# Reconstructing fully-resolved trees from triplet cover distances

Katharina T. Huber
School of Computing Sciences
University of East Anglia
Norwich, U.K.
k.huber@uea.ac.uk

Mike Steel
Department of Mathematics and Statistics
University of Canterbury
Christchurch, New Zealand
mike.steel@canterbury.ac.nz

Submitted: May 17, 2013; Accepted: Apr 8, 2014; Published: May 2, 2014
Mathematics Subject Classifications: 05C88, 05C89

**Abstract**

It is a classical result that any finite tree with positively weighted edges, and without vertices of degree 2, is uniquely determined by the weighted path distance between each pair of leaves. Moreover, it is possible for a (small) strict subset $\mathcal{L}$ of leaf pairs to suffice for reconstructing the tree and its edge weights, given just the distances between the leaf pairs in $\mathcal{L}$. It is known that any set $\mathcal{L}$ with this property for a tree in which all interior vertices have degree 3 must form a *cover* for $T$ – that is, for each interior vertex $v$ of $T$, $\mathcal{L}$ must contain a pair of leaves from each pair of the three components of $T - v$. Here we provide a partial converse of this result by showing that if a set $\mathcal{L}$ of leaf pairs forms a cover of a certain type for such a tree $T$ then $T$ and its edge weights can be uniquely determined from the distances between the pairs of leaves in $\mathcal{L}$. Moreover, there is a polynomial-time algorithm for achieving this reconstruction. The result establishes a special case of a recent question concerning 'triplet covers', and is relevant to a problem arising in evolutionary genomics.

**Keywords:** $X$-tree; tree metric; tree reconstruction; shellability; triplet cover.

## 1 Introduction

Any tree $T$ with positively weighted edges, induces a metric $d$ on the set of leaves by considering the weighted path distance in $T$ between each pair of leaves. Moreover, provided $T$ has no vertices of degree 2, and that we ignore the labeling of interior vertices, both $T$ and its edge weights are uniquely determined by the metric $d$. This uniqueness result has been known since the 1960s and fast algorithms exist for reconstructing both the tree and its edge weights from $d$ (for further background the interested reader may consult [1] and [10] and the references therein).

The uniqueness result and the algorithms are important in evolutionary biology for reconstructing an evolutionary tree of species from genetic data [6]. However in this setting one frequently may not have $d$-values available for all pairs of species, due to the patchy nature of genomic coverage [9].

This raises a fundamental mathematical question – for which subsets of pairs of leaves of a tree do we need to know the $d$-values in order to uniquely recover the tree and its edge weights? In general this appears a difficult question (indeed determining whether such a partial $d$-metric is realized by *any* tree is NP-hard [5]). However, some sufficient conditions (as well as some necessary conditions) for uniqueness to hold have been found, in [3, 8, 13], and more recently in [4], and [7]. In this paper we consider the uniqueness question for trees that are 'fully-resolved' (i.e. all the interior vertices have degree 3) as these trees are of particular importance in evolutionary biology, and because the uniqueness question is easier to study for this class of trees.

The structure of this paper is as follows. First we introduce some background terminology and concepts, and then we define the particular type of subsets of leaf pairs (called 'stable triplet covers') which we show suffice to uniquely determine a fully-resolved tree. Moreover, we show how this comes about by establishing two combinatorial properties of stable triplet covers - a 'shellability' property and a graph-theoretic property related to tree-width, which we show is quite different to shellability. We conclude by providing a proof that a polynomial-time algorithm will reconstruct a tree and its edge weights for any set of leaf pairs that contains a stable triplet cover (or more generally a shellable subset). Our result answers a special case of the question posed at the end of [4] of whether every 'triplet cover' of a fully-resolved tree determines the tree and its edge weights.

# 2 Preliminaries

We now introduce some precise definitions required to state and prove our main results. We mostly follow the notation and terminology of [10] and [4].

## 2.1 $X$-trees, edge-weightings and distances

For the rest of the paper, assume that $|X| \geqslant 3$. An $X$-tree $T = (V, E)$ is a graph theoretical tree whose leaf set is $X$ and which does not have any vertices of degree 2. We call an $X$-tree *fully-resolved* if every *interior vertex* of $T$, that is, every non-leaf vertex of $T$, has degree three. Moreover, we call two distinct leaves $x$ and $y$ of $T$ a *cherry* of $T$, denoted by $x, y$, if the parent of $x$ is simultaneously the parent of $y$. For any subset $Y \subseteq X$, we denote by $T|Y$ the restriction of $T$ to the leaf set $Y$; this is the $Y$-tree obtained from the minimal subtree of $T$ that connects $Y$ by suppressing any resulting degree two vertices that arise [10].

An example of a fully-resolved $X$-tree for $X = \{a, b, c, d, e, f, g\}$, and having two cherries, is shown in Fig. 1(i).

In case $|Y| = 4$, say $Y = \{a, b, c, d\}$, and the path from $a$ to $b$ does not share a vertex with the path from $c$ to $d$ in $T|Y$, we refer to $T|Y$ as a *quartet tree* and denote it by $ab||cd$. Note that by deleting any edge $e \in E$ from $T$ the leaf sets $A_e$ and $B_e := X - A_e$ of the resulting two trees induce a bipartition of $X$. We refer to such a bipartition as $X-$*split* and denote it by $A|B$ where $A := A_e$ and $B := B_e$ and $A_e$ and $B_e$ are as above. We say that two $X$-trees $T = (V, E)$ and $T' = (V', E')$ are *equivalent* if there exists a bijection $\phi : V \to V'$ that is the identity on $X$ and extends to a graph isomorphism from $T$ to $T'$.

Suppose for the following that $T = (V, E)$ is an $X$-tree. Then we call a map $w : E \to \mathbb{R}_{\geqslant 0}$ that assigns a *weight*, that is, a non-negative real number, to every edge of $T$ an *edge-weighting* for $T$. Note that this definition allows that some of the edges of $T$ might have weight zero. We denote an $X$-tree $T$ together with an edge-weighting $w$ by the pair $(T, w)$ and call an edge-weighting that assign non-zero weight to every edge of $T$ that is not incident with a leaf of $T$ *proper*. Note that for any edge-weighting $w$ of $T$, taking the sum of the weights of the edges on the shortest path from some $x \in X$ to some $y \in X$ induces a distance between $x$ and $y$ and thus a distance $d = d_{(T,w)}$ on $X$.

For example, in the tree in Fig. 1(i), if each edge has weight 1, then $d(a, b) = 2, d(c, e) = 4$, and $d(c, f) = 5$.

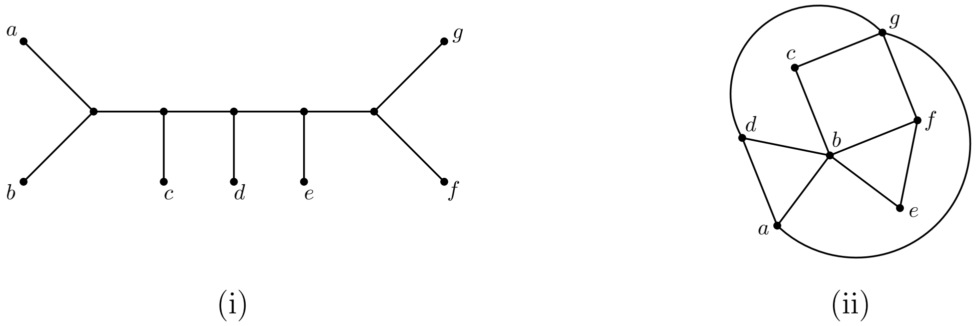

Figure 1: (i) A fully-resolved tree $X$-tree $T$ for $X = \{a, b, c, d, e, f, g\}$; (ii) the graph $(X, \mathcal{L})$ corresponding to a strong lasso $\mathcal{L}$ for $T$ (discussed further in Example 1).

## 2.2 Lassos

We call a subset of $X$ of size two a *cord* of $X$ and, for $a, b \in X$ distinct write $ab$ rather than $\{a, b\}$ for the cord containing $a$ and $b$. Also, for any non-empty set $\mathcal{L} \subseteq \binom{X}{2}$ of cords of $X$, we denote the edges of the graph $(X, \mathcal{L})$ whose vertex set is $X$ and whose edge set is the set $\{\{a, b\} : ab \in \mathcal{L}\}$ by $ab$ rather than $\{a, b\}$, $ab \in \mathcal{L}$.

Suppose for the following that $\mathcal{L} \subseteq \binom{X}{2}$ is a non-empty set of cords of $X$. If $T' = (V', E')$ is a further $X$-tree and $w$ and $w'$ are edge-weightings for $T$ and $T'$, respectively, such that $d_{(T,w)}(x, y) = d_{(T',w')}(x, y)$ holds for all $xy \in \mathcal{L}$ then we say that $(T, w)$ and $(T', w')$ are *$\mathcal{L}$-isometric*. Moreover we say that $\mathcal{L}$ is

(i) an *edge-weight lasso* for $T$ if for any two proper edge-weightings $w$ and $w'$ for $T$ such that $(T, w)$ and $(T, w')$ are $\mathcal{L}$-isometric we have that $w = w'$.

(ii) a *topological lasso* for $T$ if for any other $X$-tree $T'$ and any two proper edge-weightings $w$ and $w'$ for $T$ and $T'$, respectively, such that $(T, w)$ and $(T', w')$ are $\mathcal{L}$-isometric we have that $T$ and $T'$ are equivalent.

(iii) a *strong lasso* for $T$ if $\mathcal{L}$ is simultaneously an edge-weight and a topological lasso for $T$.

If $\mathcal{L}$ is a strong lasso for an $X$-tree then the graph $(X, \mathcal{L})$ must be connected and non-bipartite [4]. An example of a strong lasso $\mathcal{L}$ of the tree in Fig. 1(i) is the set of cords corresponding to the edges of the graph in Fig. 1(ii).

## 2.3 Shellability and 2d-trees

Suppose we have a subset $\mathcal{L}$ of $\binom{X}{2}$ with $X = \bigcup \mathcal{L}$ and an $X$-tree $T$. Then we say that $\binom{X}{2} \setminus \mathcal{L}$ is *$T$–shellable* if there exists an ordering of the cords in $\binom{X}{2} \setminus \mathcal{L}$ as, say, $a_1b_1, a_2b_2, \dots, a_mb_m$ such that, for every $\mu \in \{1, 2, \dots, m\}$, there exists a pair $x_\mu, y_\mu$ of 'pivots' for $a_\mu b_\mu$, i.e., two distinct elements $x_\mu, y_\mu \in X - \{a_\mu, b_\mu\}$, for which the tree $T|Y_\mu$ obtained from $T$ by restriction to $Y_\mu := \{a_\mu, b_\mu, x_\mu, y_\mu\}$, is the quartet tree $a_\mu x_\mu || y_\mu b_\mu$, and all cords in $\binom{Y_\mu}{2}$ except $a_\mu b_\mu$ are contained in $\mathcal{L}_\mu := \mathcal{L} \cup \big\{a_{\mu'} b_{\mu'} : \mu' \in \{1, 2, \dots, \mu - 1\}\big\}$. Any such ordering of $\binom{X}{2} \setminus \mathcal{L}$ will also be called a *shellable ordering* of $\binom{X}{2} \setminus \mathcal{L}$, and any subset $\mathcal{L}$ of $\binom{X}{2}$ for which a shellable ordering of $\binom{X}{2} \setminus \mathcal{L}$ exists will also be called an *shellable lasso for $T$*. In [4, Theorem 6], it was established that every shellable lasso for an $X$-tree is in particular a strong lasso for that tree.

A concept that is seemingly similar to shellability but, as we will see later on, quite distinct is that of a *2d-tree* where a graph $G = (V, E)$ is called a *2d-tree* if there exists an ordering $x_1, x_2, \dots, x_n$ of $V$ such that $\{x_1, x_2\} \in E$ and, for $i = 3, \dots, n$ the vertex $x_i$ has degree 2 in the subgraph of $G$ induced by $\{x_1, x_2, \dots, x_i\}$. 2d-trees are examples of $k$d-trees which were characterized in [11] and also studied in e. g. [7].

## 2.4 Example 1

Consider the seven-taxon tree, shown in Fig. 1(i), and the lasso $\mathcal{L} = \{ab, bd, ad, bc, bf, ag, dg, eb, ef, fg, gc\}$ (the edges of the graph in Fig. 1(ii)). The remaining ten chords in $\binom{X}{2} \setminus \mathcal{L}$ have a shellable ordering, described as follows:

$$bg, cd, ac, cf, ce, af, df, ae, eg, ed,$$

where the corresponding cord pivots are:

$$(a, d), (b, g), (b, d), (b, g), (b, f), (b, g), (b, g), (b, f), (a, f), (b, f),$$

and so $\mathcal{L}$ is a shellable (and hence strong) lasso for $T$. By considering the vertex ordering $a, b, d, g, c, f, e$, it is easy to check that the graph in Fig. 1(ii) is also a 2d-tree. □

### 2.5 Covers, triplet covers

A necessary condition for $\mathcal{L} \subseteq \binom{X}{2}$ to be an edge-weight lasso or a topological lasso for a fully-resolved $X$-tree is that $\mathcal{L}$ forms a *cover* for $T$ – that is, for each interior vertex $v$ of $T$, $\mathcal{L}$ contains a pair of leaves from each pair of the three components of $T - v$. However this condition is not sufficient for $\mathcal{L}$ to be either an edge-weight lasso or a topological lasso (examples are given in [4]).

A particular type of cover for a fully-resolved $X$-tree is a *triplet cover* which is defined as any subset $\mathcal{L}$ of $\binom{X}{2}$ with the property that for each interior vertex $v$ of $T$ we can select leaves $a, b, c$ from each of the three components of $T - v$ so that $ab, ac, bc \in \mathcal{L}$. It can be shown that if $\mathcal{L}$ is a triplet cover for a fully-resolved $X$-tree $T$ then $\mathcal{L}$ is an edge-weight lasso. However it is not known whether or not every triplet cover of every such $T$ is also a topological (and thereby a strong) lasso for $T$.

## 3 A special class of triplet covers

Suppose that $T = (V, E)$ is a fully-resolved $X$-tree, and let

$$\mathrm{clus}(T) := \bigcup_{e \in E} \{A_e, X - A_e\},$$

where $A_e|(X - A_e)$ denotes the $X-$split associated with edge $e \in E$. We call the elements in clus$(T)$ 'clusters' (in biology, they are also sometimes referred to as 'clans' [12]). Thus a cluster is a subset of $X$ that corresponds to the leaf labels on one side of some edge of $T$.

Given a collection $\mathcal{C}$ of non-empty subsets of $X$ we say that any function $f : \mathcal{C} \to X$ is a *stable transversal* for $\mathcal{C}$ if it satisfies the two properties:

- (transversality) $f(A) \in A$, for all $A \in \mathcal{C}$;
- (stability) $f(A) \in B \subseteq A \Longrightarrow f(A) = f(B)$ for all $A, B \in \mathcal{C}$.

Mostly we will be concerned with stable transversals for clus$(T)$, which were introduced in [2], though for a different purpose.

### 3.1 Example 2

An example of a stable transversal for clus$(T)$ is as follows: Consider *any* stable transversal $g$ for $2^X$ (equivalently, the function $g(A) = \min A$ under some total ordering of $X$), and consider *any* proper edge weighting $w$ of $T$. For a cluster $A \in \mathrm{clus}(T)$, consider the subset $A_w$ of leaves of $T$ in $A$ that are a closest to the edge $e$ whose deletion induces the split $A|(X \setminus A)$. Here 'closest' refers to the path distance in $T$ from each leaf in $A$ to $e$ under the edge weighting $w$. If we let $f(A) = g(A_w)$, for each $A \in \mathrm{clus}(T)$ then $f$ is a stable transversal for clus$(T)$. Notice that this holds also for the corresponding function in which 'closest' is replaced by 'furthest' throughout. □

### 3.2 Example 3

Consider the fully-resolved $X$-tree shown in Fig. 2(i), and the function $f$ defined as follows: $f(\{x\}) = x$ for all $x$ in $X$, and

$$f(\{a, a'\}) = a, f(\{b, b'\}) = b, f(\{c, c'\}) = c$$

and

$$f(X \setminus \{a, a'\}) = b, f(X - \{b, b'\}) = c, f(X \setminus \{c, c'\}) = a.$$

Then $f$ is a stable transversal for $T$. Note that the choices of $b, c, a$ in the last line could be replaced by, for example, $c, a, b$ or $c, c, a$ and we would still have a stable transveral. □

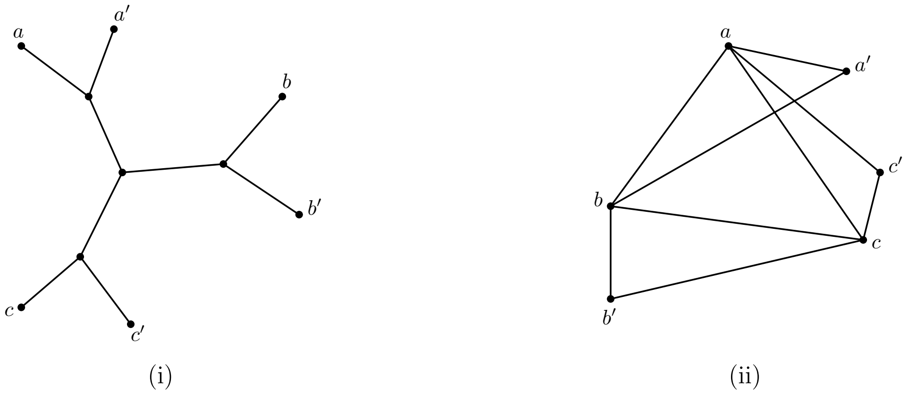

Figure 2: (i) A fully-resolved $X$-tree for the set $X = \{a, a', b, b', c, c'\}$; the graph $(X, \mathcal{L})$ where $\mathcal{L} = \mathcal{L}_{(T,f)}$ forms a stable triplet cover for $T$, and where $f$ is as defined in Example 3.

# 4 Stable triplet covers are minimal strong lassos for $T$

Given a fully-resolved $X$-tree $T$, a stable transversal $f$ of clus($T$) defines a triplet cover for $T$ as follows: For each interior vertex $v$ of $T$, consider the three components of the graph $T - v$, and let $A_v^1, A_v^2, A_v^3$ denote their leaf sets. Then let

$$\mathcal{L}_{(T,f)} := \bigcup_{v \in V_{\text{int}}} \{f(A_v^1)f(A_v^2), f(A_v^2)f(A_v^3), f(A_v^3)f(A_v^1)\}$$

where $V_{\text{int}}$ denotes the set of interior vertices of $T$. We say that $\mathcal{L}$ is a *stable triplet cover* (generated by $f$) if $\mathcal{L} = \mathcal{L}_{(T,f)}$ for some stable transversal $f$ of clus($T$). For example, for the pair $(T, f)$ described in Example 3, we have:

$$\mathcal{L}_{(T,f)} = \{ab, ac, bc, aa', a'b, bb', b'c, cc', c'a\},$$

and the graph $(X,\mathcal{L})$ for $\mathcal{L}=\mathcal{L}_{(T,f)}$ is shown in Fig. 2(ii). Notice that not all triplet covers are stable; indeed the set of triplet covers of a fully-resolved $X$-tree $T$ is precisely the set of subsets of $\binom{X}{2}$ of the form $\mathcal{L}_{(T,f)}$ where $f$ is required to satisfy only the transversality property above for some $f:\mathrm{clus}(T)\to X$.

Interestingly, Fig. 2(ii) shows that for the set $\mathcal{L}=\mathcal{L}_{(T,f)}$ with $T$ and $f$ from Example 3, the graph $(X,\mathcal{L})$ is a 2d-tree as $a,b,c,a',b',c'$ is an acceptable vertex ordering for the graph in that figure. Theorem 1 below establishes that both observations are not a coincidence.

## 4.1 Main result

We can now state our first main result which relates stable triplet covers with 2d-trees and shellable lassos.

**Theorem 1.** *If $\mathcal{L}$ is a stable triplet cover of a fully-resolved $X$-tree $T$ with $n:=|X|\geqslant 3$, then*

*(i) $(X,\mathcal{L})$ is a 2d-tree.*

*(ii) $\mathcal{L}$ is a shellable lasso for $T$, and so $\mathcal{L}$ is a strong lasso for $T$.*

*(iii) $|\mathcal{L}|=2n-3$, and so $\mathcal{L}$ is also a strong lasso for $T$ of minimal size.*

*Proof.* We prove parts (i)–(iii) simultaneously by induction on $n=|X|$. Shellability holds trivially for $n=3$ (since then $\binom{X}{2}\setminus\mathcal{L}=\emptyset$), so suppose that it holds when $n=k\geqslant 3$, and that $T$ is a fully-resolved tree with $k+1$ leaves, and that $\mathcal{L}$ is a triplet cover for $T$ generated by a stable transversal $f$ of clus$(T)$. Select any cherry $x,y$ of $T$. Without loss of generality, we may suppose that $f(\{x,y\})=x$. Let

$$z:=f(X\setminus\{x,y\}), X':=X-\{y\}, T':=T|X', \mathcal{L}':=\mathcal{L}\setminus\{xy,yz\},$$

and define $f':\mathrm{clus}(T)\to X$ by setting

$$f'(A)=\begin{cases} f(A), \text{ if } x\notin A;\\ f(A\cup\{y\}), \text{ if } x\in A.\end{cases}$$

Note that, since $f$ is a stable transversal for clus$(T)$, it follows that $y$ is not an element of any cord of $\mathcal{L}'$, and so $\mathcal{L}'\subseteq\binom{X'}{2}$. Moreover, $y\neq f'(A)$ for any $A\in\mathrm{clus}(T')$, and so $f':\mathrm{clus}(T')\to X'$. It can now be checked that $f'$ is a stable transversal for clus$(T')$ and so $\mathcal{L}'$ is a stable triplet cover of $T'$, generated by $f'$. By the inductive hypothesis (applied to $T'$ and $\mathcal{L}'$) it follows with regards to (i) that $(X',\mathcal{L}')$ is a 2d-tree. Clearly adding $y$ to the vertex set of that graph and $xy$ and $zy$ to its edge set preserves the 2d-tree property. By the definition of $\mathcal{L}'$ it is easy to see that the resulting graph is $(X,\mathcal{L})$.

Note that regarding (ii) and (iii) the induction hypothesis implies that $|\mathcal{L}'|=2k-3$, and so $|\mathcal{L}|=2(k+1)-3$ and that $\binom{X'}{2}\setminus\mathcal{L}'$ is shellable. So let us fix an ordering of $\binom{X'}{2}\setminus\mathcal{L}'$ that provides such a shelling. This will form the initial segment of a shellable ordering of $\binom{X}{2}\setminus\mathcal{L}$.

To describe this extended ordering, let $v$ be the interior vertex of $T$ adjacent to leaves $x$ and $y$, and let $u$ be the interior vertex of $T$ adjacent to $v$. Consider the three components of the graph $T - u$. One component contains $x, y$, and we will denote the leaf sets of the other two components by $X_2$ and $X_3$, where, without loss of generality, $z \in X_3$. Notice that $\binom{X}{2} \setminus \mathcal{L}$ is the disjoint union of the three sets:

$$\binom{X'}{2} \setminus \mathcal{L}', \{ty : t \in X_2\} \text{ and } \{ty : t \in X_3 \setminus \{z\}\}.$$

We order $\binom{X}{2} \setminus \mathcal{L}$ as follows: the elements of $\binom{X'}{2} \setminus \mathcal{L}'$ come first, ordered by their shellable ordering, followed by the elements $ty$ with $t \in X_2$ (in any order), followed by the elements $ty$ with $t \in X_3 \setminus \{z\}$ (in any order).

We claim that any such ordering provides a shellable ordering of $\binom{X}{2} \setminus \mathcal{L}$. To see this, observe first that, for any leaf $t \in X_2$, the elements $x, z$ provide 'pivots' for the pair $t, y$, since $T|\{x, y, z, t\} = xy||zt$ and all cords in $\binom{\{x,y,z,t\}}{2}$ except $ty$ are contained in $\mathcal{L} \cup (\binom{X'}{2} \setminus \mathcal{L}')$. Also, for any leaf $t \in X_3$, if we select any leaf $z' \in X_2$ then the pair $x, z'$ provides a 'pivot' for $t, y$, since $T|\{x, y, z', t\} = xy||z't$, and all cords in $\binom{\{x,y,z',t\}}{2}$ except $ty$ are contained in $\mathcal{L} \cup (\binom{X'}{2} \setminus \mathcal{L}') \cup \{t'y : t' \in X_2\}$. In all cases, the cords required for pivoting come earlier in the ordering.

Thus, we have established that $\mathcal{L}'$ is an shellable lasso for $T$, and so, by Theorem 6 of [4], $\mathcal{L}$ is also a strong lasso for $T$. Moreover, we showed that $|\mathcal{L}| = 2|X| - 3$, and since this equals the number of edges in any fully-resolved $X$–tree, linear algebra ensures that no smaller subset of $\mathcal{L}'$ could be an edge weight-lasso for $T$. Hence, $\mathcal{L}$ is a minimum size strong lasso for $T$, which completes the proof of the induction step, and thereby of the theorem.

□

## 4.2 Remarks

(1) Just because a graph $(X, \mathcal{L})$ is a 2d-tree, it does not follow that $\mathcal{L}$ forms a strong (let alone a shellable) lasso for every given fully-resolved $X$-tree $T$. A simple example is furnished by $X = \{a, b, c, d\}$ and $\mathcal{L} = \{ab, ac, bc, ad, bd\}$, for which $(X, \mathcal{L})$ is a 2d-tree, and yet $\mathcal{L}$ fails to be a strong lasso for $T = ab||cd$.

However, if $(X, \mathcal{L})$ forms a 2d-tree, or more generally if $\mathcal{L}$ contains a subset $\mathcal{L}'$ such that $(X, \mathcal{L}')$ is a 2d-tree, then $\mathcal{L}$ is a strong lasso for at least one fully-resolved $X$-tree. The proof is constructive based on the ordering $x_1, x_2, \ldots, x_n$ in the definition of a 2d-tree: Start with the tree consisting of leaves $x_1$ and $x_2$, and construct a fully-resolved tree as follows: for each $i > 2$, if $x_i$ is adjacent to $x_j$ and $x_k$ in $(X, \mathcal{L}')$ (where $j, k < i$) then let $x_i$ be the leaf that is attached by a new edge to a *new* subdivision vertex on the path connecting $x_j$ and $x_k$ in the tree so-far constructed.

It should be noted however that, in general, the concept of shellability and 2d-tree are quite distinct. For example consider the graph $(X, \mathcal{L})$ in Fig. 3(ii).

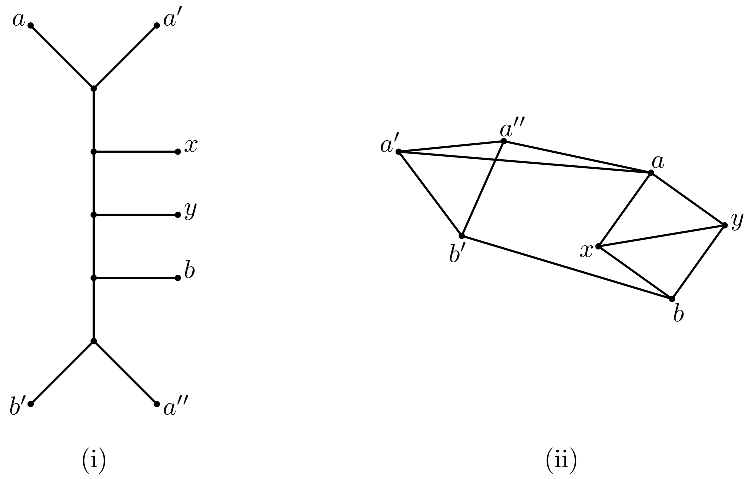

Figure 3: (i) A fully-resolved tree $X$-tree $T$ for $X = \{x, y, a, a', a'', b, b'\}$; (ii) the graph $(X, \mathcal{L})$.

Then it is easy to check that the remaining ten cords in $\binom{X}{2} \setminus \mathcal{L}$ have a shellable ordering for the tree in Fig. 3(i) given by:

$$ab, ab', b'x, b'y, xa', xa'', ya', ya'', ba', ba''$$

where the corresponding cord pivots are:

$$(x, y), (a', a''), (a, b), (a, b), (a, b'), (a, b'), (a, b'), (a, b'), (a, b'), (a, b').$$

But $(X, \mathcal{L})$ is not a 2d-tree as any such graph must necessarily contain a degree two vertex which is not the case for $(X, \mathcal{L})$. Moreover, there exists no subset $\mathcal{L}' \subsetneq \mathcal{L}$ such that $(X, \mathcal{L}')$ is a 2d-tree since any 2d-tree on seven leaves must have $2 \times 7 - 3 = |\mathcal{L}|$ edges.

(2) Suppose that $T$ is a fully-resolved $X$-tree, and $\mathcal{L} \subseteq \binom{X}{2}$ contains a stable triplet cover. A natural setting in which this situation arises is the following. Suppose $(T, w)$ is a properly edge-weighted fully-resolved $X$-tree, and $\mathcal{L} \subseteq \binom{X}{2}$ has the property that, for any interior vertex, $v$, $\mathcal{L}$ contains every chord $xy$ for which $x$ is a closest leaf to $v$ in one subtree of $T - v$ and $y$ is a closest leaf to $v$ in another subtree of $T - v$. Then, as noted in Example 2 above, $\mathcal{L}$ contains a stable triplet cover.

Now, when $\mathcal{L}$ contains a stable triplet cover for $T$, it follows by Theorem 1 that $\mathcal{L}$ is a shellable, and thereby also a strong lasso for $T$ (since any superset of a strong lasso for a tree is also a strong lasso for that tree). However, it is perhaps not clear how one might efficiently construct $(T, w)$ from the distances induced by $\mathcal{L}$, particularly when the subset of $\mathcal{L}$ corresponding to the stable triplet cover is not also given explicitly. Thus, in the next section we describe a polynomial-time algorithm for reconstructing $(T, w)$ whenever $\mathcal{L}$ contains some (unknown) shellable lasso for $T$.

## 5 An algorithm for reconstructing $(T, w)$ from $d_{(T,w)}|\mathcal{L}$ when $\mathcal{L}$ contains an shellable lasso for $T$.

Suppose that $\mathcal{L} \subseteq \binom{X}{2}$ and that $T$ is a fully-resolved $X$-tree, $w$ is a proper edge-weighting of $T$ and $d = d_{(T,w)}$. Starting with $\mathcal{L}^* = \mathcal{L}$ add cords to $\mathcal{L}^*$ and extend the domain of $d$ to those cords, by repeated application of the following extension rule $(\mathcal{R})$, described in [7] (Section 6.2, page 246):

$(\mathcal{R})$ Whenever $x, y, z, u \in X$ and

$$\binom{\{x, y, u, z\}}{2} - \{xz\} \subseteq \mathcal{L}^*, xz \notin \mathcal{L}^*, \text{ and}$$

$$d(x, y) + d(u, z) < d(x, u) + d(y, z)$$

add $xz$ to $\mathcal{L}^*$, and let $d(x, z) := d(x, u) + d(y, z) - d(y, u)$.

Let $\mathrm{cl}_{\mathcal{R}}(\mathcal{L})$ be the set of resulting set of cords obtained from the initial set $\mathcal{L}$ when this extension rule no longer yields any new cords.

Note that $\mathrm{cl}_{\mathcal{R}}(\mathcal{L})$ can be computed in polynomial time, and that $d-$values are assigned for all cords in $\mathrm{cl}_{\mathcal{R}}(\mathcal{L})$. Moreover, if $\mathrm{cl}_{\mathcal{R}}(\mathcal{L}) = \binom{X}{2}$, then $\mathrm{cl}_{\mathcal{R}}(\mathcal{L})$ is a strong lasso for $T$, however the converse does not hold (Example 6.2 of [4] provides a counterexample).

**Theorem 2.** *If $\mathcal{L} \subseteq \binom{X}{2}$ contains an shellable lasso for a fully-resolved $X$-tree $T$, and $d = d_{(T,w)}$, for some proper edge weighting $w$, then $\mathrm{cl}_{\mathcal{R}}(\mathcal{L}) = \binom{X}{2}$. Consequently, $T$ and $w$ can be reconstructed in polynomial time from the restriction of $d$ to $\mathcal{L}$.*

*Proof.* Suppose that $\mathcal{L}' \subseteq \mathcal{L}$ is a shellable lasso for $T$; we will show that $\mathrm{cl}_{\mathcal{R}}(\mathcal{L}') = \binom{X}{2}$ and so $\mathrm{cl}_{\mathcal{R}}(\mathcal{L}) = \binom{X}{2}$. Suppose to the contrary that $\mathrm{cl}_{\mathcal{R}}(\mathcal{L}')$ is a strict subset of $\binom{X}{2}$, and consider any shelling $a_1b_1, \ldots, a_mb_m$ of the cords in $\binom{X}{2} \setminus \mathcal{L}'$ (such a shelling exists by the assumption that $\mathcal{L}'$ is an shellable lasso for $T$). Let $j \in \{1, \ldots, m\}$ be the smallest index for which $a_jb_j \notin \mathrm{cl}_{\mathcal{R}}(\mathcal{L}')$. Then the condition on the shelling ensures that there exists pivots $x_j, y_j \in X - \{a_j, b_j\}$ so that for $Y = \{a_j, b_j, x_j, y_j\}$ we have $T|Y$ is the quartet tree $a_jx_j||b_jy_j$ and that each cord in $\binom{Y}{2} - \{a_jb_j\}$ either is an element of $\mathcal{L}'$ or it occurs earlier in the ordering for the shelling than $a_jb_j$, and so, by the minimality assumption concerning $j$, all these cords lie in $\mathrm{cl}_{\mathcal{R}}(\mathcal{L}')$. Consequently, $a_jb_j \in \mathrm{cl}_{\mathcal{R}}(\mathrm{cl}_{\mathcal{R}}(\mathcal{L}')) = \mathrm{cl}_{\mathcal{R}}(\mathcal{L}')$, a contradiction. Thus, our assumption that $\mathrm{cl}_{\mathcal{R}}(\mathcal{L}')$ is a strict subset of $\binom{X}{2}$ is not possible, as required.

Finally, to efficiently recover $(T, w)$, once $d$ has been defined on all of $\binom{X}{2}$, one can apply standard distance-based reconstruction methods for fully-resolved trees, such as the Neighbor-Joining method [6]. □

## Acknowledgements

M.S. thanks the Royal Society of NZ under its James Cook Fellowship scheme. Both authors thank the organizers of the "Structure Discovery in Biology: Motifs, Networks & Phylogenies", Schloss Dagstuhl, Germany, meeting where this paper was conceived. They also thank A.-A. Popescu for providing them with the example in Section 4.2(1), and an anonymous reviewer.

# References

[1] J.-P. Barthélemy and A. Guéoche, Trees and Proximity Representations. John Wiley and Sons, 1991.

[2] S. Böcker, A.W.M. Dress, and M. Steel, Patching up $X$–trees. Annals of Combinatorics 3 (1999) 1-12.

[3] S. Chaiken, A.K. Dewdney, and P.J. Slater, An optimal diagonal tree code. SIAM J. Alg. Disc. Math. 4(1) (1983) 42–49.

[4] A. Dress, K.T. Huber, and M. Steel, 'Lassoing' a tree I: Basic properties, shellings and covers, J. Math. Biol. 65(1) (2011) 77-105.

[5] M. Farach, S. Kannan, and T. Warnow, A robust model for finding optimal evolutionary trees. Algorithmica 13 (1995) 155–179.

[6] J. Felsenstein, Inferring phylogenies. Sinauer Associates, Sunderland, MA, 2004.

[7] A. Guénoche, B. Leclerc, and V. Markarenkov, On the extension a partial metric to a tree metric. Discr. Appl. Math. 276 (2004) 229–248.

[8] B. Leclerc, Minimum spanning trees for tree metrics: abridgements and adjustments. J. Classification 12 (1995) 207-241.

[9] M.J. Sanderson, M.M. McMahon, and M. Steel, Phylogenomics with incomplete taxon coverage: the limits to inference. BMC Evolutionary Biology 10 (2010) 155.

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

[11] P. Todd, A $k$-tree generalization that characterizes consistency of dimensioned engineering drawings. SIAM J. Discrete Math. 2(2) (1989) 255-261.

[12] M. Wilkinson, J.O. McInerney, R.P. Hirt, P.G. Foster, and T.M. Embley, Of clades and clans: terms for phylogenetic relationships in unrooted trees. Trends in Ecology and Evolution, 22 (2007) 114–115.

[13] S.V. Yushmanov, Representation of a tree with hanging vertices by elements of its distance matrix, Mat. Zametki, 35(6) (1984) 877–887.