Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in DISCRETE MATHEMATICS

Electronic Notes in Discrete Mathematics 50 (2015) 337–342 www.elsevier.com/locate/endm

Compatibility, Incompatibility, Tree-Width, and Forbidden Phylogenetic Minors

Sudheer Vakati[1,3]

Amazon.com
Seattle, WA, USA

David Fernández Baca[2,4]

Department of Computer Science
Iowa State University
Ames, IA 50010, USA

Abstract

A collection $\mathcal{P}$ of phylogenetic trees is compatible if there is a tree that displays all the relationships among species exhibited by the trees in $\mathcal{P}$. We give a simple characterization of compatibility based on graph triangulation. We then study how to deal with incompatibility through edge contraction and tree deletion, and introduce the notion of a phylogenetic minor.

Keywords: Triangulations, Treewidth, Phylogenetics

[1] Supported in part by the National Science Foundation under grant CCF-1017189.

[2] Supported in part by the National Science Foundation under grants CCF-1017189 and CCF-1422134.

[3] Email: `svakati@amazon.com`

[4] Email: `fernande@iastate.edu`

http://dx.doi.org/10.1016/j.endm.2015.07.056

1 Introduction

A *phylogenetic tree* T is an unrooted tree with no degree two-vertices, whose leaves are in one-to-one correspondence with a label set $\mathcal{L}(T)$. The labels represent species —also known as *taxa*—, and T represents the evolutionary history of these species. The following problem arises when constructing phylogenetic trees for a collection of species. We are given a collection of input trees with partially overlapping leaf sets. The question is to determine if there is a tree that exhibits the information from each of them — i.e., whether or not the information present in the input trees is compatible. To explain this precisely, we need some definitions.

Given a set $Y \subseteq \mathcal{L}(T)$, let $T|Y$ denote the tree obtained by suppressing vertices of degree two in the minimal subtree of T connecting the leaves with labels in Y. Let S be another phylogenetic tree with label set $\mathcal{L}(S) \supseteq \mathcal{L}(T)$. Then, S *displays* T if T can be derived from $S|\mathcal{L}(T)$ by edge contraction.

Throughout the rest of the paper, $\mathcal{P} = \{T_1, T_2, \ldots, T_k\}$ denotes a collection of k phylogenetic trees; we refer to $\mathcal{P}$ as a *profile*. We write $\mathcal{L}(\mathcal{P})$ to denote $\bigcup_{i\in[k]} \mathcal{L}(T_i)$. A *supertree* for $\mathcal{P}$ is a phylogenetic tree S with label set $\mathcal{L}(S) = \mathcal{L}(\mathcal{P})$. $\mathcal{P}$ is *compatible* if there exists a supertree of $\mathcal{P}$ that displays every tree in $\mathcal{P}$. Deciding whether a profile is compatible is called the *tree compatibility problem*. While this problem is NP-Complete [5], it is fixed-parameter tractable in the number of trees [3].

Display graphs, modified display graphs, and tree-width

The *display graph* of $\mathcal{P}$ is the graph $G(\mathcal{P})$ obtained from the disjoint union of the trees in $\mathcal{P}$ by identifying the leaves with the same labels [3]. The *modified display graph* of $\mathcal{P}$, denoted $G_{\mathsf{M}}(\mathcal{P})$, is the graph obtained from $G(\mathcal{P})$ by doing the following for each leaf v in $G(\mathcal{P})$: (i) insert edges to make the neighbors of v a clique and (ii) delete v. We refer to the edges inserted in step (i) as *added edges*; every other edge of $G_{\mathsf{M}}(\mathcal{P})$ is a *tree edge*. See Fig. 1. Unlike $G(\mathcal{P})$, none of the vertices of $G_{\mathsf{M}}(\mathcal{P})$ is labeled. Thus, two profiles with different display graphs and different label sets may have isomorphic modified display graphs, where the isomorphism distinguishes tree edges from added edges.

A *tree decomposition* of a graph G is a pair (T, B) where T is a tree and B is a function from $V(T)$ to subsets of $V(G)$ such that: (i) for every $v \in V(G)$, there is an $x \in V(T)$ where $v \in B(x)$, (ii) for every edge $uv \in E(G)$, there is an $x \in V(T)$ where $\{u, v\} \subseteq B(x)$, and (iii) for any $v \in V(G)$, the set $\{x : v \in B(x)\}$ induces a subtree in T. The *width* of a tree decomposition (T, B) of G is the maximum value of $|B(x)| - 1$ over all vertices $x \in V(T)$. The