



Treewidth distance on phylogenetic trees

Steven Kelk^a, Georgios Stamoulis^{a,*}, Taoyang Wu^b

^a Department of Data Science and Knowledge Engineering (DKE), Maastricht University, Maastricht, the Netherlands

^b School of Computing Sciences, University of East Anglia, Norwich, United Kingdom



ARTICLE INFO

Article history:

Received 26 June 2017

Received in revised form 5 January 2018

Accepted 5 April 2018

Available online 18 April 2018

Communicated by F.V. Fomin

Keywords:

Graph theory

Phylogenetics

Treewidth

Algorithmic graph theory

Computational biology

ABSTRACT

In this article we study the treewidth of the *display graph*, an auxiliary graph structure obtained from the fusion of phylogenetic (i.e., evolutionary) trees at their leaves. Earlier work has shown that the treewidth of the display graph is bounded if the trees are in some formal sense topologically similar. Here we further expand upon this relationship. We analyze a number of reduction rules, commonly used in the phylogenetics literature to obtain fixed parameter tractable algorithms. In some cases (the *subtree reduction*) the reduction rules behave similarly with respect to treewidth, while others (the *cluster reduction*) behave very differently, and the behavior of the *chain reduction* is particularly intriguing because of its link with graph separators and forbidden minors. We also show that the gap between treewidth and Tree Bisection and Reconnect (TBR) distance can be infinitely large, and that unlike, for example, planar graphs the treewidth of the display graph can be as much as linear in its number of vertices. A number of other auxiliary results are given. We conclude with a discussion and list a number of open problems.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Phylogenetic trees are used extensively within computational biology to model the history of a set of species (known as taxa) X ; the internal nodes represent evolutionary diversification events such as speciation [39]. Within the field of phylogenetics there has long been interest in quantifying the topological dissimilarity of phylogenetic trees and understanding whether this dissimilarity is biologically significant. This has led to the development of many *incongruency measures* such as Subtree Prune and Regraft (SPR) distance and Tree Bisection and Reconnect (TBR) distance [1]. Most of these measures are NP-hard to compute and this is indeed true for SPR, TBR distances. More recently such measures have also attracted attention because of their importance in methods which merge dissimilar trees into *phylogenetic networks*; phylogenetic networks are simply the generalization of trees to graphs [31].

Parallel to such developments there has been growing interest in the role of the graph-theoretic parameter *treewidth* within phylogenetics. Treewidth is an intensely studied parameter in algorithmic graph theory and it indicates, at least in an algorithmic sense, how far an undirected graph is from being a tree (see e.g. [7,11,12] for background). The enormous focus on treewidth is closely linked to the fact that a great many NP-hard optimization problems become (fixed parameter) tractable on graphs of bounded treewidth [18]. A seminal paper by Bryant and Lagergren [16] linked phylogenetics to treewidth by demonstrating that, if a set of trees (not necessarily all on the same set of taxa X) can simultaneously be topologically embedded within a single “supertree”—a property known as *compatibility*—then an auxiliary graph known as the *display graph* has bounded treewidth. Since this paper a small but growing number of papers at the interface of

* Corresponding author.

E-mail addresses: steven.kelk@maastrichtuniversity.nl (S. Kelk), georgios.stamoulis@maastrichtuniversity.nl (G. Stamoulis), taoyang.wu@uea.ac.uk (T. Wu).

graph theory and phylogenetics have explored this relationship further. Much of this literature focuses on the link between compatibility and (restricted) triangulations of the display graph (e.g. [41,29,24,42]), but more recently the algorithmic dimension has also been tentatively explored [5,27,33]. In the spirit of the original Bryant and Lagergren paper, which used heavy meta-theoretic machinery to derive a theoretically efficient algorithm for the compatibility problem, Kelk et al. [34] showed that the treewidth of the display graph of two trees is bounded as a linear function of the TBR distance (equivalently, the size of a Maximum Agreement Forest—MAF [1]) between the two trees, and then used this insight to derive theoretically efficient algorithms for computation of many different incongruency measures. In that article it was empirically observed that in practice the treewidth of the display graph is often much smaller than the TBR distance (and thus also the many incongruency measures for which TBR is a lower bound). This raises two natural questions. First, in how far can this apparently low treewidth be exploited to yield genuinely practical dynamic programming algorithms running over low-width tree decompositions? There has been some progress in this direction in the compatibility literature (notably, [5]) but there is still much work to be done. Second, how *exactly* does the treewidth of the display graph behave, both in the sense of extremal results (e.g. how large can the treewidth of a display graph get?) and in the sense of understanding when and why the treewidth differs significantly from measures such as TBR.

Here we focus primarily on the second question. We begin with a more structural perspective. We show that, given an arbitrary (multi)graph G on n vertices with maximum degree k , one can construct two unrooted binary trees $T_1(G)$ and $T_2(G)$ such that their display graph $D = D(T_1(G), T_2(G))$ has at most $O(nk)$ vertices and edges and G is a minor of D . We combine this with the known fact that cubic expanders (a special family of 3-regular graphs) on n vertices have treewidth $\Omega(n)$ to yield the result that display graphs on n vertices can also (in the worst case) have treewidth linear in n . This contrasts, for example, with planar graphs on n vertices which have treewidth at most $O(\sqrt{n})$ [20]. We also show how a more specialized construction can be used to embed arbitrary grid minors [17] into display graphs with a much smaller inflation in the number of vertices and edges.

We then continue by analyzing how reduction rules often used in the computation of incongruency measures impact upon the treewidth of the display graph. Not entirely surprisingly the *common pendant subtree* reduction rule [1] is shown to preserve treewidth. The *cluster* reduction [4,36,14], however, behaves very differently for treewidth than for many other incongruency measures. Informally speaking, if both trees can be split by deletion of an edge into two subtrees on X' and X'' , many incongruency measures combine additively around this *common split*, while treewidth behaves (up to additive terms) like the maximum function. We use this later in the article to explicitly construct a family of tree pairs such that the treewidth of their display graph is 3, but the TBR distance of the trees (and their MP distance—a measure based on the phylogenetic principle of parsimony [25,37,33]) grows to infinity. The third reduction rule we consider is the *chain rule*, which collapses common caterpillar-like regions of the trees into shorter structures. For incongruency measures it is often the case that truncation of such chains to $O(1)$ length preserves the measure [1,15,45], although sometimes the weaker result of truncation to length $f(k)$ [44,43] (for some function that depends only on the incongruency parameter k) is the best known. We show that truncation of common chains to length $f(tw)$, where tw is the treewidth of the display graph, indeed preserves treewidth; this uses asymptotic results on the number of vertices and edges in forbidden minors for treewidth. Proving that truncation to $O(1)$ -length preserves treewidth remains elusive; we prove the intermediate result that truncation to length 2 can cause the treewidth to decrease by at most 1. The case when the chain is not a separator of the display graph seems to be a particularly challenging bottleneck in removing the “−1” term from this result. Although intuitively reasonable, it remains unclear whether truncation to length $O(1)$ is treewidth-preserving, for some universal constant.

In the last two mathematical sections of the paper we prove that, if two trees have TBR- or MP-distance 1, then the treewidth of their display graph is 3. However, the converse certainly does not hold: we construct the aforementioned “infinite gap” examples where the display graph has treewidth 3 but both TBR distance and MP-distance spiral off to infinity.

Finally, we reflect on the wider context of these results and discuss a number of open problems.

In conclusion, we observe that for (algorithmic) graph theorists the interface between treewidth and phylogenetics continues to yield many new questions which will likely require a new “phylo-algorithmic” graph theory to be answered. For phylogeneticists the appeal remains structural-algorithmic: can we convert the apparently low treewidth of display graphs into competitive, or even superior, algorithms for computation of incongruency measures?

2. Preliminaries

An *unrooted binary phylogenetic tree* T on a set of leaf labels (known as *taxa*) X is an undirected tree where all internal vertices have degree three and the leaves are bijectively labeled by X . If we (exceptionally) allow some internal vertices of T to have degree two, then we call these vertices *roots* (abusing slightly the usual root meaning). When it is understood from the context we will often drop the prefix “unrooted binary phylogenetic” for brevity.

Let $Y \subseteq X$. Then, for a tree T on X we denote by $T|Y$ the tree which is obtained by forming a minimal subgraph T' of T that spans all leaves labeled by Y , and suppressing any vertices of degree 2.

In this manuscript the *display graph* of two binary phylogenetic trees plays a central role (Fig. 1):

Download English Version:

<https://daneshyari.com/en/article/6875492>

Download Persian Version:

<https://daneshyari.com/article/6875492>

[Daneshyari.com](https://daneshyari.com)