Advances in Applied Mathematics 99 (2018) 59–82

Contents lists available at ScienceDirect

Advances in Applied Mathematics

www.elsevier.com/locate/yaama

# Combinatorial properties of triplet covers for binary trees

Stefan Grünewald [a], Katharina T. Huber [b,*], Vincent Moulton [b], Mike Steel [c]

[a] *CAS-MPG Partner Institute for Computational Biology, Chinese Academy of Sciences, Key Laboratory of Computational Biology, Shanghai, China*
[b] *School of Computing Sciences, University of East Anglia, Norwich, UK*
[c] *Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand*

A R T I C L E   I N F O

*Article history:*
Received 14 September 2017
Accepted 10 April 2018
Available online xxxx

*MSC:*
05C05
92D15

*Keywords:*
Phylogenetic tree
Triplet cover
Tree-distances
Hall's theorem
Ample patchwork
Shellability

A B S T R A C T

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'.



* Corresponding author.
*E-mail addresses:* stefan@picb.ac.cn (S. Grünewald), K.Huber@uea.ac.uk (K.T. Huber), V.Moulton@uea.ac.uk (V. Moulton), Mike.Steel@canterbury.ac.nz (M. Steel).

https://doi.org/10.1016/j.aam.2018.04.002
0196-8858/

## 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 [9].

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 [5]; more recent results appear in [2], 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 [6]. That paper also established that in this case the resulting triplet cover is 'shellable', complementing other recent work into phylogenetic 'lasso' sets [2,7], as well as a Hall-type characterization of the median function on trees in [4].

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 $x, y, z \in X$.