all  $y \in Y$ .

**Case B:** There exists some  $y \in Y$  for which the trivial  $X$ -split  $y|X - y$  is absent from  $T$ .

(We will show that neither case can arise, which will furnish the required contradiction).

In Case A, let  $e'_v$  denote the pendant edge of  $T'$  that is incident with  $v$ , and for each  $y \in Y$ , let  $\ell(e_y)$  be the length of the pendant edge of  $T$  incident with the leaf labelled by (only)  $y$ . Let

$$\ell_Y = \min[\{\ell(e_y) : y \in Y\} \cup \{\ell'(e'_v)\}].$$

If  $\ell'(e'_v) = \ell_Y$  then collapse edge  $e'_v$  in  $T'$ , otherwise reduce the length of  $e'_v$  by  $\ell_Y$ . Similarly, for each  $y \in Y$  for which  $\ell(e_y) = \ell_Y$  collapse edge  $e_y$  of  $T$ , otherwise reduce the length of  $e_y$  by  $\ell_Y$ . Again the resulting modified pair of trees still consists of a binary  $X$ -tree and an  $X$ -tree, both with strictly positive edge lengths, but either  $T$  or  $T'$  has at least one (pendant) edge fewer than before. Moreover, the distance between two distinct elements  $x, x'$  of  $X$  in either tree is either unchanged (if neither  $x$  nor  $x'$  is in  $Y$ ) or is reduced by  $\ell_Y$  for both trees, when  $|\{x, x'\} \cap Y| = 1$ . For the remaining case where  $x, x' \in Y$  the

distances in the modified trees may differ, however we also have that  $xx' \notin \mathcal{T}$ , since the distance between  $x$  and  $x'$  in  $T$  and in  $T'$  is different (it is zero in  $T'$  and non-zero in  $T$ ). It follows that the modified trees again provide a smaller counterexample (in terms of the sum of the number of edges in  $T$  and in  $T'$ ) violating the minimality assumption. This shows that Case A cannot arise.

For Case B, there exists some element  $y \in Y$  that labels either an interior vertex of  $T$ , or else  $y$  together with another element of  $X$ , say  $x$ , labels a leaf of  $T$ . In either case, since  $\mathcal{T}$  is a triplet cover for  $T$  there is a triplet  $xyz$  with  $\{xy, xz, yz\} \subseteq \mathcal{T}$  and with  $d_{(T,\ell)}(x,y) + d_{(T,\ell)}(y,z) = d_{(T,\ell)}(x,z)$ . Thus, since  $v$  is a leaf of  $T'$  we must have that  $x$  or  $z$  is in  $Y$ . Without loss of generality we may assume that  $x \in Y$ . Then since  $xy \in \mathcal{T}$  it follows that  $d_{(T,\ell)}(x,y) = d_{(T',\ell)}(x,y) = 0$ . In particular,  $\{x,y\}$  is contained in the label set of a leaf  $u$  of  $T$ .

Finally, delete label  $x$  to obtain a set  $X' = X - \{x\}$  of size  $|X| - 1$ , and form modified  $X'$ -trees  $\tilde{T}$  and  $\tilde{T}'$  from  $T$  and  $T'$  (respectively) by deleting the label  $x$  from  $u$  and  $v$ . Then  $\tilde{T}$  is a binary  $X'$ -tree,  $\tilde{T}'$  an  $X'$ -tree, and both trees inherit strictly positive edge lengths from  $T$  and  $T'$ . Consider the modified set  $\tilde{\mathcal{T}} \subseteq \binom{X'}{2}$  obtained from  $\mathcal{T}$  by deleting  $xy \in \mathcal{T}$  and replacing each remaining occurrence of  $x$  by  $y$  in any pair  $xz \in \mathcal{T}$ . Then  $\tilde{\mathcal{T}}$  is a triplet cover for  $\tilde{T}$ , and if  $x'x'' \in \tilde{\mathcal{T}}$  then the distance between  $x'$  and  $x''$  in  $\tilde{T}$  is the same as it is in  $\tilde{T}'$ . Consequently,  $(\tilde{\mathcal{T}}, \tilde{T}, \tilde{T}')$  provides a counterexample to Theorem 2.3. However, this new counterexample has a label set  $X'$  of size  $|X'| = n - 1$  which is one less than the starting counterexample  $(\mathcal{T}, T, T')$ . If  $n - 1 = 3$  this is impossible, since the Theorem holds when the label set has size 3, while if  $n - 1 \geq 4$  we have violated the minimality assumption regarding the choice of  $(\mathcal{T}, T, T')$ . This shows that Case B cannot arise, thereby completing the proof. ■

### 3. Properties of minimal triplet covers

In this section, we collect together a number of definitions, observations, and results (extending earlier work from [4]) that are required for establishing some further main results later in the paper.

#### 3.1. Preliminaries

In the remainder of this paper, unless stated to the contrary, we will assume that  $T$  refers to a **binary** phylogenetic  $X$ -tree. We will also write  $\mathring{V}$  for  $\mathring{V}(T)$  (the set of interior vertices of  $T$ ) when  $T$  is clear.

Suppose that  $\mathcal{T}$  is a triplet cover of  $T$ . The following terminology and result is from [4].

- For  $x \in X$  the *multiplicity*  $\mu_{\mathcal{T}}(x)$  of  $x$  (relative to  $\mathcal{T}$ ) is the number of cords in  $\mathcal{T}$  that contain  $x$ . The *multiplicity of  $\mathcal{T}$*  is  $\mu(\mathcal{T}) = \min_{x \in X} \mu_{\mathcal{T}}(x)$ .
- $|\mathcal{T}| \geq 2|X| - 3$  (for a direct proof see [4, Proposition 3]). We call  $\mathcal{T}$  *minimum* if  $|\mathcal{T}| = 2|X| - 3$ .
- We call a triplet cover  $\mathcal{T}$  of  $T$  *minimal* if  $\mathcal{T} - \{t\}$  is not a triplet cover for  $T$  for all  $t \in \mathcal{T}$ .

Note that there exist triplet covers that are minimal but not minimum (an example is given in Fig. 3).

The following lemma summarizes some results established in [4] (namely, Corollary 1, Proposition 2, and Corollary 2 of that paper, respectively).

**Lemma 3.1.** *Suppose that  $\mathcal{T}$  is a minimal triplet cover of  $T$ . Then*

- (i)  $2|X| - 3 \leq |\mathcal{T}| \leq 3|X| - 6$ .
- (ii)  $2 \leq \mu(\mathcal{T}) \leq 5$ .
- (iii) *If  $\mathcal{T}$  is a minimum triplet cover, then  $\mu(\mathcal{T}) = 2$ .*

Given a subset  $\mathcal{T} \subseteq \binom{X}{2}$  and  $v \in \mathring{V}$ , we let  $S_v(\mathcal{T})$  be the subset of  $\binom{X}{3}$  which consists of precisely those triples which support  $v$ . We call  $S_v(\mathcal{T})$  the *support of  $v$  (relative to  $\mathcal{T}$ )*.

Note that if  $abc \in S_v(\mathcal{T})$ , some  $v \in \mathring{V}$ , then  $v = \text{med}_T(a, b, c)$ , where for all  $xyz \in \binom{X}{3}$ ,  $\text{med}_T(x, y, z)$  denotes the *median* of  $x, y, z$  (i.e. the unique vertex that lies on all of the shortest paths between  $x, y$  and  $z$ ).

**Lemma 3.2.**

- (i)  $\mathcal{T}$  is a triplet cover of  $T$  if and only if  $|S_v(\mathcal{T})| \geq 1$  for all  $v \in \mathring{V}$ .
- (ii) If  $\mathcal{T}$  is a triplet cover of  $T$  and  $v, w \in \mathring{V}$  distinct, then  $S_v(\mathcal{T}) \cap S_w(\mathcal{T}) = \emptyset$ .

*Proof:* The proof of Part (i) is straightforward. For Part (ii), if this were not the case, then for  $xyz \in S_v(\mathcal{T}) \cap S_w(\mathcal{T})$  we would have  $v = \text{med}_T(x, y, z) = w$ , a contradiction. ■

3.2. The cover graph  $\Gamma(\mathcal{T})$  and triplet set  $\mathcal{C}(\mathcal{T})$

Given subset  $\mathcal{T} \subseteq \binom{X}{2}$ , the *cover graph (of  $\mathcal{T}$ )*, denoted  $\Gamma(\mathcal{T})$ , is the graph with vertex set  $X$  and edge set  $\mathcal{T}$ . This graph, introduced in [3], has played an important role in subsequent papers [7], [5], [4].

We now consider a set which will be useful for understanding the triangles (i.e. 3-cycles) in a cover graph. Given a subset  $\mathcal{T} \subseteq \binom{X}{2}$ , we define

$$\mathcal{C}(\mathcal{T}) = \dot{\bigcup}_{v \in \mathring{V}} S_v(\mathcal{T}) \subseteq \binom{X}{3}.$$

Note that the union in this definition is disjoint by Lemma 3.2(ii). In addition we note some other useful properties of the set  $\mathcal{C}(\mathcal{T})$ .

**Lemma 3.3.**

- (i) If  $\mathcal{T}$  is a triplet cover for  $T$ , then  $\bigcup \mathcal{C}(\mathcal{T}) = X$ , and  $|\mathcal{C}(\mathcal{T})| \geq |X| - 2$ .
- (ii) If  $\mathcal{T}$  is a minimal triplet cover, then every cord in  $\mathcal{T}$  is a subset of some element of  $\mathcal{C}(\mathcal{T})$ .

*Proof:* For Part (i), if  $x \in X$ , then let  $v$  be the vertex in  $T$  adjacent to  $x$ . If  $A \in S_v(\mathcal{T})$ , then clearly  $x \in A$ . The inequality now follows from Lemma 3.2(ii). For Part (ii), suppose that  $\mathcal{T}$  is minimal and that there is a cord  $xy \in \mathcal{T}$  that is not a subset of any element in  $\mathcal{C}(\mathcal{T})$ . Then, for all  $v \in \mathring{V}$ ,  $xy \in \mathcal{T}$  is not a subset of any element of  $S_v(\mathcal{T})$ . It follows by Lemma 3.2(i) that  $\mathcal{T} - \{xy\}$  is a triplet cover for  $T$ , a contradiction. ■

We now collect together some important properties of the cover graph.

**Theorem 3.4.** *Suppose that  $\mathcal{T}$  is a triplet cover for  $T$ .*

- (i) *The triangles in the cover graph  $\Gamma(\mathcal{T})$  are in bijective correspondence with the elements of  $\mathcal{C}(\mathcal{T})$ .*
- (ii)  *$\Gamma(\mathcal{T})$  is 2-connected.<sup>1</sup>*
- (iii) *If  $\mathcal{T}$  is a minimal triplet cover for  $T$ , then every cord in  $\mathcal{T}$  is the edge of some triangle in  $\Gamma(\mathcal{T})$ .*

*Proof: Part (i):* Suppose  $xyz \in \mathcal{C}(\mathcal{T})$ , and so  $xyz \in S_v(\mathcal{T})$ , for some  $v \in \mathring{V}$ . Then clearly  $x, y, z$  is a triangle in  $\Gamma(\mathcal{T})$ , since  $xy, yz, xz \in \mathcal{T}$ . Thus, we have a map  $\psi$  that takes elements in  $\mathcal{C}(\mathcal{T})$  to triangles in  $\Gamma(\mathcal{T})$ . Clearly this map is injective. Moreover, if  $x, y, z$  is a triangle in  $\Gamma(\mathcal{T})$ , then for  $v = \text{med}_T(x, y, z)$ , we have  $xyz \in S_v(\mathcal{T})$  and  $\psi(xyz) = x, y, z$ . Thus  $\psi$  is surjective.

*Part (ii):* The statement clearly holds if  $|X| = 3$ , and so we assume  $|X| \geq 4$ . Suppose  $x \in X$ . Let  $v \in V(T)$  be the vertex in  $T$  adjacent to  $x$ . Let  $w \neq x$  be a vertex adjacent to  $v$  in  $T$ , and let  $T_w$  be the tree which is the connected component of  $T$  minus the edge  $\{v, w\}$  that contains  $w$ . Let  $Y_w = V(T_w) \cap X$ . We claim that the graph induced by  $\Gamma(\mathcal{T})$  on  $Y_w$  is connected.

---

<sup>1</sup>The connectivity of  $\Gamma(\mathcal{T})$  also follows from [3] (Proposition 1 and Corollary 3).

Consider  $T_w$  as being a rooted, directed tree, with root  $w$  and all edges directed away from  $w$ . For  $u \in V(T_w)$  we let  $Y_u$  denote the set of leaves  $x$  in  $X$  for which  $u$  lies on the path in  $T_w$  from the root vertex  $w$  to  $x$ . We now prove the claim using induction on  $|Y_u|$ . If  $|Y_u| = 1$ , then clearly the graph induced by  $\Gamma(\mathcal{T})$  on  $Y_u$  is connected. Now, suppose  $u \in V(T_w)$  with  $|Y_u| > 1$ . Let  $u', u''$  be the children of  $u$  in  $T_w$ . By induction, we can assume that the graphs induced by  $\Gamma(\mathcal{T})$  on  $Y_{u'}$  and on  $Y_{u''}$  are connected. But as  $\mathcal{T}$  is a triplet cover for  $T$ , there must exist a cord  $y'y'' \in \mathcal{T}$  with  $y' \in Y_{u'}$  and  $y'' \in Y_{u''}$ . So the graph induced by  $\Gamma(\mathcal{T})$  on  $Y_u = Y_{u'} \cup Y_{u''}$  is connected. The claim now follows as  $w \in V(T_w)$ .

Now, let  $xyz$  be a triple in the support  $S_v(\mathcal{T})$  of  $v$ , for some  $y, z \in X$ , which must exist as  $\mathcal{T}$  is a triplet cover for  $T$ . Let  $w', w''$  denote the two vertices in  $T$  adjacent to  $v$  that are not equal to  $x$ . Then as  $xyz \in S_v(\mathcal{T})$ , it follows without loss of generality that  $y \in Y_{w'}$  and  $z \in Y_{w''}$ . Moreover, as the graphs induced by  $\Gamma(\mathcal{T})$  on  $Y_{w'}$  and  $Y_{w''}$  are both connected by the above claim,  $x \notin Y_{w'}, Y_{w''}$  and  $\{y, z\}$  is an edge in  $\Gamma(\mathcal{T})$ , it immediately follows that the graph  $\Gamma(\mathcal{T}) - x$  obtained by removing  $x$  and the edges which contain it is connected. Note that such edges must exist as  $xy, xz \in \mathcal{T}$ . As the selection of  $x \in X$  was arbitrary, it follows that  $\Gamma(\mathcal{T}) - x$  is connected for all  $x \in X$ , i.e.  $\Gamma(\mathcal{T})$  is 2-connected.

*Part (iii):* This follows by Part (i) and Lemma 3.3(ii). ■

### 3.3. Sparse triplet covers and Hall-type subsets of triples

We call a triplet cover  $\mathcal{T}$  of  $T$  *sparse* if  $|\mathcal{C}(\mathcal{T})| = |X| - 2$  (cf. Lemma 3.3(i)). Note that if  $\mathcal{T}$  is a sparse cover of  $T$ , then the function  $f_{\mathcal{T}} : \mathcal{C}(\mathcal{T}) \rightarrow \mathring{V}$ , which maps  $xyz \in \mathcal{C}(\mathcal{T})$  to  $\text{med}_{\mathcal{T}}(x, y, z)$  is a bijection. From this observation, it is possible to construct minimal triplet covers that are not sparse (an example is provided in Fig. 3).

Note also that there are sparse triplet covers that are not minimal (an example is to add a new cord  $eh$  into the triplet cover  $\mathcal{T}$  in Fig. 4).

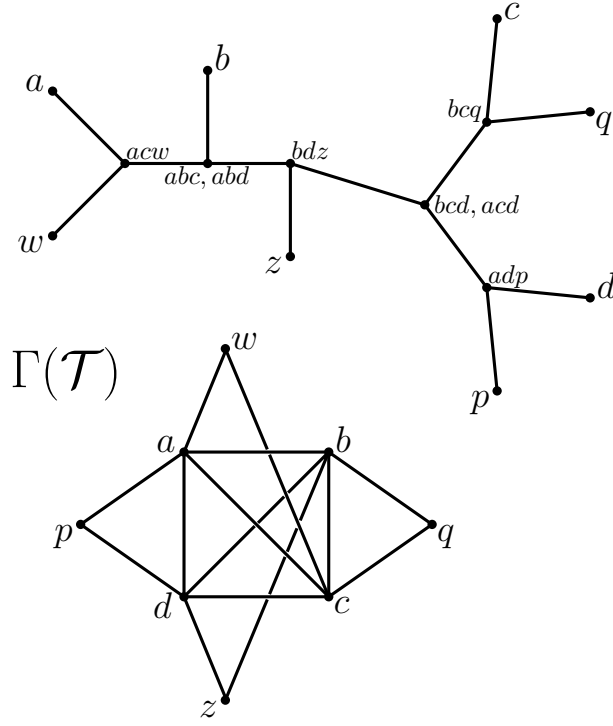


Figure 3: Top: A minimal triplet cover that is not sparse. Bottom: The associated cover graph for  $\mathcal{T}$ .

We say that a subset  $\mathcal{C} \subseteq \binom{X}{3}$  is of *Hall-type* if  $\bigcup \mathcal{C} = X$  and  $\mathcal{C}$  satisfies the following property (cf. [6]): For all non-empty subsets  $\mathcal{C}' \subseteq \mathcal{C}$ ,

$$\left| \bigcup \mathcal{C}' \right| \geq |\mathcal{C}'| + 2.$$

For example, for the triplet cover  $\mathcal{T}$  in Fig. 4, the set  $\mathcal{C}(\mathcal{T})$  is of Hall-type.

**Lemma 3.5.** *Suppose  $\mathcal{T}$  is a triplet cover of  $T$ . If  $\mathcal{T}$  is sparse, then  $\mathcal{C}(\mathcal{T})$  is of Hall-type.*

*Proof:* If  $\mathcal{T}$  is sparse then the map  $f_{\mathcal{T}}$  is a bijection. Hence, since  $\bigcup \mathcal{C}(\mathcal{T}) = X$  by Lemma 3.3(ii), it follows from [6, Theorem 1.1] that  $\mathcal{C}(\mathcal{T})$  is of Hall-type. ■



### 3.4. Sections

A subset  $\mathcal{C} \subseteq \mathcal{C}(\mathcal{T})$  is called a *section* (of  $\mathcal{C}(\mathcal{T})$ ) if  $|\mathcal{C} \cap S_v(\mathcal{T})| = 1$  for all  $v \in \mathring{V}$ . Note that  $\mathcal{C}(\mathcal{T})$  always contains a section and that if  $\mathcal{C}$  is a section, then  $|\mathcal{C}| = |X| - 2$

We now define a set whose properties will be useful later on. Given a subset  $\mathcal{C} \subseteq \binom{X}{3}$ , we define the *cord set* of  $\mathcal{C}$  to be

$$Co(\mathcal{C}) = \{xy \in \binom{X}{2} : xy \subset A \text{ some } A \in \mathcal{C}\}.$$

Note that clearly  $Co(\mathcal{C}(\mathcal{T})) \subseteq \mathcal{T}$ , but that in general  $\mathcal{T}$  is not necessarily a subset of  $Co(\mathcal{C}(\mathcal{T}))$  (for example, if we add in a new cord  $eh$  to the triplet cover  $\mathcal{T}$  in Fig. 4, we obtain a new triplet cover  $\mathcal{T}'$  for which  $\mathcal{T}'$  is not a subset of  $Co(\mathcal{C}(\mathcal{T}'))$ ). Moreover, if  $\mathcal{C}$  is a section of  $\mathcal{C}(\mathcal{T})$ , then  $Co(\mathcal{C}) \subseteq \mathcal{T}$ .

**Proposition 3.6.** *Suppose that  $\mathcal{T}$  is a triplet cover for  $T$ . The following hold:*

- (a) *If  $\mathcal{T}$  is a minimal triplet cover for  $T$ , then  $Co(\mathcal{C}(\mathcal{T})) = \mathcal{T}$ .*
- (b)  *$\mathcal{T}$  is sparse if and only if  $\mathcal{C}(\mathcal{T})$  is of Hall-type.*
- (c) (i) *If  $\mathcal{C}$  is a section of  $\mathcal{C}(\mathcal{T})$ , then  $Co(\mathcal{C})$  is a triplet cover for  $T$ . In particular,  $\bigcup \mathcal{C} = X$ .*
- (ii) *If  $\mathcal{C}$  is a section of  $\mathcal{C}(\mathcal{T})$ , then  $\mathcal{C}$  is of Hall-type.*
- (iii)  *$\mathcal{T}$  is minimal if and only if  $Co(\mathcal{C}) = \mathcal{T}$  for every section  $\mathcal{C}$  of  $\mathcal{C}(\mathcal{T})$ .*
- (iv)  *$\mathcal{T}$  is sparse if and only if  $\mathcal{C}(\mathcal{T})$  has a unique section.*

*Proof: Part (a):* Since  $\mathcal{T}$  is minimal, for all  $xy \in \mathcal{T}$ ,  $xy \subset A$  for all  $A \in S_v(\mathcal{T})$ , some  $v \in \mathring{V}$  (otherwise we could remove  $xy$  from  $\mathcal{T}$  and still have a triplet cover). So  $\mathcal{T} \subseteq Co(\mathcal{C}(\mathcal{T}))$ . As remarked above, the reverse inequality is obvious.

*Part (b):* By Lemma 3.5 it suffices to prove that if  $\mathcal{C}(\mathcal{T})$  is of Hall-type, then  $\mathcal{T}$  is sparse. Suppose for contradiction that  $\mathcal{T}$  is not sparse. Let  $\mathcal{C}$  be a section of  $\mathcal{C}(\mathcal{T})$ , so that  $|\mathcal{C}| = |X| - 2$  and so (since  $\mathcal{C}(\mathcal{T})$  is of Hall-type) we have  $\bigcup \mathcal{C} = X$ .

Since  $\mathcal{T}$  is not sparse  $|\mathcal{C}(\mathcal{T})| > |X| - 2$ . Hence there is some  $t \in \mathcal{C}(\mathcal{T})$  that is not in  $\mathcal{C}$ . Let  $\mathcal{C}' = \mathcal{C} \cup \{t\}$ . Since  $\mathcal{C}(\mathcal{T})$  is of Hall-type and  $\mathcal{C}' \subseteq \mathcal{C}(\mathcal{T})$ ,

$$|X| = \left| \bigcup \mathcal{C}' \right| \geq |\mathcal{C}'| + 2 = (|X| - 2 + 1) + 2 = |X| + 1,$$

a contradiction.

*Part (c-i):* If  $v \in \mathring{V}(T)$ , then there exists  $xyz \in \mathcal{C}$ ,  $x, y, z \in X$ , with  $xyz \in S_v(\mathcal{T})$ , and so  $xy, yz, zx \in Co(\mathcal{C})$ . Hence  $S_v(Co(\mathcal{C})) \neq \emptyset$ . The statement now follows from Lemma 3.2(i).

*Part (c-ii):* We can think of a section  $\mathcal{C}$  of  $\mathcal{C}(\mathcal{T})$  as being a bijective map  $f_{\mathcal{C}} : \mathring{V} \rightarrow \mathcal{C}$  which for each  $v \in \mathring{V}$  selects some element in  $S_v(\mathcal{T})$  (the inverse of  $f_{\mathcal{C}}$  is the map which takes each  $xyz \in \mathcal{C}$ ,  $x, y, z \in X$ , to  $\text{med}_T(x, y, z)$ ). Statement (ii) now follows immediately from [6, Theorem 1.1].

*Part (c-iii):* Suppose  $\mathcal{T}$  is minimal. In view of Proposition 3.6(c-i)  $\mathcal{T} = Co(\mathcal{C})$  must clearly hold whenever  $\mathcal{T}$  is minimal.

Conversely, suppose that  $Co(\mathcal{C}) = \mathcal{T}$  for every section  $\mathcal{C}$  of  $\mathcal{C}(\mathcal{T})$ . Suppose  $\mathcal{T}$  is not minimal. Then there exists some  $xy \in \mathcal{T}$ ,  $x, y \in X$ , such that  $\mathcal{T} - \{xy\}$  is a triplet cover for  $T$ . Thus for all  $v \in \mathring{V}$ , we have  $xy$  is not contained in some  $A \in S_v(\mathcal{T})$ . So, we can choose a section  $\mathcal{C}$  of  $\mathcal{C}(\mathcal{T})$  in which  $xy$  is not a subset of any element of  $\mathcal{C}$ . But then  $xy \notin Co(\mathcal{C})$ , which contradicts the assumption that  $Co(\mathcal{C}) = \mathcal{T}$  for all sections  $\mathcal{C}$  of  $\mathcal{C}(\mathcal{T})$ .

*Part (c-iv):* Clearly, if  $\mathcal{T}$  is sparse, then  $\mathcal{C}(\mathcal{T})$  contains a unique section. Conversely, if  $\mathcal{C}(\mathcal{T})$  has a unique section, then  $|\mathcal{C}(\mathcal{T})| = |X| - 2$ , and so  $\mathcal{T}$  is sparse. ■

### 3.5. The range of $\mu(\mathcal{T})$

We saw in Lemma 3.1 that when  $\mu(\mathcal{T})$  is a minimum triplet cover for  $T$  then  $\mu(\mathcal{T}) = 2$ . However, for minimal triplet covers,  $\mu(\mathcal{T})$  can be larger. For

example, there exists a sparse triplet cover  $\mathcal{T}$  for some  $T$  with  $\mu(\mathcal{T}) = 4$  (see Fig. 4; note that in this example  $\mathcal{T}^{-x}$  is not a triplet cover for  $T - x$  for any  $x \in X$ ). Our main result of this section is that for a minimal triplet cover  $\mathcal{T}$  for  $T$  then  $\mu(\mathcal{T})$  must lie within these two extreme values.

**Proposition 3.7.** *If  $\mathcal{T}$  is a minimal triplet cover for  $T$ , then  $2 \leq \mu(\mathcal{T}) \leq 4$ .*

*Proof:* We use an argument similar to the proof of (3)  $\Rightarrow$  (2) of [6, Theorem 1.1].

Let  $\mathcal{C}$  be a section of  $\mathcal{C}(\mathcal{T})$ , which is of Hall-type by Proposition 3.6(c-ii). Put  $n = |X|$ . For  $x \in X$ , let  $n_{\mathcal{C}}(x)$  be the number of triplets in  $\mathcal{C}$  containing  $x$ . If there exists some  $x \in X$  such that  $n_{\mathcal{C}}(x) = 1$ , then  $\mu_{\mathcal{T}}(x) = 2$  by the ‘only if’ direction of Proposition 3.6(c-iii). Hence  $\mu(\mathcal{T}) = 2$ . Thus we may suppose that  $n_{\mathcal{C}}(x) > 1$  for all  $x \in X$ . Let  $\Omega = \{(x, S) \in X \times \mathcal{C} : x \in S\}$ . Then

$$|\Omega| = \sum_{x \in X} n_{\mathcal{C}}(x) \geq 2k + 3(n - k),$$

where  $k = |\{x \in X : n_{\mathcal{C}}(x) = 2\}|$ , and, since  $\mathcal{C}$  is a section of  $\mathcal{C}(\mathcal{T})$ ,

$$|\Omega| = 3|\mathcal{C}| = 3(n - 2).$$

Hence  $2k + 3(n - k) \leq 3n - 6$ , and so  $k \geq 6$ . Hence there exists some  $x \in X$  with  $n_{\mathcal{C}}(x) = 2$ . Thus  $\mu_{\mathcal{T}}(x) \leq 4$  and so  $\mu(\mathcal{T}) \leq 4$  by again invoking the ‘only if’ direction of Proposition 3.6(c-iii). ■

#### 4. 2-tree decompositions

A graph  $H = (V, E)$  with  $|V| \geq 3$  is called a *2-tree* if there exists an ordering  $v_1, v_2, \dots, v_q$  of  $V$  such that  $\{v_1, v_2\} \in E$  and, for  $i = 3, \dots, q$ , the vertex  $v_i$  has degree 2 and belongs to a unique triangle in the subgraph induced by  $H$  on the set  $\{v_1, v_2, \dots, v_i\}$  [2, p.235]. We let  $\Delta(H)$  denote the set consisting of the triangles in  $H$ . Note that  $|E| = 2|V| - 3$  and  $|\Delta(H)| = |V| - 2$  [8, p.227].

A *2-tree decomposition* of a graph  $H = (W, F)$  is a collection  $\{H_i = (W_i, F_i)\}_{i=1}^m$ ,  $m \geq 1$ , of subgraphs of  $H$  (called *blocks*) such that the following hold:

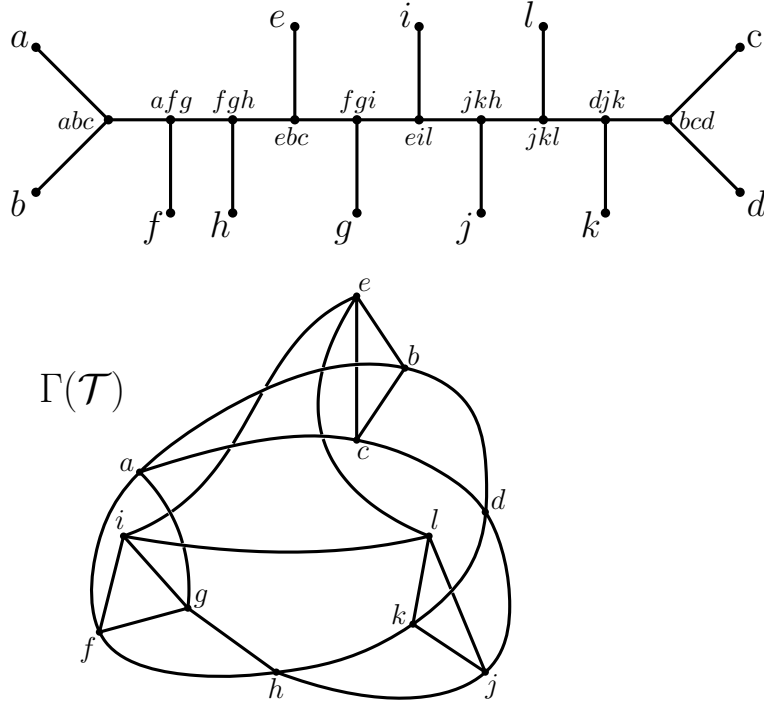


Figure 4: Top: A sparse minimal triplet cover with  $\mu(\mathcal{T}) = 4$ . Bottom: The associated cover graph.

- $H_i$  is a 2-tree, and its vertex set  $W_i$  has size at least 3;
- the vertex sets of  $H_1, \dots, H_m$  cover every vertex of  $H$  (i.e.  $W = \bigcup_{i=1}^m W_i$ );
- the edge sets of  $H_1, \dots, H_m$  partition the edge set of  $H$  (i.e.  $F = \bigcup_{i=1}^m F_i$  and the sets  $F_i$  are pairwise disjoint).

We call the decomposition *strict* if the edge set of every triangle in  $H$  is a subset of some  $F_i$ ,  $1 \leq i \leq m$ .

Note that if  $H = (V, E)$  is a 2-tree, then it can be easily shown that any 2-tree decomposition of  $H$  has one block, namely,  $H$  itself.

We now state our second main theorem.

**Theorem 4.1.** *Suppose that  $\mathcal{T}$  is a triplet cover for  $T$ .*

- (a) If  $\mathcal{T}$  is minimal and  $\mathcal{C}$  is a section of  $\mathcal{C}(\mathcal{T})$ , then  $\mathcal{C} = \dot{\cup}_{i=1}^m \Delta(H_i)$  for  $\{H_i = (W_i, F_i)\}_{i=1}^m$  some 2-tree decomposition of  $\Gamma(\mathcal{T})$ . Moreover,  $\{H_i = (W_i, F_i)\}_{i=1}^m$  is the only 2-tree decomposition of  $\Gamma(\mathcal{T})$  with  $\mathcal{C} = \dot{\cup}_{i=1}^m \Delta(H_i)$  and, if  $\mathcal{T}$  is sparse, then this decomposition is a strict 2-tree decomposition.
- (b) The following statements are equivalent:
- (i)  $\Gamma(\mathcal{T})$  has a strict 2-tree decomposition;
  - (ii)  $\Gamma(\mathcal{T})$  has a unique strict 2-tree decomposition;
  - (iii)  $\mathcal{T}$  is minimal and sparse.
- (c) If  $\mathcal{T}$  is minimal, then the following statements are equivalent:
- (i)  $\Gamma(\mathcal{T})$  has a unique 2-tree decomposition;
  - (ii)  $\mathcal{T}$  is sparse.

*Proof: Part (a):* Suppose that  $\mathcal{T}$  is a minimal triplet cover for  $T$ , and let  $\mathcal{C}$  be a section of  $\mathcal{C}(T)$ . We first construct a 2-tree decomposition of  $\Gamma(\mathcal{T})$ .

Pick some element  $t_1 \in \mathcal{C}$ . We now select a sequence of elements from  $\mathcal{C}$  starting with  $t_1$  as follows. Suppose that some sequence  $t_1, t_2, \dots, t_i \in \mathcal{C}$ ,  $i \geq 1$ , of elements in  $\mathcal{C}$  has been selected. To select  $t_{i+1}$  check if there exists some  $t \in \mathcal{C} - \{t_1, t_2, \dots, t_i\}$  and some  $1 \leq j \leq i$  with  $|t_j \cap t| = 2$ . If such a  $t$  exists, then put  $t_{i+1} = t$ , and repeat this process for the new sequence, otherwise stop. This process will clearly stop yielding a sequence  $t_1, t_2, \dots, t_k$  with  $1 \leq k \leq |\mathcal{C}|$ .

Now put  $(W_1 = \bigcup_{i=1}^k t_i, F_1 = \bigcup_{i=1}^k Co(\{t_i\}))$ . If  $\mathcal{C} = \{t_1, t_2, \dots, t_k\}$  then stop. Otherwise pick some element in  $\mathcal{C} - \{t_1, t_2, \dots, t_k\}$ , and repeat the above process for  $\mathcal{C} - \{t_1, t_2, \dots, t_k\}$ , to obtain a new pair  $(W_2, F_2)$ . Then, if necessary, keep repeating this whole process until all elements in  $\mathcal{C}$  have been selected. This results in a collection of pairs  $\{(W_1, F_1), \dots, (W_m, F_m)\}$ , with  $1 \leq m \leq |\mathcal{C}|$ .

Now, note that by construction  $H_i = (W_i, F_i)$  is a subgraph of  $\Gamma(\mathcal{T})$ ,  $|W_i| \geq 3$ , and  $H_i$  a 2-tree for all  $1 \leq i \leq m$ . Moreover, since  $\bigcup \mathcal{C} = X$  (see Proposition 3.6(c-i)), it follows that  $X = \bigcup_{i=1}^m W_i$ , and since  $\mathcal{T}$  is minimal,  $Co(\mathcal{C}) = \mathcal{T}$  (see Proposition 3.6(c-iii)) and so  $\mathcal{T} = \bigcup_{i=1}^m F_i$ .

We now observe that if  $e$  is any element of  $\mathcal{T}$ , then by construction of the pairs  $(W_i, F_i)$ , there must be some  $1 \leq l \leq m$  with  $e \in F_l$  and  $e \notin F_k$  for any  $k < l$ . Moreover, by construction if  $t \in \mathcal{C}$  with  $e \in Co(\{t\})$ , then  $Co(\{t\}) \subseteq F_l$ . In particular, again by construction, it follows that  $e \notin F_k$  for any  $k > l$ . Hence, the sets  $F_i$  are pairwise disjoint. Moreover, by construction,  $\mathcal{C} = \dot{\cup}_{i=1}^m \Delta(H_i)$ , and so  $|X| - 2 = |\mathcal{C}| = \sum_{i=1}^m |\Delta(H_i)|$  as  $\mathcal{C}$  is a section. Hence  $\{H_i\}_{i=1}^m$  is a 2-tree decomposition of  $\Gamma(\mathcal{T})$ , with  $\mathcal{C} = \dot{\cup}_{i=1}^m \Delta(H_i)$ .

To see that the uniqueness statement holds, suppose that  $\{H'_i = (W'_i, F'_i)\}_{i=1}^q$  is any 2-tree decomposition of  $\Gamma(\mathcal{T})$  with  $\mathcal{C} = \dot{\cup}_{i=1}^q \Delta(H'_i)$ . Suppose that  $v_1, \dots, v_b$  is an ordering of  $W'_1$ , which can be used to construct the 2-tree  $H'_1$ . Let  $\delta$  be the triangle  $v_1, v_2, v_3$  in  $H'_1$  (which must exist as  $|W'_1| \geq 3$ ). Then  $\delta$  must be contained in  $H_i$  for some  $1 \leq i \leq m$ , since  $\mathcal{C} = \dot{\cup}_{i=1}^m \Delta(H_i)$ .

Now, note that  $H'_1$  is a subgraph of  $H_i$ . Indeed, if not then there must exist some  $4 \leq l \leq b$  so that the triangle added to form a 2-tree on  $v_1, \dots, v_l$  which is subgraph of  $H'_1$  is not contained in  $H_i$ , but the 2-tree obtained from the sequence  $v_1, \dots, v_{l-1}$  is in  $H_i$ . So the triangle containing  $v_l$  which is added at stage  $l$  to  $H'_1$  must be contained in some  $H_j$ ,  $j \neq i$ . But this contradicts  $F_i \cap F_j = \emptyset$ .

Using similar reasoning, it follows that  $H_i$  is a subgraph of  $H'_1$  (since we can also use the triangle  $\delta$  as the first three elements in an ordering for constructing  $H_i$ ). Thus  $H'_1$  is equal to  $H_i$ .

Now, we can repeat this process for  $H'_2$ , considering now a triangle that provides the first three elements in an ordering for constructing  $H'_2$  that is in the set  $\mathcal{C} - \Delta(H'_1)$ , and keep repeating this whole process until finally come to considering a triangle in  $H'_q$  in the set  $\mathcal{C} - \dot{\cup}_{i=1}^{q-1} \Delta(H'_i)$ . In this way we see that  $q = m$  and  $H'_i$  is equal to some  $H_j$  for all  $1 \leq i \leq m$ . Hence the 2-tree decomposition  $\{H_i\}_{i=1}^m$  is the unique 2-tree decomposition of  $\Gamma(\mathcal{T})$  with  $\mathcal{C} = \dot{\cup}_{i=1}^m \Delta(H_i)$ .

To complete the proof of Part (a), note that if  $\mathcal{T}$  is sparse, then  $\mathcal{C} = \mathcal{C}(\mathcal{T})$  by Proposition 3.6 (c-ii). It follows by Theorem 3.4(i) and construction of  $\{H_i\}_{i=1}^m$  that there is no triangle in  $\Gamma(\mathcal{T})$  which is not contained in some  $H_i$ . Hence

$\{H_i\}_{i=1}^m$  is a strict 2-tree decomposition of  $\Gamma(\mathcal{T})$ .

From now on, we denote the 2-tree decomposition associated to a section  $\mathcal{C}$  given in Part (a) by  $\mathcal{H}_{\mathcal{C}}$ .

*Part (b):* The implication (ii)  $\Rightarrow$  (i) is obvious.

(iii)  $\Rightarrow$  (i) follows from Part (a) by taking the 2-tree decomposition  $\mathcal{H}_{\mathcal{C}(\mathcal{T})}$ .

(i)  $\Rightarrow$  (iii) Let  $\{H_i = (W_i, F_i)\}_{i=1}^m$  be a strict 2-tree decomposition for  $\Gamma(\mathcal{T})$ . Then  $|X| - 2 = \sum_{i=1}^m |\Delta(H_i)|$ . But every triangle in  $\Gamma(\mathcal{T})$  is contained in precisely one of the blocks  $H_i$  since  $\{H_i = (W_i, F_i)\}_{i=1}^m$  is strict. Hence by Theorem 3.4(i),

$$|\mathcal{C}(\mathcal{T})| = |\Delta(\Gamma(\mathcal{T}))| = \sum_{i=1}^m |\Delta(H_i)| = |X| - 2,$$

and so  $\mathcal{T}$  is sparse. Moreover,  $\mathcal{T}$  is minimal. Indeed, every edge  $xy$  of  $\Gamma(\mathcal{T})$  must be contained in some triangle of  $\Gamma(\mathcal{T})$  (by definition of a 2-tree decomposition). Hence, its removal would imply  $|\mathcal{C}(\mathcal{T} - xy)| = |\Delta(\Gamma(\mathcal{T} - xy))| < |\Delta(\Gamma(\mathcal{T}))| = |X| - 2$ , and so  $\mathcal{T} - xy$  would not be a triplet cover of  $T$ .

(i)  $\Rightarrow$  (ii) Suppose that  $\{H_i = (W_i, F_i)\}_{i=1}^m$  is any strict 2-tree decomposition for  $\Gamma(\mathcal{T})$ . Then  $\mathcal{C}(\mathcal{T}) = \Delta(\Gamma(\mathcal{T})) = \dot{\cup}_{i=1}^m \Delta(H_i)$ , and since  $\mathcal{T}$  is sparse  $\mathcal{C}(\mathcal{T})$  has a unique section (by Proposition 3.6(c-iv)), namely  $\mathcal{C}(\mathcal{T})$ . Statement (ii) now follows from Part (a) and the fact that  $\mathcal{T}$  is minimal.

*Part (c):* (i)  $\Rightarrow$  (ii) Suppose that  $\mathcal{T}$  is not sparse. Then  $\mathcal{C}(\mathcal{T})$  must have at least two distinct sections  $\mathcal{C}$  and  $\mathcal{C}'$ . Then  $\mathcal{H}_{\mathcal{C}} \neq \mathcal{H}_{\mathcal{C}'}$  are 2-tree decompositions of  $\Gamma(\mathcal{T})$ . Statement (i) now follows immediately.

(ii)  $\Rightarrow$  (i): This follows from Part (b). ■

Note that there exists a minimal triplet cover  $\mathcal{T}$  for some  $T$ , and a 2-tree decomposition of  $\Gamma(\mathcal{T})$  which is not of the form  $\mathcal{H}_{\mathcal{C}}$  as defined in the proof of the last theorem for any  $\mathcal{C}$  a section of  $\mathcal{C}(\mathcal{T})$ . Namely, take the tree and minimal

triplet cover in Fig. 3, add in a new leaf  $r$  to the edge adjacent to the cherry  $\{p, d\}$  to get  $T$ , and add the cords  $rc$  and  $rd$  to get  $\mathcal{T}$ . Then for the 2-tree  $H_1$  with triangles consisting of the vertex sets  $\{p, a, d\}$ ,  $\{a, b, d\}$ ,  $\{a, b, c\}$ ,  $\{b, c, q\}$ ,  $\{a, c, w\}$ ,  $\{d, b, z\}$  and the 2-tree  $H_2$  consisting of the triangle  $\{r, d, c\}$ ,  $\{H_1, H_2\}$  is a 2-tree decomposition, but cannot arise from a section since  $abc$  and  $abd$  are both in the support of the same vertex of  $T$ .

#### 4.1. Further observations

We now show that 2-tree decompositions have some attractive properties, which also allow us to obtain previous results on triplet covers in a rather natural way.

**Proposition 4.2.** *Suppose that  $H = (W, F)$  is a graph that has a 2-tree decomposition into  $m \geq 1$  blocks. Then  $|F| = 2|W| - 4 + m$ .*

*Proof:* Let  $\{H_i = (W_i, F_i)\}_{i=1}^m$  be a 2-tree decomposition of  $H$ . Then, since the number of triangles in a 2-tree equals the size of its vertex set minus 2

$$(1) \quad |W| - 2 = \sum_{i=1}^m |\Delta(H_i)| = \sum_{i=1}^m (|W_i| - 2).$$

Now, as  $\{H_i\}_{i=1}^m$  is a 2-tree decomposition,  $|F| = \dot{\cup}_{i=1}^m |F_i|$ . Hence, as the number of edges in a 2-tree is equal to twice the number of its vertices minus 3, we have

$$(2) \quad |F| = \sum_{i=1}^m (2|W_i| - 3).$$

Using Equations (1) and (2) it immediately follows that  $|F| = 2|W| - 4 + m$ . ■

Proposition 4.2 leads directly to the following result which shows that the size of the 2-tree decomposition  $\mathcal{H}_{\mathcal{C}}$  associated to a section  $\mathcal{C}$  is independent of the choice of  $\mathcal{C}$ .



**Corollary 4.3.** *Suppose that  $\mathcal{T}$  is a minimal triplet cover for  $T$ , for which  $\Gamma(\mathcal{T})$  has a 2-tree decomposition into  $m \geq 1$  blocks. Then  $|\mathcal{T}| = 2|X| - 4 + m$ . In particular, it follows that if  $\mathcal{C}$  is any section of  $\mathcal{C}(\mathcal{T})$ , then the 2-tree decomposition  $\mathcal{H}_{\mathcal{C}}$  of  $\Gamma(\mathcal{T})$  has  $|\mathcal{T}| - 2|X| + 4$  blocks.*

We pause to mention two consequences of Corollary 4.3. First, suppose that  $\mathcal{T}$  is a triplet cover for  $T$ . Let  $\mathcal{T}'$  be a minimal triplet cover for  $T$  contained in  $\mathcal{T}$ , and let  $\mathcal{C}$  be some section of  $\mathcal{C}(\mathcal{T}')$ . Let  $m = |\mathcal{H}_{\mathcal{C}}|$ . Then by Corollary 4.3,  $|\mathcal{T}| \geq |\mathcal{T}'| = 2|X| - 4 + m$ . Thus, since  $m \geq 1$ ,  $|\mathcal{T}| \geq 2|X| - 3$ . (This recovers [4, Proposition 1]).

Second, suppose that  $\mathcal{T}$  is a minimum triplet cover for  $T$ . Let  $\mathcal{C}$  be a section of  $\mathcal{C}(\mathcal{T})$ , and  $|\mathcal{H}_{\mathcal{C}}| = m$ . By Corollary 4.3, we then have  $2|X| - 3 = |\mathcal{T}| = 2|X| - 4 + m$ , and so  $m = 1$ . Therefore,  $\Gamma(\mathcal{T})$  is a 2-tree (this recovers [4, Theorem 1]). Moreover, as  $\Gamma(\mathcal{T})$  is a 2-tree,  $|\mathcal{C}(\mathcal{T})| = |X| - 2$ , and so  $\mathcal{T}$  is sparse.

**Remark**

The results above suggests the following natural question: If a graph  $H = (V, F)$  is 2-connected and  $H$  has a strict 2-tree decomposition in which every pair of blocks intersect in at most 2 vertices, then does there exist a minimal, sparse triplet cover  $\mathcal{T}$  for some phylogenetic  $X$ -tree  $T$ , with  $\Gamma(\mathcal{T})$  isomorphic to  $H$ ? Note that this can be shown to hold in case  $H$  is a 2-tree.

**5. Shellings and ample patchworks**

The concept of a shellable triplet cover was introduced in [3], and has proved helpful in subsequent papers. In order to define it, one requires the notion of a quartet tree. Suppose that  $X = \{a, b, c, d\}$  and that  $T$  is a phylogenetic tree for which the path joining  $a$  and  $b$  does not share a vertex with the path joining

$c$  and  $d$ . In that case we say that  $T$  is a *quartet tree* and denote it by writing  $ab|cd$ .

Given a triplet cover  $\mathcal{T}$  of a phylogenetic  $X$ -tree  $T$ , we say that  $\mathcal{T}$  is  $T$ -shellable if either  $|X| = 3$  or  $|X| \geq 4$  and there exists an ordering of the cords in  $\binom{X}{2} - \mathcal{T}$ , say  $a_1b_1, a_2b_2, \dots, a_mb_m$  such that for every  $1 \leq i \leq m$ , there exists a pair  $x_i, y_i$  of distinct elements in  $X - \{a_i, b_i\}$  such that the restriction of  $T$  to the set  $Y_i = \{a_i, b_i, x_i, y_i\}$  is the quartet tree  $x_ia_i|y_ib_i$ , and all cords in  $\binom{Y_i}{2}$  except  $a_ib_i$  are contained in  $\mathcal{T}_i = \mathcal{T} \cup \{a_jb_j : 1 \leq j \leq i-1\}$ .

For example, the triplet cover (indicated in terms of the median) in Fig. 1 is shellable since  $ae, bd, ad$  is a shelling for it.

Although this combinatorial definition of shellability seems somewhat involved, its motivation rests on it being a sufficient condition for recursively determining the distances between all pairs of leaves (when the edges of  $T$  are assigned arbitrary positive edge lengths) starting with just the distance values for the pairs in the triplet cover. In other words, if a triplet cover  $\mathcal{T}$  of  $T$  is shellable then the pairs of elements from  $X$  that are not already present in  $\mathcal{T}$  can be ordered in a sequence so that the distance in  $T$  between the leaves in each pair is uniquely determined from the distances values on pairs that are either (i) present as an element of  $\mathcal{T}$  or (ii) appear earlier in the sequence.

Note that a triplet cover of a tree  $T$  need not be  $T$ -shellable, even if  $\mathcal{T}$  is sparse. An example is shown in Fig. 5. Thus, it is of interest to better understand those triplet covers which are shellable.

The following lemma recalls some basic properties of shellability established in Proposition 4 of [4].

**Lemma 5.1.**

- (i) *Suppose that  $x \in X$  and  $\mathcal{T}$  is a triplet cover of  $T$  such that  $\mathcal{T}^{-x}$  is a triplet cover of  $T - x$ . If  $\mathcal{T}^{-x}$  is  $(T - x)$ -shellable, then  $\mathcal{T}$  is  $T$ -shellable.*
- (ii) *Suppose that  $\mathcal{T}, \mathcal{T}'$  are triplet covers of  $T$  and that  $\mathcal{T}' \subseteq \mathcal{T}$ . If  $\mathcal{T}'$  is  $T$ -shellable, then so is  $\mathcal{T}$ .*

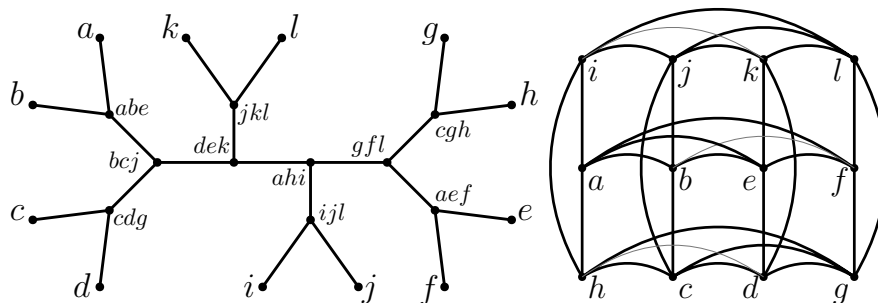


Figure 5: A non-shellable triplet cover (left). Its associated cover graph is described by the bold edges in the graph on the right. The three additional (lightly shaded) edges  $ik, bf, dh$  form the initial sequence of a shelling (based on the quartets  $ij|kl, ab|ef$ , and  $cd|gh$ , respectively), which does not extend to a full shelling.

Let  $M$  be a finite set. A *hierarchy* on  $M$  is a collection  $\mathcal{H}$  of non-empty subsets of  $M$  which satisfies the property that for all  $A, B \in \mathcal{H}$ ,  $A \cap B \in \{\emptyset, A, B\}$ . The hierarchy  $\mathcal{H}$  is *maximal* if there is no element  $H$  of  $2^M - \mathcal{H}$  such that  $\mathcal{H} \cup \{H\}$  is a hierarchy.

From [9], a collection  $\mathbf{P}$  of subsets of  $M$  forms a *patchwork* if it satisfies the following property

$$A, B \in \mathbf{P} \text{ and } A \cap B \neq \emptyset \Rightarrow A \cup B \in \mathbf{P}.$$

A patchwork  $\mathbf{P}$  on  $M$  is called *ample* if  $M$  is in  $\mathbf{P}$ , for all  $m \in M$ ,  $\{m\}$  is in  $\mathbf{P}$ , and  $\mathbf{P}$  contains a maximal hierarchy on  $M$ .

If  $\mathcal{T}$  is a triplet cover of  $T$  and  $\mathcal{C}$  is a section of  $\mathcal{C}(\mathcal{T})$ , we define  $\mathbf{P}(\mathcal{C})$  to be the collection of non-empty subsets  $\mathcal{C}'$  of  $\mathcal{C}$  that satisfy  $|\bigcup \mathcal{C}'| = |\mathcal{C}'| + 2$ . Note that  $\mathbf{P}(\mathcal{C})$  is a patchwork by Proposition 3.6(c-ii) and [6, Lemma 1.2].

Note that for the sparse minimal triplet cover  $\mathcal{T}$  of  $T$  in Fig. 4,  $\mathcal{T}$  is  $T$ -shellable, but  $\mathbf{P}(\mathcal{C}(\mathcal{T}))$  is not an ample patchwork.

**Proposition 5.2.** *Suppose that  $\mathcal{T}$  a triplet cover for  $T$  and  $\mathcal{C}$  is a section of  $\mathcal{C}(\mathcal{T})$ .*

- (i) *If  $\text{Co}(\mathcal{C})$  shellable, then so is  $\mathcal{T}$ .*

- (ii) If  $\mathcal{C}', \mathcal{C}'' \in \mathbf{P}(\mathcal{C})$  and  $\mathcal{C}' \cap \mathcal{C}'' = \emptyset$ , then  $|\bigcup \mathcal{C}' \cap \bigcup \mathcal{C}''| \leq 2$ . Moreover, if  $\mathcal{C}' \cup \mathcal{C}'' \in \mathbf{P}(\mathcal{C})$  also holds, then  $|\bigcup \mathcal{C}' \cap \bigcup \mathcal{C}''| = 2$ .

*Proof:* Part (i) follows by Lemma 5.1(ii). For Part (ii), let  $\mathcal{C}', \mathcal{C}'' \in \mathbf{P}(\mathcal{C})$  be such that  $\mathcal{C}' \cap \mathcal{C}'' = \emptyset$  and let  $Y = \bigcup \mathcal{C}'$  and  $Z = \bigcup \mathcal{C}''$ . Then since  $\mathcal{C}$  is a section of  $\mathcal{C}(\mathcal{T})$  and  $\mathcal{C}' \cup \mathcal{C}'' \subseteq \mathcal{C}$  we obtain

$$\begin{aligned}
|Y| + |Z| - |Y \cap Z| &= |Y \cup Z| \\
&= |\bigcup (\mathcal{C}' \cup \mathcal{C}'')| \\
&\geq |(\mathcal{C}' \cup \mathcal{C}'')| + 2 \\
&= |\mathcal{C}'| + |\mathcal{C}''| + 2 \\
&= (|Y| - 2) + (|Z| - 2) + 2 = |Y| + |Z| - 2
\end{aligned}$$

in view of Lemma 3.6(c-ii). Moreover, if  $\mathcal{C}' \cup \mathcal{C}'' \in \mathbf{P}(\mathcal{C})$ , then equality holds in the third line. ■

Suppose  $\mathcal{T}$  is a triplet cover for  $T$ . For  $A$  a non-empty subset of  $X$  of size at least three, we define  $T|_A$  to be the subtree of  $T$  spanned by the leaves in  $A$  (suppressing degree 2 vertices). Clearly,  $T|_A$  is a phylogenetic  $A$ -tree. We also define  $\mathcal{T}|_A$  to be the subset of  $\mathcal{T}$  consisting of those cords  $xy \in \mathcal{T}$  with  $x, y \in A$ .

**Lemma 5.3.** *Suppose  $\mathcal{T}$  is a triplet cover for  $T$  and that  $\mathcal{C}$  is a section of  $\mathcal{C}(\mathcal{T})$ .*

- (i) *If  $\mathcal{C}' \in \mathbf{P}(\mathcal{C})$  then, for  $Y = \bigcup \mathcal{C}'$ ,  $\mathcal{T}|_Y$  is a triplet cover for  $T|_Y$ .*
- (ii) *Suppose  $\mathcal{C}', \mathcal{C}'' \in \mathbf{P}(\mathcal{C})$ ,  $A = \bigcup \mathcal{C}'$ ,  $B = \bigcup \mathcal{C}''$ ,  $A \cup B = X$  and  $A \cap B = \{x, y\}$  for some  $x \neq y \in X$ . If  $a \in A - \{x, y\}$  and  $b \in B - \{x, y\}$ , then the quartet tree induced by  $T$  on  $\{a, b, x, y\}$  is either  $ax|yb$  or  $ay|bx$ .*

*Proof:* Part (i): Note that (considering  $T|_Y$  as a subtree of  $T$ )

$$(3) \quad \{v \in \mathring{V} : v = \text{med}_T(x, y, z) \text{ and } xyz \in \mathcal{C}'\} \subseteq \mathring{V}(T|_Y).$$

But  $|\mathring{V}(T|_Y)| = |Y| - 2$  (since  $T|_Y$  is a phylogenetic  $Y$ -tree), and, since  $\mathcal{C}$  is a section of  $\mathcal{C}(\mathcal{T})$  and  $\mathcal{C}' \in \mathbf{P}(\mathcal{C})$  we have:

$$|\{v \in \mathring{V} : v = \text{med}_T(x, y, z) \text{ and } xyz \in \mathcal{C}'\}| = |\mathcal{C}'| = |\bigcup \mathcal{C}'| - 2 = |Y| - 2.$$

Therefore, equality holds in (3), from which Lemma 5.3(i) immediately follows.

*Part (ii):* Suppose that  $\mathcal{C}', \mathcal{C}'' \in \mathbf{P}(\mathcal{C})$ ,  $A = \bigcup \mathcal{C}'$ ,  $B = \bigcup \mathcal{C}''$ ,  $A \cup B = X$  and  $A \cap B = \{x, y\}$  for some  $x \neq y \in X$ . If  $a \in A - \{x, y\}$  and  $b \in B - \{x, y\}$ , then we claim that the quartet tree induced by  $T$  on  $\{a, b, x, y\}$  is either  $ax|yb$  or  $ay|bx$ .

*Proof of Claim:* Let  $v \neq x, y$  be any vertex in  $T$  on the path between  $x$  and  $y$  in  $T$ . Let  $u \in V(T)$  be the vertex in  $T$  that is adjacent to  $v$  but not on the path between  $x$  and  $y$ . Consider the subtree  $T_u$  of  $T$  which is the component obtained by removing the edge  $\{u, v\}$  from  $T$  that contains  $u$ . We will show that either  $V(T_u) \cap X \subseteq A$  or  $V(T_u) \cap X \subseteq B$ . The statement then follows immediately.

Consider  $T_u$  as being a directed rooted tree with root  $u$  and all edges directed away from  $u$ . For  $w \in V(T_u)$ , let  $T_w$  denote the directed, rooted subtree of  $T_u$  with root  $w$ , and let  $Y_w = V(T_w) \cap X$ . We show by induction on  $|Y_w|$  that  $Y_w \subseteq A$  or  $Y_w \subseteq B$  for all  $w \in V(T_u)$ .

If  $|Y_w| = 1$ , then since  $A \cup B = X$ , clearly  $Y_w \subseteq A$  or  $Y_w \subseteq B$ . Suppose that  $|Y_w| > 1$ . Let  $w', w''$  be the children of  $w$  in  $T_w$ . Then, by induction,  $Y_{w'}$  is a subset of  $A$  or  $B$  and so is  $Y_{w''}$ . Suppose without loss of generality that  $Y_{w'}$  is a subset of  $A$ . We need to show that  $Y_{w''}$  is also a subset of  $A$ . Since  $\mathcal{T}$  is a triplet cover of  $T$ , there must exist a cord  $pq \in \mathcal{T}$  with  $p \in Y_{w'}$  and  $q \in Y_{w''}$ , and some  $r \in X - \{p, q\}$  with  $pqr \in S_w(\mathcal{T})$ . Hence,  $pqr \in B$  or  $pqr \in A$ . But since  $p \in A - \{x, y\}$  and  $A \cap B = \{x, y\}$  it follows that  $q \in A$ . Hence,  $Y_{w''} \subseteq A$ . ■

We can now state our third main theorem.

**Theorem 5.4.** *If  $\mathcal{T}$  is a triplet cover of  $T$  and there exists a section  $\mathcal{C}$  of  $\mathcal{C}(\mathcal{T})$  such that  $\mathbf{P}(\mathcal{C})$  is an ample patchwork, then  $\mathcal{T}$  is shellable.*

*Proof:* Since  $\mathbf{P}(\mathcal{C})$  is an ample patchwork, there exists a maximal hierarchy  $M$  in  $\mathbf{P}(\mathcal{C})$ . Suppose that  $\mathcal{C}' \in \mathbf{P}(\mathcal{C})$ ,  $\mathcal{C}' \neq \{t\}$ , some  $t \in \mathcal{C}$ , and  $\mathcal{C}_1, \mathcal{C}_2$  are the children of  $\mathcal{C}'$  in  $M$ . Let  $Y = \bigcup \mathcal{C}'$ , and  $Y_i = \bigcup \mathcal{C}_i$ ,  $i = 1, 2$ , so that in particular

$Y = Y_1 \cup Y_2$ . By Lemma 5.3(i),  $\mathcal{T}|_Y$  is a triplet cover for  $T|_Y$ , and  $\mathcal{T}|_{Y_i}$  is a triplet cover of  $T|_{Y_i}$ , for  $i = 1, 2$ . We now show that if  $\mathcal{T}|_{Y_i}$  is  $T|_{Y_i}$ -shellable, for  $i = 1, 2$ , then  $\mathcal{T}|_Y$  is  $T|_Y$ -shellable. This will complete the proof of Theorem 5.4 since induction can be used in a bottom-up fashion on  $M$  to see that  $\mathcal{T}$  is shellable.

Since  $\mathcal{T}|_{Y_i}$  is  $T|_{Y_i}$ -shellable, for  $i = 1, 2$ , it follows that for each  $i$  there must be some ordering of the cords in  $\mathcal{T}|_{Y_i}$  which satisfies the definition of  $T|_{Y_i}$ -shellability. Hence we can assume that we have added all cords in  $\binom{Y_i}{2}$ ,  $i = 1, 2$  into  $\mathcal{T}|_Y$  to obtain a new set of cords  $\mathcal{T}'$ .

Now, by Proposition 5.2 (applied with  $\mathcal{C}' = \mathcal{C}_1$  and  $\mathcal{C}'' = \mathcal{C}_2$ ),  $|Y_1 \cap Y_2| = 2$ . Let  $Y_1 \cap Y_2 = \{y, z\}$ ,  $y, z \in X$ . Note that  $yz \in \mathcal{T}'$  by our assumption on  $\mathcal{T}'$ . Now if  $p \in Y_1 - \{y, z\}$  and  $q \in Y_2 - \{y, z\}$  and  $pq \notin \mathcal{T}'$ , then  $\{yz, yp, yq, zp, zq\} \subseteq \mathcal{T}'$ , and so by Lemma 5.3(ii) applied to  $T|_Y$  and the sets  $Y_1, Y_2$ , we can add the cord  $pq$  into  $\mathcal{T}'$ . This can be repeated until we obtain all cords in  $\binom{Y}{2}$ , in such a way that it follows that  $\mathcal{T}|_Y$  is  $T|_Y$ -shellable.  $\blacksquare$

**Corollary 5.5.** *If  $\mathcal{T}$  is a sparse triplet cover of  $T$  such that  $\mathbf{P}(\mathcal{C}(\mathcal{T}))$  is ample, then  $\mathcal{T}$  is shellable.*

**Corollary 5.6.** *If  $\mathcal{T}$  is a triplet cover for  $T$  for which  $\Gamma(\mathcal{T})$  has a strict 2-tree decomposition into 2 or fewer blocks then  $\mathcal{T}$  is shellable.*

*Proof:* First note that, by Theorem 4.1(b) and Proposition 3.6(c-iv),  $\mathcal{C}(\mathcal{T})$  is the unique section of  $\Gamma(\mathcal{T})$ . Now suppose that  $\Gamma(\mathcal{T})$  has one block. Then  $\Gamma(\mathcal{T})$  is a 2-tree. Pick some ordering  $t_1, t_2, \dots, t_p$ ,  $p \geq 1$ , of the triangles in  $\Gamma(\mathcal{T})$  so that  $\Gamma(\mathcal{T})$  can be constructed by adding in one triangle at a time in the given ordering. Then it is straightforward to see that the set

$$\{\{t_1\}, \dots, \{t_p\}, \{t_1, t_2\}, \{t_1, t_2, t_3\}, \dots, \{t_1, t_2, \dots, t_p\}\}$$

is a maximal hierarchy in  $\mathbf{P}(\mathcal{C}(\mathcal{T}))$ . Now apply Theorem 5.4.

If the 2-tree decomposition of  $\Gamma(\mathcal{T})$  has two blocks, let  $t_1, t_2, \dots, t_p$  and  $t'_1, t'_2, \dots, t'_q$  be orderings of the triangles in each of the blocks so that each

block can be constructed by adding in one triangle at a time in the given ordering. Put  $\mathcal{C}' = \{\{t_1\}, \dots, \{t_p\}, \{t_1, t_2\}, \{t_1, t_2, t_3\}, \dots, \{t_1, t_2, \dots, t_p\}\}$ ,  $p \geq 1$ , and  $\mathcal{C}'' = \{\{t'_1\}, \dots, \{t'_q\}, \{t'_1, t'_2\}, \{t'_1, t'_2, t'_3\}, \dots, \{t'_1, t'_2, \dots, t'_q\}\}$ ,  $q \geq 1$ . Note that  $\{t_1, t_2, \dots, t_p\} \cup \{t'_1, t'_2, \dots, t'_q\} \in \mathbf{P}(\mathcal{C}(\mathcal{T}))$  by Theorem 3.4(ii) and Proposition 5.2(ii). Now it is straightforward to see that the set

$$\mathcal{C}' \cup \mathcal{C}'' \cup \{\{t_1, t_2, \dots, t_p, t'_1, t'_2, \dots, t'_q\}\}$$

is a maximal hierarchy in  $\mathbf{P}(\mathcal{C}(\mathcal{T}))$ . Now apply Theorem 5.4 again. ■

Note that the case for one block in the last corollary was also shown to hold in [4].

## 6. Acknowledgements

KTH and VM thank the London Mathematical Society and the Biomathematics Research Center, University of Canterbury, for their support. MS thanks the (former) Allan Wilson Centre for funding support.

## References

## References

- [1] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.* 4 (1987) 406–425.
- [2] A. Guénoche, B. Leclerc, V. Makarenkov, On the extension of a partial metric to a tree metric, *Discrete Mathematics* 276 (2004) 229–248.
- [3] A. Dress, K. T. Huber, M. Steel, ‘lassoing’ a phylogenetic tree i: basic properties, shellings and covers, *J. Math. Biol.* 65 (2012) 77–105.
- [4] K. T. Huber, V. Moulton, M. Steel, Minimum triplet covers of binary phylogenetic  $x$ -trees, *J. Math. Biol.* (2017). Doi:10.1007/s00285-017-1117-6.

- [5] K. T. Huber, M. Steel, Reconstructing fully-resolved trees from triplet cover distances, *Electron. J. Combin.* 21 (2014) P2.15.
- [6] A. Dress, M. Steel, A hall-type theorem for triplet systems based on medians in trees, *App. Math. Lett.* 22 (2009) 1789–1793.
- [7] A. Dress, K. Huber, M. Steel, A matroid associated with a phylogenetic tree, *Discrete Mathematics and Theoretical Computer Science* 16 (2014) 41–56.
- [8] B. Leclerc, V. Makarenkov, On some relations between 2-trees and tree metrics, *Discrete Math.* 192 (1998) 223–249.
- [9] A. Dress, S. Böcker, Patchworks, *Adv. Math.* 157 (2001) 1–21.