Journal of Computer and System Sciences 117 (2021) 165–181

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Maximum parsimony distance on phylogenetic trees: A linear kernel and constant factor approximation algorithm

Mark Jones [a,b,*], Steven Kelk [c], Leen Stougie [b,d,e]

[a] Delft Institute of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE, Delft, the Netherlands
[b] Centrum Wiskunde & Informatica (CWI), 1098 XG Amsterdam, the Netherlands
[c] Department of Data Science and Knowledge Engineering (DKE), Maastricht University, 6200 MD Maastricht, the Netherlands
[d] Vrije Universiteit Amsterdam, 1081 HV Amsterdam, the Netherlands
[e] INRIA-Erable, France

ARTICLE INFO

Article history:
Received 7 April 2020
Received in revised form 23 October 2020
Accepted 26 October 2020
Available online 7 December 2020

Keywords:
Phylogenetics
Maximum parsimony
Fixed parameter tractability
Maximum agreement forest

ABSTRACT

Maximum parsimony distance is a measure used to quantify the dissimilarity of two unrooted phylogenetic trees. It is NP-hard to compute, and very few positive algorithmic results are known due to its complex combinatorial structure. Here we address this shortcoming by showing that the problem is fixed parameter tractable. We do this by establishing a linear kernel i.e., that after applying certain reduction rules the resulting instance has size that is bounded by a linear function of the distance. As powerful corollaries to this result we prove that the problem permits a polynomial-time constant-factor approximation algorithm; that the treewidth of a natural auxiliary graph structure encountered in phylogenetics is bounded by a function of the distance; and that the distance is within a constant factor of the size of a maximum agreement forest of the two trees, a well studied object in phylogenetics.

1. Introduction

Phylogenetics is the science of inferring and comparing trees (or more generally, graphs) that represent the evolutionary history of a set of species [34]. In this article we focus on trees. The inference problem has been comprehensively studied: given only data about the species in X (such as DNA data) construct a *phylogenetic tree* which optimizes a particular objective function [17,40]. Informally, a phylogenetic tree is simply a tree whose leaves are bijectively labelled by X. Due to different objective functions, multiple optima and the phenomenon that certain genomes are the result of several evolutionary paths (rather than just one) we are often confronted with multiple "good" phylogenetic trees [32]. In such cases we wish to formally quantify how dissimilar these trees really are. This leads naturally to the problem of defining and computing the *distance* between phylogenetic trees [36]. Many such distances have been proposed, some of which can be computed in polynomial-time, such as *Robinson-Foulds* (RF) distance [33], and some of which are NP-hard, such as *Subtree Prune and Regraft* (SPR) distance [9] or *Tree Bisection and Reconnection* (TBR) distance [1].

Interestingly, distances are not only relevant as a numerical quantification of difference: they also appear in constructive methods for the inference of phylogenetic networks [20], which generalise trees to graphs, and phylogenetic supertrees,

* Corresponding author at: Delft Institute of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE, Delft, the Netherlands.
E-mail address: M.E.L.Jones@tudelft.nl (M. Jones).

https://doi.org/10.1016/j.jcss.2020.10.003

which seek to merge multiple trees into a single summary tree [42]. In recent decades NP-hard phylogenetic distances have attracted quite some attention from the discrete optimization and parameterized complexity communities, see e.g. [12,16].

In this article we focus on a relatively new distance measure, *maximum parsimony distance*, henceforth denoted d_{MP}. Let T_1 and T_2 be two unrooted (i.e. undirected) binary phylogenetic trees, with the same set of leaf labels X. Consider an arbitrary assignment of colours ("states") to X; we call such an assignment a *character*. The *parsimony score* of T_1 with respect to the character is the minimum number of bichromatic edges in T_1, ranging over all possible colourings of the internal vertices of T_1. The parsimony distance of T_1 and T_2 is the maximum absolute difference between parsimony scores of T_1 and T_2, ranging over all characters [18,31].

The distance has several attractive properties; it is a metric, and (unlike e.g. RF distance) it is not confounded by the influence of horizontal evolutionary events [18]. Furthermore, the concept of parsimony, which lies at the heart of d_{MP}, is fundamental in phylogenetics since it articulates the idea that explanations of evolutionary history should be no more complex than necessary. Alongside its historical significance for applied phylogenetics [17], the study of character-based parsimony has given rise to many beautiful combinatorial and algorithmic results; we refer to e.g. [37,29,38,2,30] for overviews.

Unfortunately, it is NP-hard to compute d_{MP} [22]. A simple exponential-time algorithm is known [26], which runs in time $O(\phi^n \cdot \text{poly}(n))$, where $|X| = n$ and $\phi \approx 1.618$ is the golden ratio, but beyond this few positive results are known. This is frustrating and surprising, since a number of results link d_{MP} to the well-studied TBR distance, henceforth denoted d_{TBR}. Namely, it has been proven that d_{MP} is a lower bound on d_{TBR} [18], which, informally, asks for the minimum number of topological rearrangement operations to transform one tree into the other; an empirical study has suggested that in practice the distances are often very close [23]. Also, d_{MP} has been used to prove the tightness of the best-known kernelization results for d_{TBR} [24,25]. What, exactly, is the relationship between d_{MP} and d_{TBR}? This is a pertinent question, which transcends the specifics of TBR distance because, crucially, d_{TBR} can be characterized using the powerful *maximum agreement forest* abstraction.

Distances based on agreement forests have been intensively and successfully studied in recent years, as the use of the agreement forest abstraction almost always yields fixed parameter tractability and constant-factor approximation algorithms [10], many of which are effective in practice. We refer to [41,39,14,35] for recent overviews of the agreement forest literature, and books such as [15] for an introduction to fixed parameter tractability. In particular, d_{TBR} can be computed in $O(3^{d_{TBR}} \cdot \text{poly}(n))$ time [13], permits a polynomial-time 3-approximation algorithm, and a kernel of size $11d_{TBR} - 9$ [25].

In contrast, prior to this paper very little was known about d_{MP}: nothing was known about the approximability of d_{MP}; it was not known whether it is fixed parameter tractable (where d_{MP} is the parameter); and, while, as mentioned above, it is known that $d_{MP} \leq d_{TBR}$, it remained unclear how much smaller d_{MP} can be than d_{TBR} in the worst case. Despite promising partial results it even remained unclear whether questions such as "Is $d_{MP} \geq k$?" can be solved in *polynomial* time when k is a constant [8,23]. This is another important difference with distances such as d_{TBR}, where corresponding questions are trivially polynomial time solvable for fixed k. The apparent extra complexity of d_{MP} seems to stem from the unusual max-min definition of the problem, and the fact that unlike d_{TBR}, which is based on topological rearrangements of subtrees, d_{MP} is based only on characters.

In this article we take a significant step forward in understanding the deeper complexity of d_{MP} and resolve all of the above questions. Our central result is that we prove that two common polynomial-time reduction rules encountered in phylogenetics, the *subtree* and *chain* reductions [1], are sufficient to produce a *linear kernel* for d_{MP}. This means that, after exhaustive application of these rules, which preserve d_{MP}, the reduced trees will have at most $\alpha \cdot (d_{MP} + 1)$ leaves, with $\alpha = 560$. The fixed parameter tractability of computing d_{MP} (parameterized by itself) then follows, by solving the kernel using the exact algorithm from [26]. The fact that the reduction rules preserve d_{MP} was already known [23]. However, proving the bound on the size of the reduced trees requires rather involved combinatorial arguments, which have a very different flavour to the arguments typically encountered in the maximum agreement forest literature. The main goal of this article is to present these arguments as clearly as possible, rather than to optimize the resulting constants.

The kernel confirms that questions such as "Is $d_{MP} \geq k$?" can, indeed, be solved in polynomial time: it is striking that here the proof of fixed parameter tractability has preceded the weaker result of polynomial-time solveability for fixed k.

Next, by producing a modified, constructive version of the bounding argument underpinning the kernelization, we are able to demonstrate a polynomial-time $\alpha(1 + 1/r)$-factor approximation algorithm for computation of d_{MP} for any constant r, placing the problem in APX.

A number of other powerful corollaries result from the kernelization. We leverage the fact that the reduction rules also preserve d_{TBR}, to show that $1 \leq \frac{d_{TBR}}{d_{MP}} \leq 2\alpha$, which limits how much smaller d_{MP} can be than d_{TBR}. Subsequently, we show that the treewidth of an auxiliary graph structure known as the *display graph* [11] is bounded by a linear function of d_{MP}, resolving an open question posed several times [28,23]. The treewidth bound, and the existence of a non-trivial approximation algorithm for d_{MP}, were specified as sufficient conditions for proving the fixed parameter tractability of d_{MP} via Courcelle's Theorem [23]; our linear kernel implies *them*. Summarising, our central result shows how kernelization can open the gateway to a host of strong auxiliary results and bypass intermediate steps in the algorithm design process.

The structure of the paper is as follows. In Section 2 we give formal definitions and insightful preliminary results. In Section 3 we prove our main result: the linear kernel. The section starts with Subsection 3.1 that gives a high-level overview of how a sequence of lemmas and theorems lead to the kernel, whereas in the rest of the section these lemmas and theorems are proved. Interesting corollaries of the existence of a linear kernel are derived in Section 4: A constant approximation algorithm in Section 4.1; A bound on the ratio between d_{MP} and d_{TBR} in Section 4.2; A bound on the

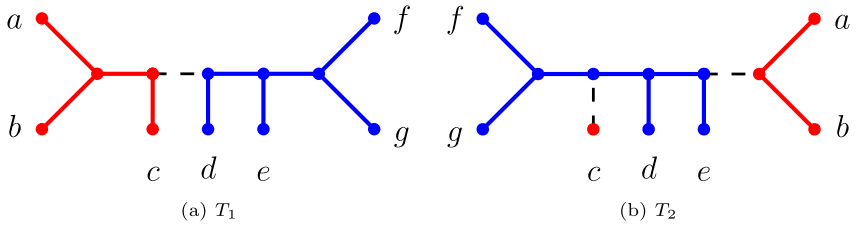


Fig. 1. Two unrooted binary phylogenetic trees T_1, T_2 on $X = \{a, \ldots, g\}$. Solid edges are monochromatic and dashed edges are bichromatic under an optimal extension for the character $\chi : X \to \{\text{RED}, \text{BLUE}\}$, where $\chi(a) = \chi(b) = \chi(c) = \text{RED}$, $\chi(d) = \chi(e) = \chi(f) = \chi(g) = \text{BLUE}$. As there is one bichromatic edge in T_1 and two in T_2, we have that $l_\chi(T_1) = 1$, $l_\chi(T_2) = 2$, proving that $d_{MP}(T_1, T_2) \geq |1 - 2| = 1$. In fact, it can be verified that no character can cause the parsimony scores of these two trees to differ by more, so $d_{MP}(T_1, T_2) = 1$. We will show in Section 4.2 that $d_{TBR}(T_1, T_2) = 2$, because a maximum agreement forest of these two trees contains three blocks [23]. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

treewidth of the so-called display graph in terms of d_{MP} in Section 4.3. Section 5 concludes with some directions for future research.

2. Definitions and preliminaries

An *unrooted binary phylogenetic tree* on a set of species (or taxa) X is an undirected tree in which all internal vertices have degree 3, and the degree-1 vertices (the *leaves*) are bijectively labelled with elements from X. For brevity we will refer to unrooted binary phylogenetic trees as *phylogenetic trees*, or even shorter *trees*. See Fig. 1 for an example.

Given a set $S \subseteq X$ and a tree T on X, we denote by $T[S]$ the *spanning subtree on S in T*, that is, the minimal connected subgraph T' of T such that T' contains every element of S. The *induced subtree $T|_S$ by S in T* is the tree derived from $T[S]$ by suppressing any vertices of degree 2.

Given a subset $S \subseteq X$ and a tree T on X, we say that S has *degree d in T* if there are exactly d edges uv in T for which u is in $T[S]$ and v is not; in other words, d is the number of edges separating $T[S]$ from the rest of T. We call these edges *pending edges of S in T*.

For two disjoint subsets $S_1, S_2 \subseteq X$, we say S_1 and S_2 are *spanning-disjoint* in T if the spanning subtrees $T[S_1]$ and $T[S_2]$ are edge-disjoint. (Observe that as T is binary, this also implies that $T[S_1]$ and $T[S_2]$ are vertex-disjoint.) Similarly, we say a collection $S_1, \ldots S_m$ of subsets of X are *spanning-disjoint* in T if S_i, S_j are spanning-disjoint in T for any $i \neq j$.

2.1. Characters and parsimony

A *character* on X is a function $\chi : X \to \mathbf{C}$, where $\mathbf{C}$ is a set of *states*. In this paper there is no limit on the size of $\mathbf{C}$, in contrast to some contexts where $|\mathbf{C}|$ is assumed to be quite small (for example, in genetic data the nucleobases A, C, G, T). Think of the states as colours, say $1, 2, \ldots, t =: [t]$.

For a given character χ and tree T on X, the *parsimony score* measures how well T fits χ. It is defined in the following way. Call a colouring $\phi : V(T) \to [t]$ an *extension* of χ to T if $\phi(x) = \chi(x)$ for all $x \in X$. Denote by $\Delta_T(\phi)$ the number of bichromatic edges uv in T, i.e. for which $\phi(u) \neq \phi(v)$. We usually omit subscript T when the tree is clear from context. The *parsimony score* for T with respect to χ is defined as

$$l_\chi(T) = \min_\phi \Delta_T(\phi)$$

where the minimum is taken over all possible extensions ϕ of χ to T. An extension ϕ that achieves this bound is called an *optimal extension* of χ to T. An optimal extension, and thus the parsimony score, can be easily computed in polynomial time using dynamic programming or e.g. Fitch's algorithm [19].

Observe that for any T and χ, the parsimony score for T with respect to χ is at least $|\chi(X)| - 1$, i.e. the number of colours assigned by χ minus 1. If $l_\chi(T)$ is exactly $|\chi(X)| - 1$, we say that T is a *perfect phylogeny* for χ. For trees T_1, T_2 and a character χ on X, the *parsimony distance with respect to χ* is defined as

$$d_{MP\,\chi}(T_1, T_2) = |l_\chi(T_1) - l_\chi(T_2)|.$$

Now we are ready to define the *maximum parsimony distance* between two trees (see also Fig. 1). For two trees T_1, T_2 on X, the maximum parsimony distance is defined as

$$d_{MP}(T_1, T_2) = \max_\chi d_{MP\,\chi}(T_1, T_2)$$

where the maximum is taken over all possible characters χ on X [18,31]. Equivalently, we may write it as

$$d_{MP}(T_1, T_2) = \max_\chi |\Delta(\phi_1) - \Delta(\phi_2)|$$

where ϕ_1 is an optimal extension of χ to T_1, and ϕ_2 an optimal extension of χ to T_2. This measure satisfies the properties of a distance metric on the space of unrooted binary phylogenetic trees [18,31]. For two trees on n taxa it is known that d_{MP} is at most $n - 2\sqrt{n} + 1$ [18]. A weaker bound of $n - 1$ is easily obtained by observing that the parsimony score of a character on a tree is at least 0 and at most $n - 1$.

Given a tree T on X and a colouring $\phi : V(T) \to [t]$, the *forest induced by* ϕ is derived from T by deleting every bichromatic edge under ϕ. Observe that the number of connected components in the forest induced by ϕ is exactly $\Delta(\phi) + 1$.

Lemma 1. *If $\chi : X \to [t]$ is a character with $S_i = \chi^{-1}(i) \neq \emptyset$ (i.e. at least one taxa is coloured i) for each $i \in [t]$, and T is a tree on X, then*

$$l_T(\chi) \geq t - 1$$

with equality if and only if $S_1, \ldots S_t$ are spanning-disjoint in T.

Proof. To see that $l_T(\chi) \geq t - 1$, consider an optimal extension ϕ of χ to T, and let F be the forest induced by ϕ. As each connected component in F is monochromatically coloured by ϕ, there must be at least t connected components, and thus $\Delta(\phi) \geq t - 1$, which implies $l_\chi(T) \geq t - 1$.

Now suppose that $S_1, \ldots, S_t$ are spanning-disjoint in T. Then construct an extension ϕ of χ to T by first setting $\phi(u) = i$ for every vertex u in $T[S_i]$, for each $i \in [t]$. (As the spanning trees are edge-disjoint and thus vertex-disjoint in T, this is well-defined). For any remaining unassigned vertices v, if v has a neighbour u for which $\phi(u)$ is defined, then set $\phi(v) = \phi(u)$. Repeat this process until every vertex is assigned a colour by ϕ. Now observe that by construction, the vertices assigned colour i by ϕ form a connected subtree for each $i \in [t]$. Thus the forest induced by ϕ has exactly t connected components, and so $\Delta(\phi) = t - 1$.

Finally, suppose $l_\chi(T) = t - 1$, and let ϕ be an optimal extension of χ. Then the forest F induced by ϕ has exactly t connected components, which implies by the pigeonhole principle that each S_i is a subset of one connected component in F. Then as each S_i is contained within a different connected component of F, the spanning trees $T[S_i]$ are also contained within these components, and so $S_1, \ldots S_t$ are spanning-disjoint. □

2.2. Parameterized complexity and kernelization

A *parameterized problem* is a problem for which the inputs are of the form (x, k), where k is an non-negative integer, called the *parameter*. A parameterized problem is *fixed-parameter tractable* (FPT) if there exists an algorithm that solves any instance (x, k) in $f(k) \cdot |x|^{O(1)}$ time, where $f()$ is a computable function depending only on k. A parameterized problem has a *kernel* of size $g(k)$, where $g()$ is a computable function depending only on k, if there exists a polynomial time algorithm transforming any instance (x, k) into an equivalent problem (x', k'), with $|x'|, k' \leq g(k)$. If $g(k)$ is a polynomial in k then we call this a *polynomial kernel*; if $g(k) = O(k)$ then it is a *linear kernel*. It is well-known that a parameterized problem is fixed-parameter tractable if and only if it has a (not necessarily polynomial) kernel. For more information, we refer the reader to [15].

For a maximization problem Π and $\rho \geq 1$, we say Π has a *constant factor approximation* with *approximation ratio* ρ if there exists a polynomial-time algorithm such that for any instance π of Π, the following inequalities hold, where $opt(\pi)$ denotes the maximum value of a solution to π, and $alg(\pi)$ denotes the value of the solution to π returned by the algorithm:

$$1 \leq \frac{opt(\pi)}{alg(\pi)} \leq \rho$$

In this paper we study the following maximization problem:

MAXIMUM PARSIMONY DISTANCE (DMP)
Input: Two trees T_1, T_2 on a set of taxa X.
Output: A character χ on X that maximizes $|l_\chi(T_1) - l_\chi(T_2)|$.

3. Kernel bound

3.1. Overview

In this section we give an overview of the constituent parts of our kernelization result, and how they fit together.

The first step is to apply two reduction rules, the Cherry rule and the Chain rule, described in the next section. These rules correspond roughly to reduction rules that often appear in papers on computational phylogenetics. The correctness of these rules was proved in [23]; our contribution is to show that the exhaustive application of these rules grants a linear kernel, as stated in the following theorem.

Theorem 1. *There exists a constant α ($\alpha = 560$) for which the following holds. Let (T_1, T_2) be a pair of binary unrooted phylogenetic trees on X that are irreducible under Reduction Rules 1 and 2.*

Then if $|X| \geq \alpha k$, it holds that $d_{MP}(T_1, T_2) \geq k$, and we can find a witnessing character, i.e. a character χ yielding $d_{MP\chi}(T_1, T_2) \geq k$, in polynomial time.

This theorem, together with the correctness of the reduction rules as proved in [23], immediately implies a linear kernel for DMP.

To show how we prove the theorem, we will need to introduce some terminology as we go.

A *quartet* Q is any set of 4 elements in X. If $T_1|_Q \neq T_2|_Q$, we say that Q is a *conflicting quartet* for (T_1, T_2).

As a crucial step we prove that for any S large enough with respect to the degree of S in both T_1 and T_2, either there exists a conflicting quartet or one of the reduction rules applies.

Lemma 2. *Let S be a subset of X with d_1 the degree of S in T_1, and d_2 the degree of S in T_2. If $|S| > 9(d_1 + d_2) - 12$, then either $T_1|_S \neq T_2|_S$ or one of Reduction Rules 1 or 2 applies to (T_1, T_2). In particular if (T_1, T_2) is irreducible under Rules 1 or 2 and $|S| \geq 9(d_1 + d_2) - 11$, then there exists a conflicting quartet $Q \subseteq S$, and such a quartet can be found in polynomial time.*

The next result implies that if we have a large enough number of conflicting quartets that are also spanning-disjoint in both T_1 and T_2, then we are done. While it is intuitively clear that such quartets can be leveraged to create a high parsimony score in one tree, some care has to be taken to keep the parsimony score low in the other tree.

Lemma 3. *Let $\mathcal{Q} = \{Q_1, \ldots, Q_k\}$ be a set of conflicting quartets for T_1, T_2, such that $Q_1, \ldots Q_k$ are spanning-disjoint in T_1 and in T_2.*

Then $d_{MP}(T_1, T_2) \geq k$, and we can find a witnessing character in polynomial time.

In combination, Lemmas 2 and 3 allow us to show that $d_{MP}(T_1, T_2) \geq k$ provided that we can find at least k sets $S_1, \ldots S_k$ that are spanning-disjoint in both trees and satisfy the conditions of Lemma 2.

We will find k such sets as part of the construction of a character that witnesses $d_{MP}(T_1, T_2) \geq k$, for any reduced instance with $|X| \geq \alpha k$. In order to construct this character, we first create a partition of X into large subsets, as described by the following lemma.

Lemma 4. *Suppose that $|X| \geq 2ct$ for some integers c and t, and let T_1 be a phylogenetic tree on X.*

Then in polynomial time we can construct a partition $S_1, \ldots, S_t$ of X with $S_1, \ldots, S_t$ spanning-disjoint in T_1, such that $|S_i| \geq c$ for each i.

We note that there is a one-to-one correspondence between partitions and characters on X, in the following sense. Given a partition $S_1, \ldots S_t$ of X, we may define a character $\chi : X \to [t]$ such that $\chi(x) = i$ if $x \in S_i$, for each $i \in [t]$. Call such a character the character *defined* by $S_1, \ldots S_t$.

Thus let us consider the character χ on X defined by the partition described by Lemma 4. Since $S_1, \ldots S_t$ are spanning-disjoint in T_1, Lemma 1 tells that the parsimony score of T_1 with respect to χ is exactly $t - 1$.

Lemma 5. *Let χ be the character defined by the partition $S_1, \ldots, S_t$ where $S_1, \ldots, S_t$ are spanning-disjoint in T_1, let d_1, d_2 be positive integers such that $d_1 d_2 - d_1 - d_2 > 0$, and assume*

$$t \geq \left\lceil \frac{(2d_1 d_2 + d_1)}{d_1 d_2 - d_1 - d_2} \right\rceil k.$$

Then either $d_{MP\chi}(T_1, T_2) \geq k$, or in polynomial time we can find a set of indices $i_1, \ldots i_{k'}$ with $k' \geq k$ such that:

- *$S_{i_1}, \ldots S_{i_{k'}}$ are spanning-disjoint in T_2 (as well as in T_1);*
- *S_{i_j} has degree at most d_1 in T_1 for each $j \in [k']$; and*
- *S_{i_j} has degree at most d_2 in T_2 for each $j \in [k']$.*

We will prove Theorem 1 by combining these results in the following way. Fix integers d_1, d_2 to be determined later. Assume (T_1, T_2) is irreducible under Reduction Rules 1 and 2, and assume that

$$|X| \geq 2ct, \text{ where } c = 9(d_1 + d_2) - 11 \text{ and } t \geq \lceil \frac{(2d_1 d_2 + d_1)}{d_1 d_2 - d_1 - d_2} \rceil k$$

(this holds if $|X| \geq \alpha k$).

By Lemma 4, there exists a partition $S_1, \ldots S_t$ of X with $S_1, \ldots S_t$ spanning-disjoint in T_1 and $|S_i| \geq c$ for each $i \in [t]$. Let χ be the character defined by this partition. If $d_{MP\chi}(T_1, T_2) \geq k$, we may return χ. Otherwise, we may apply Lemma 5 to

get a set of indices $i_1, \dots i_k$ such that $S_{i_1}, \dots S_{i_k}$ are spanning-disjoint in T_2 (as well as in T_1), each S_{i_j} has degree at most d_1 in T_1, and each S_{i_j} has degree at most d_2 in T_2. But then each S_{i_j} satisfies the conditions of Lemma 2, and therefore for each $j \in [k]$ there exists a conflicting quartet $Q_j \subseteq S_{i_j}$. Moreover, as $S_{i_1}, \dots S_{i_k}$ are spanning-disjoint in T_1 and T_2, the quartets $Q_1, \dots Q_k$ are also spanning-disjoint in T_1 and T_2. Then Lemma 3 implies that $d_{MP}(T_1, T_2) \geq k$.

By setting $d_1 = 4$ and $d_2 = 5$, we get that $\alpha = 560$, giving the desired bound.

In the next subsections we prove each of these lemmas, and then the main theorem, in turn.

3.2. Reduction rules

We begin by stating the reduction rules for our kernelization result. In what follows, a pair (x, y) with $x, y \in X$ is a *cherry* in a tree T if there exists an internal vertex u in T adjacent to both x and y. A cherry is also sometimes known in the literature as a *sibling-pair*. A sequence of leaves $x_1, \dots x_r \in X$ is a *chain* in T if there exists a path of internal vertices $p_1, \dots, p_r$ (possibly with $p_1 = p_2$ and possibly with $p_{r-1} = p_r$), such that for each $i \in [r]$ p_i is the internal vertex adjacent to x_i. We call r the *length* of this chain.

Reduction Rule 1. *[Cherry reduction rule] If there exist $x, y \in X$ such that (x, y) is a cherry in each of T_1, T_2, then replace (T_1, T_2) with $(T_1|_{X\setminus\{x\}}, T_2|_{X\setminus\{x\}})$.*

Reduction Rule 2. *[Chain reduction rule] If there exists a sequence of leaves $x_1, \dots x_r \in X$ such that $x_1, \dots x_r$ is chain in both T_1 and T_2, and $r \geq 5$, then replace (T_1, T_2) with $(T_1|_{X\setminus\{x_5,\dots,x_r\}}, T_2|_{X\setminus\{x_5,\dots x_r\}})$ (thus, the common chain is reduced to length 4).*

The correctness of these rules (in the sense that they preserve d_{MP}) was previously proved in [23].

Theorem 2. *Let (T_1', T_2') be an instance of* DMP *derived from (T_1, T_2) by an application of Reduction Rules 1 or 2. Then*

$$d_{MP}(T_1', T_2') = d_{MP}(T_1, T_2).$$

Correctness of the chain reduction rule follows from Theorem 3.1 in [23]. Correctness of the cherry reduction rule follows as a subcase of Theorem 4.1 in [23].

Our main contribution is to show that if an instance is reduced by these rules then its size is bounded by a linear function of d_{MP}.

3.3. Small degree sets

In this section we prove Lemma 2.

Lemma 2. *Let S be a subset of X with d_1 the degree of S in T_1, and d_2 the degree of S in T_2. If $|S| > 9(d_1 + d_2) - 12$, then either $T_1|_S \neq T_2|_S$ or one of Reduction Rules 1 or 2 applies to (T_1, T_2). In particular if (T_1, T_2) is irreducible under Rules 1 or 2 and $|S| \geq 9(d_1 + d_2) - 11$, then there exists a conflicting quartet $Q \subseteq S$, and such a quartet can be found in polynomial time.*

Proof. Since unrooted binary trees are characterized by their quartets [34, Theorem 6.3.5(iii)], the last statement of the lemma follows directly.

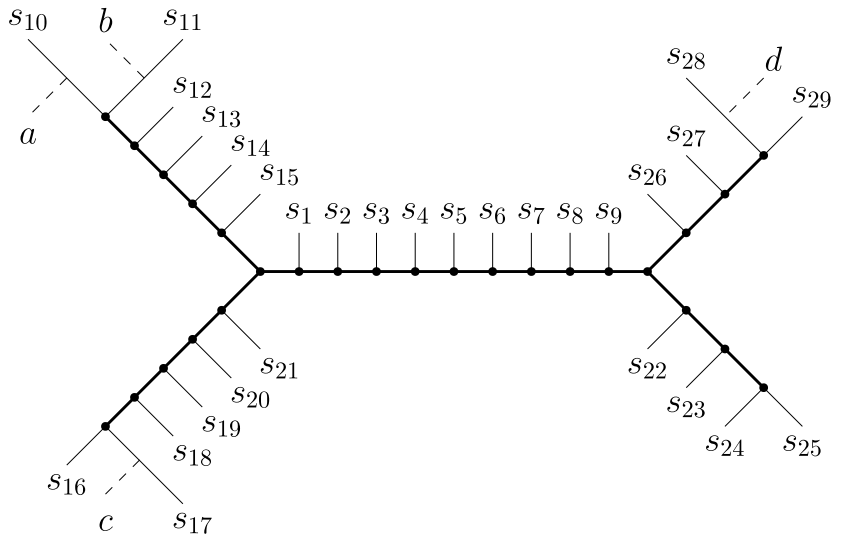
We will show that if $T_1|_S = T_2|_S$ and neither of the reduction rules applies to (T_1, T_2), then $|S| \leq 9(d_1 + d_2) - 12$. This implies the main claim of the lemma. Let us denote $T|_S = T_1|_S = T_2|_S$.

Consider the *backbone* graph of $T|_S$ obtained by deleting all leaves (see Fig. 2 for an example). Let P_C be the set of nodes having degree 1 on the backbone, which we refer to as *parents* of a cherry in $T|_S$. Let P_L be the set of nodes having degree 2 on the backbone, which we refer to as *parents* of a leaf of $T|_S$. All remaining vertices on the backbone have degree 3. Thus $|S|$, the total number of leaves of $T|_S$ is $2|P_C| + |P_L|$. We call the path between any two odd degree vertices on the backbone, having internal nodes only in P_L, a *side* of the backbone.

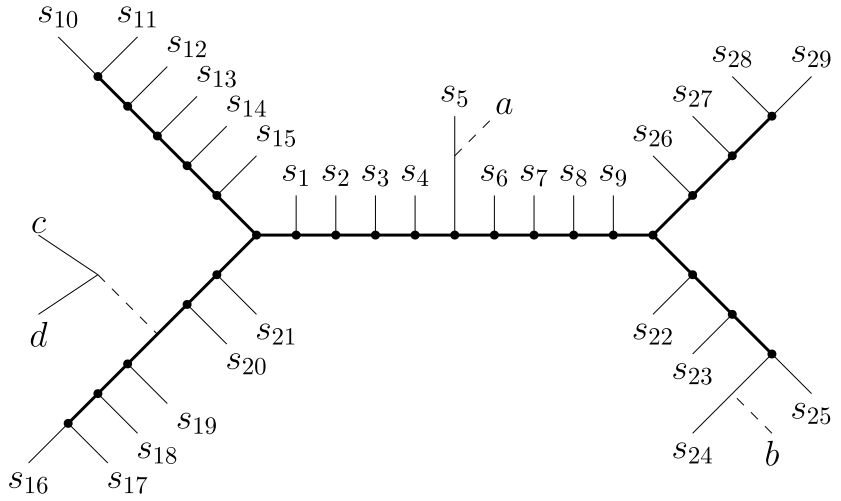
First notice that for each cherry in $T|_S$, there must exist in $T_1[S]$, the spanning tree on S in T_1, or in $T_2[S]$ a node, incident to a pending edge of S, between at least one of its two leaves and its corresponding node in P_C. Otherwise Reduction Rule 1 can be applied. In particular this implies that $|P_C| \leq d_1 + d_2$.

Thus at least P_C of the $d_1 + d_2$ pending edges must be used for "cutting" the cherries, each of them cutting 1 leaf of a cherry. Let us choose one such leaf from each cherry, and call these the *cut-leaves*.

After removing cut-leaves, every node in P_C and P_L is now the parent of 1 leaf in $T|_S$. Every side of the backbone contains at most 4 vertices in P_C and P_L, unless $T_1[S]$ or $T_2[S]$ has a node of a pending edge of S or a node adjacent to a node of a pending edge on that side. We show that every such pending edge on a side may increase the number of P_L-nodes on that side by at most 5 (see Fig. 2). Indeed, suppose a side of the backbone has in total d pending edges of S in both T_1 and T_2, but more than $4 + 5d$ nodes in P_L, i.e. at least $5(d + 1)$. Then $T|_S$ contains a chain of length $5(d + 1)$,



(a) Backbone of $T_1|_S$ within T_1.



(b) Backbone of $T_2|_S$ within T_2

Fig. 2. Example illustration of the backbone of $T|_S = T_1|_S = T_2|_S$ within T_1 and T_2, where $S = \{s_1, \ldots, s_{29}\}$. Edges and vertices of the backbone are in bold. Observe that $T|_S$ has the chain $s_1, \ldots, s_9$, but (T_1, T_2) do not have a common chain of length greater than 4, as the leaf s_5 has a sibling a in T_2.

which we can split up into $d+1$ chains of length 5. Clearly at least one of these chains has no pending edge in either T_1 or T_2, and so T_1, T_2 have a common chain of length 5, a contradiction.

Thus the total number of nodes from P_C and P_L on a side is at most five times the number of pending edges of S (in $T_1[S]$ or $T_2[S]$) on that side, plus 4. Otherwise Reduction Rule 2 can be applied. Given that we already used $|P_C|$ pending edges for cutting the cherries, we have $d_1 + d_2 - |P_C|$ pending edges left to be distributed over the sides.

The number of sides on the backbone is the number of edges in an unrooted binary tree with $|P_C|$ leaves, which is $2|P_C| - 3$. Therefore the total number of leaves of $T|_S$ is

$$\begin{aligned} |S| = 2|P_C| + |P_L| &\leq |P_C| + 4(2|P_C| - 3) + 5(d_1 + d_2 - |P_C|) \\ &\leq 4|P_C| + 5(d_1 + d_2) - 12. \end{aligned}$$

Clearly, this attains its largest value if $|P_C| = d_1 + d_2$, in which case $|S| \leq 9(d_1 + d_2) - 12$, as was to be proven. □

3.4. Combining conflicting quartets

In this section we prove Lemma 3.

Lemma 3. *Let $\mathcal{Q} = \{Q_1, \ldots, Q_k\}$ be a set of conflicting quartets for T_1, T_2, such that $Q_1, \ldots Q_k$ are spanning-disjoint in T_1 and in T_2.*

Then $d_{MP}(T_1, T_2) \geq k$, and we can find a witnessing character in polynomial time.

Proof. For a quartet Q and tree T, we say that $T|_Q = ab|cd$ if $Q = \{a, b, c, d\}$ and in T the path between a and b is edge-disjoint from the path between c and d. Without loss of generality, we may assume $Q_i = \{a_i, b_i, c_i, d_i\}$, $T_1|_{Q_i} = a_ib_i|c_id_i$ and $T_2|_{Q_i} = a_ic_i|b_id_i$ for each $i \in [k]$.

We will show how to build a character χ with two states, such that $l_\chi(T_1) \leq k$, and $l_\chi(T_2) \geq 2k$. This shows that $d_{MP\,\chi}(T_1, T_2) \geq k$, as required.

The idea is to construct χ in such a way that, for each quartet Q_i, $\chi(a_i) = \chi(b_i) \neq \chi(c_i) = \chi(d_i)$. This will ensure that $l_\chi(T_2)$ is at least $2k$, as T_2 will have at least $2k$ edge-disjoint paths (from a_i to c_i and from b_i d_i, for each $i \in [k]$) that each require at least one change in state along some edge.

For each Q_i, let e_{Q_i} denote an edge in T_1 such that in $T_1[Q_i]$, e_i is on the path that separates $\{a_i, b_i\}$ from $\{c_i, d_i\}$.

Now we construct a function $\phi : V(T_1) \to \{\text{RED}, \text{BLUE}\}$ as follows. Start by choosing an arbitrary leaf in T_1, say without loss of generality a_1, and set $\phi(a_1) = \text{RED}$. Now proceed as follows. For any edge uv in T_1 such that $\phi(u)$ is defined but $\phi(v)$ is not, we set $\phi(v) = \phi(u)$, unless $uv = e_{Q_i}$ for some i. In that case, we set $\phi(v) = \text{BLUE}$ if $\phi(u) = \text{RED}$, and set $\phi(v) = \text{RED}$ otherwise.

Now we can let χ be the restriction of ϕ to X. By construction, ϕ is an extension of χ to T_1 and $\Delta(\phi) = |e_{Q_i} : i \in [k]| = k$. This is enough to show that $l_\chi(T_1) \leq k$.

We now show that $\chi(a_i) = \chi(b_i) \neq \chi(c_i) = \chi(d_i)$, for each $i \in [k]$. To see this, consider the spanning tree $T_1[Q_i]$. By construction, $T_1[Q_i]$ contains the edge e_{Q_i} and e_{Q_i} separates $\{a_i, b_i\}$ from $\{c_i, d_i\}$. Let u_i, v_i be the vertices of e_{Q_i}, with u_i the vertex closer to a_i and b_i. Note that $T_1[Q_i]$ cannot contain e_{Q_j} for any $j \neq i$, as $T_1[Q_i]$ and $T_1[Q_j]$ are edge-disjoint. It follows that u_i, a_ib_i are all assigned the same value by ϕ and v_i, c_i, d_i are assigned the opposite value. Thus by definition of χ, we have $\chi(a_i) = \chi(b_i) = \phi(u_i) \neq \phi(v_i) = \chi(c_i) = \chi(d_i)$.

It remains to observe that as $Q_1, \ldots Q_k$ are spanning-disjoint in T_2, the $a_i - c_i$ and $b_i - d_i$ paths in T_2 are pairwise edge-disjoint for all $i \in [k]$. Then as $\chi(a_i) \neq \chi(c_i)$ and $\chi(b_i) \neq \chi(d_i)$, there exist at least $2k$ edges uv in T_2 with $\phi_2(u) \neq \phi_2(v)$, for any extension ϕ_2 of χ to T_2. It follows that $l_\chi(T_2) \geq 2k$, and so $d_{MP}(T_1, T_2) \geq d_{MP\,\chi}(T_1, T_2) = |l_\chi(T_1) - l_\chi(T_2)| \geq 2k - k = k$.

Since each edge is processed at most once in the construction of χ, it is clear that this construction takes polynomial time. □

3.5. Constructing an initial partition

In this section we prove Lemma 4.

Lemma 4. *Suppose that $|X| \geq 2ct$ for some integers c and t, and let T_1 be a phylogenetic tree on X.*

Then in polynomial time we can construct a partition $S_1, \ldots, S_t$ of X with $S_1, \ldots, S_t$ spanning-disjoint in T_1, such that $|S_i| \geq c$ for each i.

Proof. We prove the claim by induction on t. For the base case, if $t = 1$ then we may let $S_1 = X$, and we have the desired partition.

For the inductive step, assume $|X| \geq 2ct$ and that the claim is true for smaller values of t. We first fix an arbitrary rooting on T_1. That is, choose an arbitrary edge e in T_1 and subdivide it with a new (temporary) vertex r, then orient all edges in T_1 away from r. Under this rooting, let u be a lowest vertex in T_1 for which u has at least c descendants in X. Let $S_t \subseteq X$ be the set of these descendants. Note that since T_1 is binary, $|S_t| < 2c$, as otherwise one of the two children of u would be a lower vertex with at least c descendants.

Now consider the induced subtree $T_1|_{X'}$, where $X' = X \setminus S_t$. As $|S_t| < 2c$, we have $X' \geq 2c(t-1)$. Then by the inductive hypothesis, we can construct a partition $S_1, \ldots, S_{t-1}$ of X' with $S_1, \ldots, S_{t-1}$ spanning-disjoint in $T_1|_{X'}$, such that $|S_i| \geq c$ for each i. By construction it is clear that S_t is spanning-disjoint in T_1 from $S_1, \ldots, S_{t-1}$. Thus $S_1, \ldots, S_t$ is the desired partition.

As the construction of S_t can be done in polynomial time and this process is repeated $t \leq |X|$ times, the entire process takes polynomial time. □

3.6. Well-behaved sets

In this section we prove Lemma 5. We start with an observation:

Observation 1. *For any (not necessarily binary) unrooted tree T with n vertices, and any integer $d \geq 1$, the number of vertices in T with degree strictly greater than d is at most n/d.*[1]

[1] The proof of this observation is based on an argument in [3].

Proof. For each vertex v in T let $d(v)$ denote the degree of v. Recall that an unrooted tree with n vertices has exactly $n-1$ edges. It follows that

$$\sum_{v \in V(T)} d(v) = 2|E(T)| = 2n - 2.$$

Now suppose that T has $m > n/d$ vertices with degree strictly greater than d, i.e. at least $d+1$. The remaining $n-m$ vertices all have degree at least 1, from which it follows that

$$\sum_{v \in V(T)} d(v) \geq m(d+1) + n - m = md + n \geq (n/d)d + n = 2n,$$

a contradiction. □

Lemma 5. *Let χ be the character defined by the partition $S_1, \ldots, S_t$ where $S_1, \ldots, S_t$ are spanning-disjoint in T_1, let d_1, d_2 be positive integers such that $d_1 d_2 - d_1 - d_2 > 0$, and assume*

$$t \geq \left\lceil \frac{(2d_1 d_2 + d_1)}{d_1 d_2 - d_1 - d_2} \right\rceil k.$$

Then either $d_{MP\chi}(T_1, T_2) \geq k$, or in polynomial time we can find a set of indices $i_1, \ldots i_{k'}$ with $k' \geq k$ such that:

- *$S_{i_1}, \ldots S_{i_{k'}}$ are spanning-disjoint in T_2 (as well as in T_1);*
- *S_{i_j} has degree at most d_1 in T_1 for each $j \in [k']$; and*
- *S_{i_j} has degree at most d_2 in T_2 for each $j \in [k']$.*

Proof. By Lemma 1, $l_\chi(T_1) = t - 1$. If $l_\chi(T_2) \geq t + k - 1$, then $d_{MP\chi}(T_1, T_2) \geq k$ as required. So we may assume that $l_\chi(T_2) \leq t + k - 2$. Let $\delta = l_\chi(T_2) - l_\chi(T_1)$, and observe that $0 \leq \delta \leq k - 1$.

We now construct a partition $P_1, \ldots P_s$ of X which is spanning-disjoint in T_2 (see Fig. 3 for an illustration). Let ϕ_2 be an optimal extension of χ to T_2. As $l_\chi(T_2) = l_\chi(T_1) + \delta = t + \delta - 1$, the forest induced by ϕ_2 has exactly s monochromatic connected components, where $s = t + \delta$. Let $P_1, \ldots, P_s$ be the partition of X formed by taking the intersection of X with the vertex set of each tree in this forest. Observe that by construction $P_1, \ldots P_s$ are spanning-disjoint in T_2, and that furthermore each P_j is a subset of S_i for some $i \in [t]$ (as each element of P_j is assigned the same value by ϕ_2, and thus by χ).

Now let $\mathcal{I} \subseteq [t]$ denote the set of indices i in $[t]$ such that

- $S_i = P_j$ for some $j \in [s]$;
- S_i has degree at most d_1 in T_1; and
- S_i has degree at most d_2 in T_2.

Note that since $P_1, \ldots P_j$ are spanning-disjoint in T_2, the sets $\{S_i : i \in \mathcal{I}\}$ are also spanning-disjoint in T_2. Notice that it is sufficient to prove that $|\mathcal{I}| \geq k$, whence any subset of k indices from $\mathcal{I}$ satisfies the lemma. We will prove this by providing upper bounds on the number of indices in $[t]$ that do not satisfy the conditions of $\mathcal{I}$.

Let $\mathcal{I}_0$ denote the set of indices $i \in [t]$ such that $P_j \neq S_i$ for any $j \in [s]$. We first claim that $|\mathcal{I}_0| \leq \delta$. Indeed, since every P_j is a subset of some S_i and $S_1, \ldots S_t$ and $P_1, \ldots, P_s$ are both partitions of X, we have that for every $i \in \mathcal{I}_0$, there exist at least two distinct indices $j, j' \in [s]$ for which $P_j, P_{j'} \subset S_i$. Hence,

$$s \geq 2|\mathcal{I}_0| + |[t] \setminus \mathcal{I}_0| = t + |\mathcal{I}_0|.$$

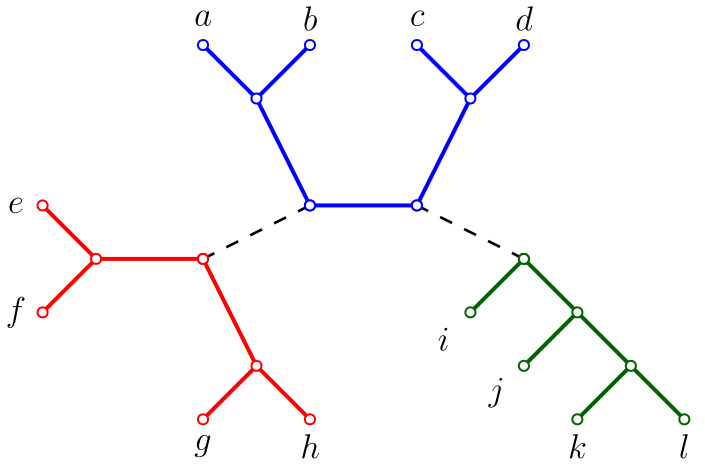
Therefore if $|\mathcal{I}_0| > \delta$ then $s > t + \delta$, contradicting the definition of s. Thus, we have $|\mathcal{I}_0| \leq \delta$.

Next, let $\mathcal{I}_{>d_1}$ denote the set of indices $i \in [t]$ for which S_i has degree greater than d_1 in T_1. We will show that $|\mathcal{I}_{>d_1}| \leq t/d_1$. For each $i \in [t]$, compress the spanning subtree $T_1[S_i]$ to a single vertex, and observe that the degree of this vertex is equal to the degree of S_i in T_1. Any vertex u which is not part of any $T_1[S_i]$ is merged with one of its neighbours. Note that this merging process can only increase the degrees of the remaining vertices. Call the resulting tree T_1'. See Fig. 4. T_1' has t vertices, each of them corresponding to a subset S_i, and having degree at least the degree of the corresponding S_i in T_1. Now by Observation 1, there are at most t/d_1 vertices in T_1' with degree greater than d_1. It follows that there are at most t/d_1 values of $i \in [t]$ for which S_i has degree greater than d_1 in T_1, and thus $|\mathcal{I}_{>d_1}| \leq t/d_1$ as we wanted to show.

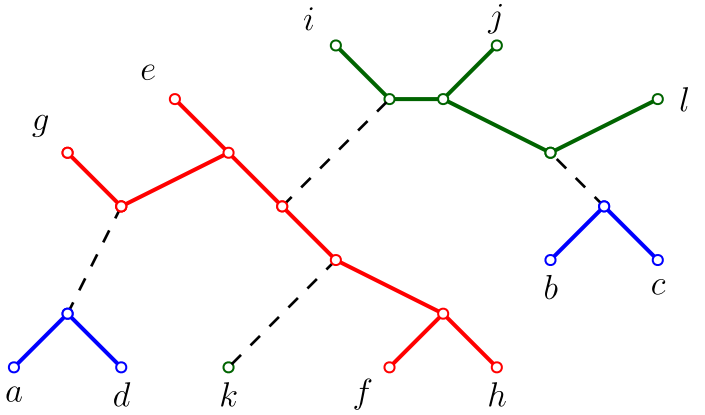
Similarly let $\mathcal{J}_{>d_2}$ denote the set of indices $j \in [s]$ for which P_j has degree greater than d_2 in T_2. By similar arguments as used for $\mathcal{I}_{>d_1}$ above, we can show that $|\mathcal{J}_{>d_2}| \leq s/d_2$.

Notice that for any $i \in [t]$, if i is not in $\mathcal{I}$, then either $i \in \mathcal{I}_0$, or $i \in \mathcal{I}_{>d_1}$, or there exists $j \in \mathcal{J}_{>d_2}$ such that $S_i = P_j$. We therefore have that

$$|\mathcal{I}| \geq t - |\mathcal{I}_0| - |\mathcal{I}_{>d_1}| - |\mathcal{J}_{>d_2}| \geq t - \delta - t/d_1 - s/d_2.$$

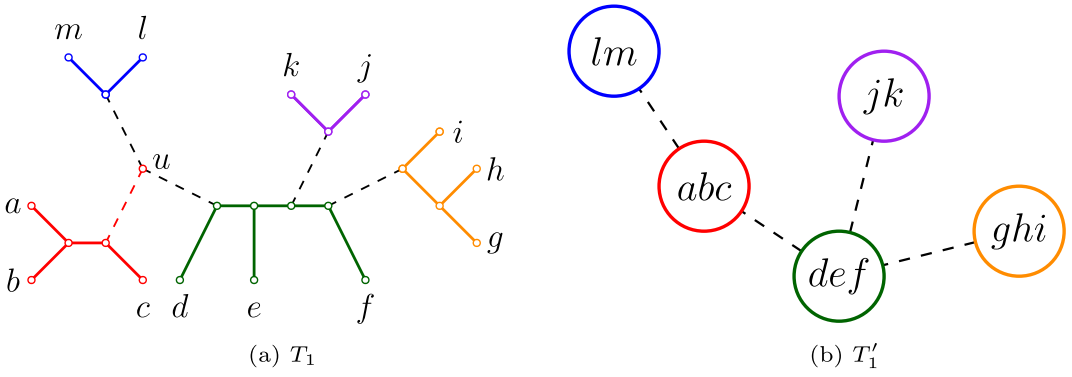


(a) The initial partition $S_1 = \{a, b, c, d\}$, $S_2 = \{e, f, g, h\}$, $S_3 = \{i, j, k, l\}$. Solid edges are monochromatic and dashed edges are bichromatic under an optimal extension of the character χ to T_1, where χ is the character corresponding to S_1, S_2, S_3.



(b) An optimal extension of χ to T_2. Note that the original partition S_1, S_2, S_3 is not spanning-disjoint in T_2, and as such, there are distinct monochromatic components assigned the same color (such as the monochromatic components on $\{a, d\}$ and $\{b, c\}$). The leaf sets of these components give us the new partition $P_1 = \{a, d\}$, $P_2 = \{b, c\}$, $P_3 = \{e, f, g, h\}$, $P_4 = \{i, j, l\}$, $P_5 = \{k\}$.

Fig. 3. Illustration of the construction of partition P_1, P_2, P_3, P_4, P_5 from S_1, S_2, S_3. Solid edges are monochromatic and dashed edges are bichromatic under an optimal extension for χ, where χ is the character induced by S_1, S_2, S_3.



(a) T_1 (b) T_1'

Fig. 4. Illustration of the construction of auxiliary tree T_1', given a partition of X with $S_1 = \{a, b, c\}$, $S_2 = \{d, e, f\}$, $S_3 = \{g, h, i\}$, $S_4 = \{j, k\}$, $S_5 = \{l, m\}$. Note that the internal vertex labelled u is not part of $T_1[S_i]$ for any i, so we merge it with an arbitrary adjacent vertex. In this case we merge u into $S_1 = \{a, b, c\}$, which is why S_1 has degree 1 in T_1 but degree 2 in T_1'.

Now, using that $t \geq \frac{(2d_1d_2+d_1)}{d_1d_2-d_1-d_2}k$, $s = t + \delta$ and $\delta \leq k-1$, we have:

$$\begin{aligned}
|\mathcal{I}| &\geq t - |\mathcal{I}_0| - |\mathcal{I}_{>d_1}| - |\mathcal{J}_{>d_2}| \\
&\geq t - \delta - t/d_1 - s/d_2 \\
&= t - \delta - t/d_1 - (t+\delta)/d_2 \\
&= \frac{d_1d_2t - d_1d_2\delta - d_2t - d_1t - d_1\delta}{d_1d_2} \\
&= \frac{(d_1d_2 - d_1 - d_2)t - (d_1d_2 + d_1)\delta}{d_1d_2} \\
&\geq \frac{(d_1d_2 - d_1 - d_2)t - (d_1d_2 + d_1)(k-1)}{d_1d_2} \\
&\geq \frac{(2d_1d_2 + d_1)k - (d_1d_2 + d_1)(k-1)}{d_1d_2} \\
&= \frac{d_1d_2k + d_1d_2 + d_1}{d_1d_2} \\
&> \frac{d_1d_2k}{d_1d_2} \\
&= k,
\end{aligned}$$

as we needed to prove. To see that $\mathcal{I}$ can be constructed in polynomial time, it suffices to observe that the partition $P_1, \ldots, P_s$ can be constructed in polynomial time (as the ϕ_2 can be found in polynomial time), and after this each S_i can be checked for membership in $\mathcal{I}$ in polynomial time. □

3.7. Proof of Theorem 1

Lemma 6. *Let d_1, d_2 be positive integers such that $d_1d_2 - d_1 - d_2 > 0$. Let (T_1, T_2) be a pair of binary unrooted phylogenetic trees on X that are irreducible under Reduction Rules 1 and 2.*

Then if $|X| \geq 2ct$, where $c = 9(d_1 + d_2) - 11$ and $t = \lceil \frac{(2d_1d_2+d_1)}{d_1d_2-d_1-d_2} \rceil k$, it holds that $d_{MP}(T_1, T_2) \geq k$, and we can find a witnessing character in polynomial time.

Proof. By Lemma 4, there exists a partition $S_1, \ldots S_t$ of X, all spanning-disjoint in T_1, and with $|S_i| \geq c$ for all $i \in [t]$. Let χ be the character defined by $S_1, \ldots, S_t$. If χ is a witness to $d_{MP}(T_1, T_2) \geq k$, then we may return χ and we are done. Otherwise, we may apply Lemma 5 to find indices $i_1, \ldots i_k$ such that:

- $S_{i_1}, \ldots S_{i_k}$ are all spanning-disjoint in T_2 (as well as in T_1);
- each S_{i_j} has degree at most d_1 in T_1; and
- each S_{i_j} has degree at most d_2 in T_2.

Now for each S_{i_j}, we have that S_{i_j} has degree $d_1^j \leq d_1$ in T_1 and $d_2^j \leq d_2$ in T_2, and that

$$|S_{i_j}| \geq c > 9(d_1 + d_2) - 11 \geq 9(d_1^j + d_2^j) - 11,$$

and also that (T_1, T_2) is irreducible under Rules 1 and 2. Thus we may apply Lemma 2, to find a conflicting quartet $Q_j \subseteq S_{i_j}$ for each i_j.

Finally, as $S_{i_1}, \ldots S_{i_k}$ are spanning-disjoint in both T_1 and T_2, and as each Q_j is a subset of S_{i_j}, we have that $Q_1, \ldots, Q_k$ are also spanning-disjoint in both T_1 and T_2. Therefore we may apply Lemma 3 to find a witnessing character for $d_{MP}(T_1, T_2) \geq k$. As each step of this process takes polynomial time, the construction of a witnessing character takes polynomial time. □

It remains to complete the proof of Theorem 1.

Theorem 1. *There exists a constant α ($\alpha = 560$) for which the following holds. Let (T_1, T_2) be a pair of binary unrooted phylogenetic trees on X that are irreducible under Reduction Rules 1 and 2.*

Then if $|X| \geq \alpha k$, it holds that $d_{MP}(T_1, T_2) \geq k$, and we can find a witnessing character, i.e. a character χ yielding $d_{MP\chi}(T_1, T_2) \geq k$, in polynomial time.

Proof. The proof boils down to choosing the appropriate values of d_1 and d_2 such that $2ct = (9(d_1 + d_2) - 11) \cdot \lceil \frac{(2d_1d_2+d_1)}{d_1d_2-d_1-d_2} \rceil k = \alpha k$. For $d_1 = 4, d_2 = 5$ we get $c = 70$ and $t = 4k$, yielding the value of $\alpha = 560$ for $\alpha k = 2ct$. □

In the appendix, we show that $d_1 = 4, d_2 = 5$ is in fact the optimal choice of values for d_1 and d_2.

As a corollary to Theorem 1 and Theorem 2, we have that DMP is fixed-parameter tractable with respect to d_{MP}. Specifically, the kernel can be solved using the exponential-time algorithm described in [26], which computes the maximum parsimony distance of two trees on n leaves in time $O(1.619^n \cdot \text{poly}(n))$.

Corollary 1. DMP *has a kernel with at most* $\alpha(k+1)$ *taxa, and can be solved in time* $O(1.619^{\alpha k} \cdot poly(\alpha k) + poly(n))$, *with* $k = d_{MP}(T_1, T_2)$.

Proof. Given an instance (T_1, T_2) of DMP where T_1, T_2 are trees on a set of taxa X, let (T_1', T_2') be the instance derived from (T_1, T_2) by exhaustively applying Reduction Rules 1 and 2. As each reduction rule can be applied in polynomial time, and each application of the rule reduces the number of taxa, (T_1', T_2') can be derived in polynomial time. Moreover (T_1', T_2') is irreducible under Rules 1 and 2, and by Theorem 2, $d_{MP}(T_1', T_2') = d_{MP}(T_1, T_2) = k$.

Let X' be the leaf set of (T_1', T_2'), and suppose for a contradiction that $|X'| \geq \alpha(k+1)$. Then Theorem 1 implies that $d_{MP}(T_1', T_2') \geq k+1$, a contradiction. Thus $|X'| < \alpha(k+1)$, and (T_1', T_2') is the desired kernel.

To see that DMP can be solved in time $O(1.619^{\alpha k} \cdot poly(\alpha k) + poly(n))$, recall that DMP has a simple exponential-time algorithm with running time $O(\phi^n \cdot \text{poly}(n))$, where $|X| = n$ and $\phi < 1.619$ is the golden ratio [26]. Applying this algorithm to our kernel, we get an algorithm with running time

$$\text{poly}(n) + O(1.619^{\alpha(k+1)} \cdot \text{poly}(\alpha(k+1))) = O(1.619^{\alpha k} \cdot \text{poly}(\alpha k) + \text{poly}(n)) \quad \square$$

For completeness, we clarify that these results also prove that the decision problems "$d_{MP} \leq k$?", "$d_{MP} \geq k$?" and "$d_{MP} = k$?" can all be answered in time $f(k) \cdot poly(n)$. To answer "$d_{MP} \leq k$?", note that if the kernel has size at least $\alpha(k+1)$ the answer is definitely NO, and otherwise the algorithm from [26] can be applied to compute d_{MP} directly; this can then be compared to k to resolve the question. The "$d_{MP} \geq k$?" question can be answered by asking "$d_{MP} \leq k-1$?" and negating the answer; and "$d_{MP} = k$?" can be answered by combining the answers to the $\leq$ and $\geq$ questions.

4. Corollaries: leveraging the kernel

4.1. A polynomial-time constant-factor approximation algorithm for DMP

We present how a constant factor approximation algorithm for DMP can be designed using Theorem 1 together with Reduction Rules 1 and 2.

In order to incorporate Reduction Rules 1 and 2 into our approximation algorithm, we require a way to construct a witnessing character for the original instance from a witnessing character for the reduced instance.

Lemma 7. *Let* (T_1', T_2') *be an instance of* DMP *derived from* (T_1, T_2) *by an application of Reduction Rule 1 or 2, with* T_1', T_2' *trees on* $X' \subset X$. *Then given a character* χ' *on* X', *we can derive a character* χ *on* X *in polynomial time such that* $d_{MP\,\chi}(T_1, T_2) \geq d_{MP\,\chi'}(T_1', T_2')$.

Proof. First observe that by definition of the reduction rules, we may assume that $T_1' = T_1|_{X'}$ and $T_2' = T_2|_{X'}$ for some $X' \subseteq X$. Assume without loss of generality that $l_{\chi'}(T_2') \geq l_{\chi'}(T_1')$, and let ϕ_1' be an optimal extension of χ' to T_1'. We will now define a function $\phi : V(T_1) \to \mathbf{C}$ such that $\phi(u) = \phi'(u)$ for all $u \in V(T_1')$, and such that $\Delta_{T_1}(\phi_1) = \Delta_{T_1'}(\phi_1') = l_{\chi'}(T_1')$.

Recall that $T_1|_{X'}$ is derived from the spanning tree $T_1[X']$ by suppressing vertices of degree 2, and therefore $T_1[X']$ can be derived from $T_1' = T_1|_{X'}$ by repeatedly subdividing edges with degree-2 vertices. Now construct ϕ_1 as follows. For each vertex v in T_1', set $\phi_1(v) = \phi_1'(v)$. For every edge $e = uv$ that gets subdivided with one or more degree-2 vertices, set $\phi_1(u') = \phi_1'(u)$ for each such degree-2 vertex u'. Thus, ϕ_1 assigns a colour to every vertex in $T_1[X']$, and by construction $\Delta_{T_1[X']}(\phi_1) = \Delta_{T_1'}(\phi_1')$.

In order to assign $\phi(v)$ to vertices v of T_1 not in $T_1[X']$, take any edge $e = uv$ in T_1 such that $\phi_1(u)$ has been assigned but $\phi_1(v)$ has not, and set $\phi_1(v) = \phi_1(u)$. After completing this process, we have that ϕ_1 assigns a colour to every vertex in T_1 (including its leaves) and $\Delta_{T_1}(\phi_1) = \Delta_{T_1'}(\phi_1')$, as required.

Now let the character χ be the restriction of ϕ_1 to X. Then by construction ϕ_1 is an extension of χ on X, whence

$$l_\chi(T_1) \leq \Delta_{T_1}(\phi_1) = l_{\chi'}(T_1').$$

Moreover, we must have that $l_\chi(T_1) \geq l_{\chi'}(T_1')$ and thus $l_\chi(T_1) = l_{\chi'}(T_1')$. Indeed, if $\Delta_{T_1}(\phi) < \Delta_{T_1}(\phi_1)$ for some extension ϕ of χ to T_1, then by considering the restriction of ϕ to $T_1[S]$, we can see that $l_{\chi'}(T_1') \leq \Delta_{T_1}(\phi) < \Delta_{T_1}(\phi_1)$, a contradiction as $\Delta_{T_1}(\phi_1) = \Delta_{T_1'}(\phi_1') = l_{\chi'}(T_1')$.

Next we show that $l_\chi(T_2) \geq l_{\chi'}(T_2')$. Consider any optimal extension ϕ_2 of χ to T_2, and take the restriction ϕ_2' of this function to $T_2' = T_2|_{X'}$. Then clearly $\Delta_{T_2'}(\phi_2') \leq \Delta_{T_2}(\phi_2)$ and therefore

$$l_{\chi'}(T_2') \leq \Delta_{T_2'}(\phi_2') \leq \Delta_{T_2}(\phi_2) = l_\chi(T_2).$$

Thus we have

$$d_{MP\chi}(T_1, T_2) \geq l_\chi(T_2) - l_\chi(T_1) \geq l_{\chi'}(T_2') - l_{\chi'}(T_1') = d_{MP\chi'}(T_1', T_2'). \quad \square$$

Theorem 3. *For any positive integer r, given an instance (T_1, T_2) of* DMP*, we can find in polynomial time a character χ such that*

$$1 \leq \frac{d_{MP}(T_1, T_2)}{d_{MP\chi}(T_1, T_2)} \leq (1 + 1/r)\alpha$$

where $\alpha = 560$. That is, DMP *has a constant factor approximation with approximation ratio $(1 + 1/r)\alpha$.*

Proof. First apply Reduction Rules 1 and 2 exhaustively, to derive an irreducible instance (T_1', T_2'). By Theorem 2, $d_{MP}(T_1', T_2') = d_{MP}(T_1, T_2)$. Let X' be the leaf set of this reduced instance. Now let k be the maximum integer such that $|X'| \geq \alpha k$, where $\alpha = 560$. If $k < r$, then we can determine a character χ' for which $d_{MP\chi'}(T_1', T_2') = d_{MP}(T_1', T_2')$ exactly in time $O(1.619^{\alpha r} \cdot poly(n))$ using the algorithm of [26]. Otherwise by Theorem 1, we can in polynomial time construct a character χ' on X' such that $d_{MP\chi'}(T_1', T_2') \geq k$. In either case, by Lemma 7 we can extend χ' to a character χ on X such that

$$d_{MP\chi}(T_1, T_2) \geq d_{MP\chi'}(T_1', T_2') \geq k.$$

We return χ.

It remains to show that

$$d_{MP}(T_1, T_2)/(1 + 1/r)\alpha \leq d_{MP\chi}(T_1, T_2) \leq d_{MP}(T_1, T_2)$$

from which the theorem follows. The second inequality is by definition of $d_{MP}(T_1, T_2)$. To see the first inequality: if $k < r$ then by construction

$$d_{MP}(T_1, T_2) = d_{MP}(T_1', T_2') = d_{MP\chi'}(T_1', T_2') \leq d_{MP\chi}(T_1, T_2).$$

So now assume that $k \geq r$, and so by construction $d_{MP\chi'}(T_1', T_2') \geq k \geq r$. As stated in the preliminaries, the number of taxa provides an upper bound on the d_{MP} of any instance. Thus, $d_{MP}(T_1', T_2') \leq |X'|$. By choice of k, we have $|X'| < \alpha(k + 1)$. Then we have

$$\begin{aligned} d_{MP}(T_1, T_2)/\alpha &= d_{MP}(T_1', T_2')/\alpha \\ &\leq |X'|/\alpha \\ &< \alpha(k+1)/\alpha = k + 1 \\ &\leq d_{MP\chi}(T_1, T_2) + 1 \\ &\leq (1 + 1/r) d_{MP\chi}(T_1, T_2) \end{aligned}$$

Thus $d_{MP}(T_1, T_2)/(1 + 1/r)\alpha \leq d_{MP\chi}(T_1, T_2)$, as required. $\square$

4.2. Bounding the distance between d_{TBR} and d_{MP}

Tree Bisection and Reconnection (TBR) distance, denoted d_{TBR}, is a distance measure defined on two unrooted binary phylogenetic trees T_1, T_2. It is defined as the minimum number of "TBR-moves" required to transform T_1 into T_2 (or vice-versa): it is a metric [1]. Informally, a TBR-move consists of deleting an edge of a tree and then reconnecting the two resulting components via a new edge. This definition is motivated by the way software for constructing phylogenetic trees heuristically navigates through tree space in search of better trees [36]. However, for algorithmic and analytical purposes d_{TBR} is most interesting because of its equivalence to the *agreement forest* abstraction. An agreement forest of T_1 and T_2 on the same set of taxa X is a partition of X into non-empty sets $S_1, S_2 \ldots, S_t$ called *blocks*, such that: (1) for each i, $T_1|_{S_i} = T_2|_{S_i}$; (2) $S_1, S_2, \ldots, S_t$ are spanning-disjoint in T_1 and in T_2. An *(unrooted) maximum agreement forest* is an agreement forest with a minimum number of blocks, and $d_{TBR}(T_1, T_2)$ is equal to this minimum, minus 1 [1]. A maximum agreement forest for the two trees in Fig. 1 consists of three blocks $\{a, b\}$, $\{f, g\}$ and $\{c, d, e\}$, so here d_{TBR} is 2.

The characterization of d_{TBR} via agreement forests is significant, because agreement forests have opened the door to a large number of positive FPT and approximation results in the phylogenetics literature, and they have also attracted attention

from outside phylogenetics. We refer to [41,16,39,14,10,35,4] for recent results. Moreover, a number of other problems have been shown to be FPT when parameterized by d_{TBR}, by leveraging properties of the d_{TBR} kernel [23] and/or showing that, via agreement forests, the treewidth of a certain auxiliary graph structure is bounded by a function of d_{TBR} [28] (see the next section). d_{TBR} is a lower bound on many phylogenetic dissimilarity measurements [28], which helps to prove FPT results for these larger parameters, but what about d_{MP}? It has previously been shown that $d_{MP}(T_1, T_2) \leq d_{TBR}(T_1, T_2)$ for any pair of trees T_1, T_2 [18,31]. However, the possibility remained that d_{MP} could be arbitrarily smaller than d_{TBR}, and this hinders our ability to bind d_{MP} to other phylogenetic parameters. Our contribution is to show that d_{MP} and d_{TBR} are in fact within a constant factor of each other: $d_{TBR}(T_1, T_2) \leq 2\alpha d_{MP}(T_1, T_2)$.

To show this, we use the fortunate fact that Reduction Rules 1 and 2, which we used to prove the kernel bound for DMP, preserve d_{TBR} as well as d_{MP} for d_{TBR}. The following theorem is, modulo a small modification, due to [1].

Theorem 4. *Let (T_1', T_2') be a pair of phylogenetic trees on X' derived from (T_1, T_2) by an application of Reduction Rule 1 or 2. Then*

$$d_{TBR}(T_1', T_2') = d_{TBR}(T_1, T_2).$$

Proof. Theorem 3.4 of [1] shows that d_{TBR} is preserved under reduction rules similar to Reduction Rules 1 and 2, except that common chains are reduced to length 3 instead of 4. For a pair of trees T_1, T_2 on X, let (T_1'', T_2'') with leaf set X'' be the instance derived from (T_1, T_2) by exhaustively applying these reduction rules. Also let (T_1', T_2') with leaf set X' be the instance derived from (T_1, T_2) by exhaustively applying Reduction Rules 1 and 2. Observe that we may assume $X'' \subseteq X' \subseteq X$, since any leaf deleted in an application of Reduction Rule 1 or 2 can also be deleted by an application of one of the reduction rules in [1]. Furthermore by Lemma 2.1 of [1], d_{TBR} distance is non-increasing on subtrees induced by subsets of X, which implies that

$$d_{TBR}(T_1'', T_2'') \leq d_{TBR}(T_1', T_2') \leq d_{TBR}(T_1, T_2).$$

As Theorem 3.4 of [1] states that $d_{TBR}(T_1'', T_2'') = d_{TBR}(T_1, T_2)$, the chain of inequalities becomes a chain of equalities and hence $d_{TBR}(T_1', T_2') = d_{TBR}(T_1, T_2)$. □

Theorem 5. *For any pair of phylogenetic trees T_1, T_2 such that $T_1 \neq T_2$, whence $d_{MP}(T_1, T_2) \geq 1$,*

$$1 \leq \frac{d_{TBR}(T_1, T_2)}{d_{MP}(T_1, T_2)} \leq 2\alpha.$$

Proof. Let (T_1', T_2') be the pair of trees derived from (T_1, T_2) by exhaustively applying Reduction Rules 1 and 2, and let X' be the leaf set of T_1' and T_2'. It is well-known that $d_{TBR}(T_1', T_2') \leq |X'| - 3$ [1]. Then by Theorems 1, 2 and 4,

$$\begin{aligned} d_{TBR}(T_1, T_2) = d_{TBR}(T_1', T_2') &< |X'| \\ &< \alpha(d_{MP}(T_1', T_2') + 1) \\ &\leq 2\alpha\, d_{MP}(T_1, T_2). \end{aligned}$$

Using $d_{MP}(T_1, T_2) \leq d_{TBR}(T_1, T_2)$ [31, Lemma 2.1], we have

$$d_{MP}(T_1, T_2) \leq d_{TBR}(T_1, T_2) \leq 2\alpha d_{MP}(T_1, T_2)$$

which, dividing by $d_{MP}(T_1, T_2)$, proves the theorem. □

4.3. The treewidth of the display graph

Let $G = (V, E)$ be an undirected graph. A *tree decomposition* of G consists of a multi-set of *bags*, $B = \{B_1, \ldots, B_t\}$ where each $B_i \subseteq V$, and a tree T whose nodes are in bijection with B, such that: (1) Every vertex $v \in V$ is in at least one bag; (2) for every edge $\{u, v\}$, at least one bag contains both u and v, and (3) for every vertex $v \in V$, the bags of T that contain v induce a connected subtree of T. The *width* of the tree decomposition is equal to the size of its largest bag, minus one, and the *treewidth* of G is the minimum width, ranging over all tree decompositions T of G [7]. Treewidth derives its importance in combinatorial optimization from the fact that many NP-hard problems on graphs become fixed parameter tractable when parameterized by the treewidth of the graph [6].

Given two phylogenetic trees T_1, T_2 on X, where $|X| \geq 3$, the *display graph* of T_1 and T_2, denoted $D(T_1, T_2)$, is the graph obtained by identifying the leaves of T_1 and T_2 with the same label. A sequence of articles have studied the treewidth of display graphs, expressed as a function of various phylogenetic parameters, and used this to prove FPT results for a number of NP-hard phylogenetics problems using Courcelle's Theorem [11,28,21] and explicit dynamic programming algorithms running over tree decompositions of the display graph [5]. However, the question remained whether the treewidth of the display graph, denoted by $tw(D(T_1, T_2))$ could be bounded by a function of $d_{MP}(T_1, T_2)$ [23].

The answer is emphatically yes: here we show, by leveraging the fact that d_{MP} and d_{TBR} are within a constant factor of each other, that the display graph has treewidth bounded by a linear function of $d_{MP}(T_1, T_2)$.

Theorem 6. *For two phylogenetic trees T_1, T_2 on X,*

$$tw(D(T_1, T_2)) \leq 2\alpha d_{MP}(T_1, T_2) + 2$$

Proof. It was shown in [28] that $tw(D(T_1, T_2)) \leq d_{TBR}(T_1, T_2) + 2$. As Theorem 5 shows $d_{TBR}(T_1, T_2) \leq 2\alpha d_{MP}(T_1, T_2)$ the theorem follows. □

Note that Theorem 7.2 of [27] shows an infinite family of trees where the treewidth of the display graph is 3 but d_{MP} is unbounded.

5. Conclusion

A natural question is how far the analysis can be tightened, or changed, to improve the existing bound on the size of the kernel. In any case, it can be shown that for these two reduction rules a bound smaller than $20k - 12$ is not possible. That is because the family of fully-reduced instances described in [24] have exactly $15k - 9$ taxa, where in this specific case $k = d_{TBR} = d_{MP}$. By replacing the length-3 chains with length-4 chains in this family we obtain the bound $20k - 12$. We expect that, *in practice*, the achieved reduction on realistic trees will be far superior to the bounds proven in this paper.

From the perspective of algorithm design it would be useful to design an explicit algorithm with FPT runtime that does not rely on kernelization; for example, by branching or by dynamic programming over an appropriately defined decomposition. Similarly, in the quest for small constant approximation factors it would be interesting to design polynomial-time approximation algorithms that do not rely on kernelization. It is unlikely that through kernelization we will be able to achieve such truly small constant ratios.

The precise relationship between d_{MP} and d_{TBR} remains intriguing. Although we have now established that they are within a constant factor of each other, we are still a long way from proving or disproving the conjecture that $d_{MP} \geq (1/2)d_{TBR}$ [23]. An infinite family of examples is known where $d_{MP} = (1/2)d_{TBR} + o(1)$ [31, Theorem 7.1], and small examples are known where $d_{MP} = (1/2)d_{TBR}$ (see e.g. Fig. 1, based on [23, Figure 5]), so $d_{MP} \geq (1/2)d_{TBR}$ would be the best possible bound.

On a slightly different note, recent publications have reduced the d_{TBR} kernel size from $28k$ to $15k - 9$ [24], and then to $11k - 9$ [25]. The $11k - 9$ kernel augments the two reduction rules discussed in this article, with five new reduction rules. Which of these new reduction rules work (possibly in a modified form) for d_{MP}, and how might this help us obtain a smaller linear kernel for d_{MP}?

Finally, we note that there are several slight variations of d_{MP} in the literature. These include the "asymmetric" version $d_{AMP}(T_1, T_2)$, defined as $\max_\chi(l_\chi(T_1) - l_\chi(T_2))$, in which T_1 is required to have the higher parsimony score, and the "restricted states" version $d^2_{MP}(T_1, T_2) := \max_\chi d_{MP\chi}(T_1, T_2)$, where the maximum is taken over all characters with at most 2 states [28,22]. Many of the results in this article will go through for $d_{AMP}(T_2, T_1)$, as the characters we construct consistently give a larger score to T_2. It is less obvious how our results impact on d^2_{MP}. In particular, it is not immediately clear whether the reduction rules described in [23] go through for d^2_{MP}, or how one would prove an analogue of Lemma 5 for d^2_{MP}. Relatedly, it is unclear how much smaller d^2_{MP} can be than d_{MP} itself. Specifically, how important are additional states when attempting to maximize the parsimony distance between trees? It is known that $7d_{MP} - 5$ states are sufficient to obtain a character that witnesses d_{MP} [8], but it is unclear what happens below this bound.

CRediT authorship contribution statement

Mark Jones: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Steven Kelk:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Leen Stougie:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Netherlands Organisation for Scientific Research (NWO) through Gravitation Programme Networks 024.002.003 and Klein Grant OCENW.KLEIN.125.

		d_2							
		2	3	4	5	6	7	8	9
d_1	2	–	34	43	52	61	70	79	88
	3	34	43	52	61	70	79	88	97
	4	43	52	61	70	79	88	97	106
	5	52	61	70	79	88	97	106	115
	6	61	70	79	88	97	106	115	124
	7	70	79	88	97	106	115	124	133
	8	79	88	97	106	115	124	133	142
	9	88	97	106	115	124	133	142	151

Fig. A.5. Values for $c = 9(d_1 + d_2) - 11$.

		d_2							
		2	3	4	5	6	7	8	9
d_1	2	–	14	9	8	7	6	6	6
	3	15	7	6	5	5	5	4	4
	4	10	6	5	4	4	4	4	4
	5	9	5	5	4	4	4	4	4
	6	8	5	4	4	4	4	3	3
	7	7	5	4	4	4	3	3	3
	8	7	5	4	4	4	3	3	3
	9	7	5	4	4	3	3	3	3

Fig. A.6. Values for $t' = \lceil \frac{2d_1d_2+d_1}{d_1d_2-d_1-d_2} \rceil$.

		d_2							
		2	3	4	5	6	7	8	9
d_1	2	–	952	774	832	854	840	948	1056
	3	1020	602	624	610	700	790	704	776
	4	860	624	610	**560**	632	704	776	848
	5	936	610	700	632	704	776	848	920
	6	976	700	632	704	776	848	690	744
	7	980	790	704	776	848	690	744	798
	8	1106	880	776	848	920	744	798	852
	9	1232	970	848	920	744	798	852	906

Fig. A.7. Values for $\alpha = 2 \cdot (9(d_1 + d_2) - 11) \cdot \lceil \frac{2d_1d_2+d_1}{d_1d_2-d_1-d_2} \rceil$. Observe that the minimum is achieved at $d_1 = 4$, $d_2 = 5$.

Appendix A. Finding optimal d_1, d_2

For the sake of completeness, we here argue that the choice of $d_1 = 4, d_2 = 5$ gives the minimum value of $\alpha = 2 \cdot (9(d_1 + d_2) - 11) \cdot \lceil \frac{2d_1d_2+d_1}{d_1d_2-d_1-d_2} \rceil$ in Theorem 1. Let $c = 9(d_1 + d_2) - 11$ and $t' = \lceil \frac{2d_1d_2+d_1}{d_1d_2-d_1-d_2} \rceil$, so that $\alpha = 2ct'$. For $d_1 = 4, d_2 = 5$, we have $c = 81 - 11 = 70$ and $t' = \lceil \frac{44}{11} \rceil = 4$, and so $\alpha = 2 \cdot 70 \cdot 4 = 560$. Figs. A.5, A.6 and A.7 gives the possible values of c, t' and α respectively, for d_1, d_2 taking values between 2 and 9 (recall that d_1, d_2 must be at least 2, as Lemma 6 requires $d_1d_2 - d_1 - d_2 > 0$).

By inspection of Fig. A.7, it is easy to see that the minimum possible value of α for $2 \leq d_1, d_2 \leq 9$ is 560. For larger values of d_1, d_2, we argue as follows: Observe that $t' = \lceil \frac{2d_1d_2+d_1}{d_1d_2-d_1-d_2} \rceil$ is at least 3 for any d_1, d_2, as

$$\frac{2d_1d_2 + d_1}{d_1d_2 - d_1 - d_2} > \frac{2d_1d_2}{d_1d_2} = 2.$$

If one of d_1, d_2 is at least 10, then $c = 9(d_1 + d_2) - 11 \geq 9(10 + 2) - 11 = 97$. But then for such values we would have $\alpha = 2ct' \geq 2 \cdot 97 \cdot 3 = 582$. Thus, the smallest value of α is in fact 560, achieved for $d_1 = 4, d_2 = 5$.

References

[1] B.L. Allen, M. Steel, Subtree transfer operations and their induced metrics on evolutionary trees, Ann. Comb. 5 (2001) 1–15.
[2] N. Alon, B. Chor, F. Pardi, A. Rapoport, Approximate maximum parsimony and ancestral maximum likelihood, IEEE/ACM Trans. Comput. Biol. Bioinform. 7 (1) (January 2010) 183–187.
[3] answer Anonymous, How many vertices of degree 3 or more can a tree have at most?, Mathematics Stack Exchange, https://math.stackexchange.com/q/388948 (version: 2013-05-12).
[4] R. Atkins, C. McDiarmid, Extremal distances for subtree transfer operations in binary trees, Ann. Comb. 23 (1) (2019) 1–26.
[5] J. Baste, C. Paul, I. Sau, C. Scornavacca, Efficient FPT algorithms for (strict) compatibility of unrooted phylogenetic trees, Bull. Math. Biol. 79 (4) (2017) 920–938.
[6] H.L. Bodlaender, A tourist guide through treewidth, Acta Cybern. 11 (1–2) (1994) 1.
[7] H.L. Bodlaender, A linear-time algorithm for finding tree-decompositions of small treewidth, SIAM J. Comput. 25 (1996) 1305–1317.

[8] O. Boes, M. Fischer, S. Kelk, A linear bound on the number of states in optimal convex characters for maximum parsimony distance, IEEE/ACM Trans. Comput. Biol. Bioinform. 14 (2) (2016) 472–477.
[9] M.L. Bonet, K. St John, On the complexity of uSPR distance, IEEE/ACM Trans. Comput. Biol. Bioinform. 7 (3) (2010) 572–576.
[10] M. Bordewich, C. Scornavacca, N. Tokac, M. Weller, On the fixed parameter tractability of agreement-based phylogenetic distances, J. Math. Biol. 74 (1–2) (2017) 239–257.
[11] D. Bryant, J. Lagergren, Compatibility of unrooted phylogenetic trees is FPT, Theor. Comput. Sci. 351 (3) (2006) 296–302.
[12] L. Bulteau, M. Weller, Parameterized algorithms in bioinformatics: an overview, Algorithms 12 (12) (2019) 256.
[13] J. Chen, J-H. Fan, S-H. Sze, Parameterized and approximation algorithms for maximum agreement forest in multifurcating trees, Theor. Comput. Sci. 562 (2015) 496–512.
[14] J. Chen, F. Shi, J. Wang, Approximating maximum agreement forest on multiple binary trees, Algorithmica 76 (4) (2016) 867–889.
[15] M. Cygan, F. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, S. Saurabh, Parameterized Algorithms, 1st edition, Springer Publishing Company, Incorporated, 2015.
[16] R. Downey, M. Fellows, Fundamentals of Parameterized Complexity, vol. 4, Springer, 2013.
[17] J. Felsenstein, Inferring Phylogenies, Sinauer Associates, Incorporated, 2004.
[18] M. Fischer, S. Kelk, On the maximum parsimony distance between phylogenetic trees, Ann. Comb. 20 (1) (2016) 87–113.
[19] W. Fitch, Toward defining the course of evolution: minimum change for a specific tree topology, Syst. Biol. 20 (4) (1971) 406–416.
[20] D. Huson, R. Rupp, C. Scornavacca, Phylogenetic Networks: Concepts, Algorithms and Applications, Cambridge University Press, 2011.
[21] R. Janssen, M. Jones, S. Kelk, G. Stamoulis, T. Wu, Treewidth of display graphs: bounds, brambles and applications, J. Graph Algorithms Appl. 23 (4) (2019).
[22] S. Kelk, M. Fischer, On the complexity of computing MP distance between binary phylogenetic trees, Ann. Comb. 21 (2017) 573–604.
[23] S. Kelk, M. Fischer, V. Moulton, T. Wu, Reduction rules for the maximum parsimony distance on phylogenetic trees, Theor. Comput. Sci. 646 (2016) 1–15.
[24] S. Kelk, S. Linz, A tight kernel for computing the tree bisection and reconnection distance between two phylogenetic trees, SIAM J. Discrete Math. 33 (3) (2019) 1556–1574.
[25] S. Kelk, S. Linz, New reduction rules for the tree bisection and reconnection distance, Ann. Comb. 24 (2020) 475–502.
[26] S. Kelk, G. Stamoulis, A note on convex characters, Fibonacci numbers and exponential-time algorithms, Adv. Appl. Math. 84 (2017) 34–46.
[27] S. Kelk, G. Stamoulis, T. Wu, Treewidth distance on phylogenetic trees, Theor. Comput. Sci. 731 (2018) 99–117.
[28] S. Kelk, L.J.J. van Iersel, C. Scornavacca, M. Weller, Phylogenetic incongruence through the lens of monadic second order logic, J. Graph Algorithms Appl. 20 (2) (2016) 189–215.
[29] F. Liers, A. Martin, S. Pape, Binary Steiner trees: structural results and an exact solution approach, Discrete Optim. 21 (2016) 85–117.
[30] S. Moran, S. Snir, Convex recolorings of strings and trees: definitions, hardness results and algorithms, J. Comput. Syst. Sci. 74 (5) (2008) 850–869.
[31] V. Moulton, T. Wu, A parsimony-based metric for phylogenetic trees, Adv. Appl. Math. 66 (2015) 22–45.
[32] L. Nakhleh, Computational approaches to species phylogeny inference and gene tree reconciliation, Trends Ecol. Evol. 28 (12) (2013) 719–728.
[33] D. Robinson, L. Foulds, Comparison of phylogenetic trees, Math. Biosci. 53 (1–2) (1981) 131–147.
[34] C. Semple, M. Steel, Phylogenetics, Oxford University Press, 2003.
[35] F. Shi, J. Chen, Q. Feng, J. Wang, A parameterized algorithm for the maximum agreement forest problem on multiple rooted multifurcating trees, J. Comput. Syst. Sci. 97 (2018) 28–44.
[36] K. St John, The shape of phylogenetic treespace, Syst. Biol. 66 (1) (2017) e83–e94.
[37] M. Steel, Phylogeny: Discrete and Random Processes in Evolution, SIAM, 2016.
[38] L. van Iersel, M. Jones, S. Kelk, A third strike against perfect phylogeny, Syst. Biol. 68 (5) (2019) 814–827.
[39] L. van Iersel, S. Kelk, N. Lekic, C. Whidden, N. Zeh, Hybridization number on three rooted binary trees is EPT, SIAM J. Discrete Math. 30 (3) (2016) 1607–1631.
[40] T. Warnow, Computational Phylogenetics: An Introduction to Designing Methods for Phylogeny Estimation, Cambridge University Press, 2017.
[41] C. Whidden, R.G. Beiko, N. Zeh, Fixed-parameter algorithms for maximum agreement forests, SIAM J. Comput. 42 (4) (2013) 1431–1466.
[42] C. Whidden, N. Zeh, R.G. Beiko, Supertrees based on the subtree prune-and-regraft distance, Syst. Biol. 63 (4) (2014) 566–581.