Bounding the SNPR Distance Between Two Tree-Child Networks Using Generalised Agreement Forests

Steven Kelk[a] Simone Linz[b] Charles Semple[c]

Submitted: Mar 18, 2025; Accepted: Aug 15, 2025; Published: Sep 19, 2025

Abstract

Agreement forests continue to play a central role in the comparison of phylogenetic trees since their introduction more than 25 years ago. More specifically, they are used to characterise several distances that are based on tree rearrangement operations and related quantifiers of dissimilarity between phylogenetic trees. In addition, the concept of agreement forests continues to underlie most advancements in the development of algorithms that exactly compute the aforementioned measures. In this paper, we introduce agreement digraphs, a concept that generalises agreement forests for two phylogenetic trees to two phylogenetic networks. Analogous to the way in which agreement forests compute the subtree prune and regraft distance between two phylogenetic trees but inherently more complex, we then use agreement digraphs to bound the subnet prune and regraft distance between two tree-child networks from above and below and show that our bounds are tight.

Mathematics Subject Classifications: 05C20, 92D15

1 Introduction

Phylogenetic trees and networks play an important role in areas of biology that investigate the relationships between biological entities such as species, viruses, and cells. A central task in the analysis of phylogenetic trees and networks is the quantification of the dissimilarity between them. Distances between phylogenetic trees that provide a measure of dissimilarity can be broadly classified into distances that are based on tree vector representations and those based on tree rearrangement operations [39]. While many of the

[a]Department of Advanced Computing Science, Maastricht University, Maastricht, The Netherlands (steven.kelk@maastrichtuniversity.nl).
[b]School of Computer Science, University of Auckland, Auckland, New Zealand (s.linz@auckland.ac.nz).
[c]School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand (charles.semple@canterbury.ac.nz).

former distances can be computed in polynomial time, the latter distances are typically NP-hard to compute. On the positive side and as summarised in Semple and Steel [40], tree distances that are based on rearrangement operations provide a framework to explore and traverse spaces of phylogenetic trees since any tree in a given space can be transformed into any other tree of the same space by a finite number of operations such as nearest neighbour interchange (NNI), subtree prune and regraft (SPR), rooted subtree prune and regraft (rSPR), and tree bisection and reconnection (TBR). In particular, the length of a shortest path between two phylogenetic trees in a given space of trees equals the distance between the two trees under the rearrangement operation that underlies the space.

Computing distances that are based on tree rearrangement operations and related dissimilarity measures such as the minimum hybridisation number for two phylogenetic trees [5] remains an active area of research (e.g. [26, 38, 41]) despite the NP-hardness of the associated optimisation problems. Indeed, recent algorithmic progress facilitates computations that exactly calculate the aforementioned measures for data sets of remarkable size [43, 44]. Notably, the concept of agreement forests, which was first introduced by Hein et al. [16], underpins almost all mathematical and algorithmic advances in this area of research. Intuitively, an agreement forest of two binary phylogenetic trees $\mathcal{T}$ and $\mathcal{T}'$ is a decomposition of $\mathcal{T}$ and $\mathcal{T}'$ into smaller and non-overlapping subtrees that have the same topology in $\mathcal{T}$ and $\mathcal{T}'$. Since the introduction of agreement forests almost 30 years ago, different variants of agreement forests have been used to characterise the rSPR distance between two rooted binary phylogenetic trees, the TBR distance between two unrooted binary phylogenetic trees, and the minimum hybridisation number of two rooted binary phylogenetic trees, as well as to establish related NP-hardness results [1, 5, 6]. Subsequent work has focussed on generalising agreement forests to collections of phylogenetic trees of arbitrary size that are not necessarily binary and on the development of fixed-parameter tractable and approximation algorithms [9, 25, 33, 38, 45]. Additional developments in the context of agreement forests include a generalisation of agreement forests to relaxed agreement forests [3] and the exploitation of agreement forests to establish extremal results on the SPR, rSPR, and TBR distances [4, 11]. Part of the success of agreement forests is due to the fact that they replace the computation of a measure of dissimilarity between two phylogenetic trees $\mathcal{T}$ and $\mathcal{T}'$ with the more static computation of an agreement forest such that the sought-after measure equates to the size of an optimal agreement forest for $\mathcal{T}$ and $\mathcal{T}'$. Moreover, agreement forests enable rigorous mathematical arguments that operate only on $\mathcal{T}$ and $\mathcal{T}'$ without knowing the topology of any intermediate tree that lies on a shortest path between $\mathcal{T}$ and $\mathcal{T}'$ in an associated space of trees.

Since phylogenetic trees are somewhat limited in the type of biological processes that they can represent, rooted phylogenetic networks are increasingly being adopted to represent evolutionary relationships between biological entities whose past does not only include divergence events such as speciation but also convergence events such as lateral gene transfer or hybridisation. Inspired by tree rearrangement operations, several network rearrangement operations have recently been developed. For example, the subnet prune and regraft (SNPR) operation generalises rSPR to two rooted binary phylogenetic networks [7].

An SNPR operation either adds or deletes a reticulation edge (i.e., an edge that is directed into a vertex of in-degree two), or prunes and regrafts a subnetwork in the spirit of rSPR. Since the introduction of SNPR, the operation has been implemented in the Python package PhyloX [20], used in studies that, for example, reconstruct reassortment networks and involve a search through the space of rooted phylogenetic networks [35, 36, 37], and analysed mathematically with regards to the neighbourhood size [27] and properties of shortest length SNPR paths between two rooted phylogenetic networks [31]. Other network rearrangement operations for rooted phylogenetic networks [13, 15, 19, 21], unrooted phylogenetic networks [14, 17, 18, 22], and semi-directed networks [2, 34] have also been developed and analysed. Like SNPR, all such operations generalise a rearrangement operation for phylogenetic trees to phylogenetic networks. For excellent summaries of rearrangement operations for phylogenetic networks, we refer the reader to two PhD theses [23, 30]. Although a slightly modified framework of agreement forests can be used to compute the SNPR distance between a rooted binary phylogenetic tree and a rooted binary phylogenetic network [31], the question of how to compute the distances that result from network rearrangement operations between two arbitrary binary phylogenetic networks remains open.

In this paper, we generalise agreement forests for two rooted binary phylogenetic trees to agreement digraphs for two rooted binary phylogenetic networks $\mathcal{N}$ and $\mathcal{N}'$ that capture the commonalities between them. Focussing on the class of tree-child networks [8] and using this novel framework of agreement digraphs and their extensions (formal definitions are given Section 2), we bound the SNPR distance $d_{\mathrm{tc}}(\mathcal{N},\mathcal{N}')$ between two tree-child networks $\mathcal{N}$ and $\mathcal{N}'$ from above and below, where not only $\mathcal{N}$ and $\mathcal{N}'$ are tree-child but also each intermediate network in an associated sequence. Both bounds are tight and within small constant factors of the minimum number $m_{\mathrm{tc}}(\mathcal{N},\mathcal{N}')$ of edges in $\mathcal{N}$ and $\mathcal{N}'$ that are not contained in an embedding of the agreement digraph and an extension, where the minimum is taken over all agreement digraphs for $\mathcal{N}$ and $\mathcal{N}'$ and their extensions. More specifically, the main result of this paper is the following theorem.

Theorem 1. *Let $\mathcal{N}$ and $\mathcal{N}'$ be two binary tree-child networks on X. Then*

$$\frac{1}{2}m_{\mathrm{tc}}(\mathcal{N},\mathcal{N}') \leqslant d_{\mathrm{tc}}(\mathcal{N},\mathcal{N}') \leqslant m_{\mathrm{tc}}(\mathcal{N},\mathcal{N}').$$

The problem of finding an extension of an agreement digraph that optimally captures the commonalities between two rooted binary phylogenetic networks $\mathcal{N}$ and $\mathcal{N}'$ may, at first sight, appear to be related to the problem of finding a maximum agreement subnetwork that $\mathcal{N}$ and $\mathcal{N}'$ have in common [10, 24, 42]. However, upon further inspection, it becomes clear that the two problems are fundamentally different since a maximum agreement subnetwork for $\mathcal{N}$ and $\mathcal{N}'$ is not necessarily a component of an optimal agreement digraph for $\mathcal{N}$ and $\mathcal{N}'$. On the other hand, our work is related to that of Klawitter [28, 29], who has developed an alternative generalisation of agreement forests for two rooted binary phylogenetic networks and for two unrooted binary phylogenetic networks. His generalisation for rooted networks gives rise to collections of agreement

subgraphs that, in comparison with our generalisation, may have unlabelled leaves of in-degree one or two and that consequently do not resemble phylogenetic networks. In the same paper, Klawitter established bounds on the SNPR distance $d^*_{\mathrm{SNPR}}(\mathcal{N}, \mathcal{N}')$ between two rooted binary phylogenetic networks $\mathcal{N}$ and $\mathcal{N}'$ in terms of collections of agreement subgraphs whose number of unlabelled vertices of degree one is minimised. Without going into detail, this minimum number is referred to as $d_{\mathrm{AD}}(\mathcal{N}, \mathcal{N}')$. In particular, Klawitter established the following theorem.

Theorem 2. *[28, Corollary 5.5] Let $\mathcal{N}$ and $\mathcal{N}'$ be two rooted binary phylogenetic networks on X. Then*

$$d_{\mathrm{AD}}(\mathcal{N}, \mathcal{N}') \leqslant d^*_{\mathrm{SNPR}}(\mathcal{N}, \mathcal{N}') \leqslant 6d_{\mathrm{AD}}(\mathcal{N}, \mathcal{N}').$$

While Theorem 2 applies to all rooted binary phylogenetic networks, it remains unknown whether or not the bounds are tight. Comparing Theorems 1 and 2, we note that $d^*_{\mathrm{SNPR}}(\mathcal{N}, \mathcal{N}')$ in Theorem 2 refers to the minimum number of SNPR operations that are necessary to transform $\mathcal{N}$ into $\mathcal{N}'$, while the quantity $d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$ in Theorem 1 does not only count SNPR operations but, additionally, weights them. More precisely, each SNPR operation that adds or deletes a reticulation edge is weighted one and each SNPR operation that prunes and regrafts a subnetwork is weighted two. Hence, $d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$ equates to the minimum sum of weights of SNPR operations that are needed to transform $\mathcal{N}$ into $\mathcal{N}'$. We discuss this further in the last section of the paper. Lastly, we note that although Klawitter's generalisation and our definition of an agreement digraph differ, both definitions generalise agreement forests in the sense that, when applied to two rooted binary phylogenetic trees, they can be used to exactly compute their rSPR distance ([28, Proposition 4.2] and Proposition 15 of the present paper).

The paper is organised as follows. In Section 2, we present basic notation and terminology for rooted phylogenetic networks. This is followed by an introduction of the new concepts of phylogenetic digraphs, which is the main definition building up to agreement digraphs, and their extensions in Section 3. Section 4 establishes several basic properties of extensions. Subsequently, in Section 5, we introduce the tree-child SNPR distance between two tree-child networks $\mathcal{N}$ and $\mathcal{N}'$ and a maximum agreement tree-child digraph for $\mathcal{N}$ and $\mathcal{N}$. In Section 6, we establish Theorem 1 and show that our bounds are tight before we finish with some concluding remarks in Section 7.

2 Preliminaries

This section provides notation and terminology that is used in the remainder of this paper. Throughout the paper, X denotes a non-empty finite set. It is also worth noting that, except for rooted phylogenetic networks, the graphs that we consider in this paper are not necessarily connected. We start by introducing a broad class of directed acyclic graphs and several definitions that apply to this class. Subsequent definitions in this and the next section consider subclasses of these directed acyclic graphs.

Directed acyclic graphs. Let D be a directed acyclic graph. We allow parallel edges in D and note that D may have several vertices with in-degree zero. Furthermore, the undirected graph that underlies D may contain more than one connected component. Let V_D denote the vertex set of D, and let E_D denote the edge set of D. We say that a vertex v in V_D is a *tree vertex* if v has in-degree one and out-degree one or two, and that v is a *reticulation* if v has in-degree two and out-degree one. Furthermore, an edge (u, v) in D is a *reticulation edge* if v is a reticulation and, otherwise, (u, v) is a *tree edge*. Lastly, for two vertices u and v in D, we say that u is a *parent* of v and v is a *child* of u precisely if there is an edge (u, v) in D.

Phylogenetic networks. Rooted phylogenetic networks generalise rooted phylogenetic trees to digraphs with underlying (but no directed) cycles. They allow vertices with in-degree greater than one, which represent non-treelike events such as hybridisation, lateral gene transfer, or recombination. Formally, a *rooted binary phylogenetic network on* X is a connected directed acyclic graph with a single vertex of in-degree zero that satisfies the following properties:

(i) the unique root ρ has in-degree zero and out-degree one,

(ii) vertices with out-degree zero have in-degree one, and the set of vertices with out-degree zero is X, and

(iii) all other vertices have either in-degree one and out-degree two, or in-degree two and out-degree one.

The vertices of $\mathcal{N}$ of out-degree zero are called *leaves*, and so X is referred to as the *leaf set* $\mathcal{L}(\mathcal{N})$ of $\mathcal{N}$. In keeping with the literature on distances between two phylogenetic networks, we allow parallel edges in rooted binary phylogenetic networks. Since all phylogenetic networks in this paper are rooted and binary, we simply refer to a rooted binary phylogenetic network on X as a *phylogenetic network on* X. Now, let $\mathcal{N}$ be a phylogenetic network on X. The vertices of out-degree zero, that is the elements in X, are called *leaves* and X is referred to as the *leaf set* of $\mathcal{N}$. If a phylogenetic network $\mathcal{N}$ has no reticulations, we call $\mathcal{N}$ a *phylogenetic* X*-tree*. Moreover, if $\mathcal{N}$ is a phylogenetic X-tree and X contains exactly three elements, say $X = \{a, b, c\}$, then we refer to $\mathcal{N}$ as a *triple* if the underlying path joining the root of $\mathcal{N}$ and c is vertex-disjoint from that joining a and b in which case, (a, b, c) or, equivalently, (b, a, c) denotes this triple.

Now, let $\mathcal{N}$ and $\mathcal{N}'$ be two phylogenetic networks on X with vertex and edge sets V and E, and V' and E', respectively. We say that $\mathcal{N}$ is *isomorphic* to $\mathcal{N}'$ if there is a bijection $\varphi : V \to V'$ such that $\varphi(x) = x$ for all $x \in X$, and $(u, v) \in E$ if and only if $(\varphi(u), \varphi(v)) \in E'$ for all $u, v \in V$. If $\mathcal{N}$ and $\mathcal{N}'$ are isomorphic, then we write $\mathcal{N} \cong \mathcal{N}'$.

Tree-child networks. Let $\mathcal{N}$ be a phylogenetic network on X. We say that $\mathcal{N}$ is *tree-child* if each non-leaf vertex has a child that is a tree vertex or a leaf. Moreover, we say that $\mathcal{N}$ contains a *stack* if there exist two reticulations that are joined by an edge and that $\mathcal{N}$ contains a pair of *sibling reticulations* if there exist two reticulations that have a

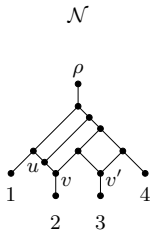

Figure 1: A phylogenetic network on $X = \{1, 2, 3, 4\}$ that is not tree-child because u and v form a stack, and v and v' are a pair of sibling reticulations.

common parent. A phylogenetic network that is not tree-child is shown in Figure 1. In studying the mathematics that underlies phylogenetic networks, tree-child networks have been proven to be particularly successful because of their combinatorial properties that are often exploited to gain traction in establishing mathematical results. At the same time, these properties are not overly restrictive from a structural perspective in comparison to level-1 networks for example whose underlying cycles are pairwise vertex disjoint. For an overview of classes of phylogenetic networks, we refer the reader to Kong et al. [32].

The following well-known equivalence follows from the definition of a tree-child network and will be freely used throughout the remainder of the paper.

Lemma 3. *Let $\mathcal{N}$ be a phylogenetic network. Then $\mathcal{N}$ is tree-child if and only if it has no stack, no pair of sibling reticulations, and no pair of parallel edges.*

The next lemma was established by Döcker et al. [12, Lemma 7] and shows that the deletion of a reticulation edge of a tree-child network results in another tree-child network.

Lemma 4. *Let $\mathcal{N}$ be a tree-child network on X, and let $e = (u, v)$ be a reticulation edge of $\mathcal{N}$. Then the network obtained from $\mathcal{N}$ by deleting e and suppressing u and v is a tree-child network on X.*

3 Phylogenetic digraphs and their extensions

In this section, we provide formal definitions of the concepts of phylogenetic digraphs and their extensions. As we will see in Section 5, these definitions generalise agreement forests for two phylogenetic trees to two phylogenetic networks. We start by providing some high-level ideas that may guide the reader in developing some intuition before providing formal definitions. Essentially, a phylogenetic digraph $\mathcal{D}$ of a phylogenetic network $\mathcal{N}$ on X and with root ρ is a collection of directed acyclic graphs that each contain at least one element in X with the exception that ρ may be a singleton in $\mathcal{D}$, whose vertices of out-degree zero are bijectively labelled with the elements in X, and for which there is a vertex-disjoint embedding of its components in $\mathcal{N}$. Let $\mathcal{M}$ be such an embedding of $\mathcal{D}$ in $\mathcal{N}$, and let v be a vertex of $\mathcal{M}$ that either has in-degree zero, or is a reticulation of $\mathcal{N}$ and has in-degree one and out-degree one. We obtain an extension $\mathcal{R}$ of $\mathcal{M}$ by starting at v and extending $\mathcal{M}$ towards the root by adding edges of $\mathcal{N}$ in a

certain algorithmic way and then repeating this process for all such vertices of $\mathcal{M}$. As suggested by the phrase *extension of* $\mathcal{M}$, $\mathcal{R}$ contains all edges of $\mathcal{M}$. Although $\mathcal{D}$ may have several embeddings in $\mathcal{N}$, each embedding is anchored at the leaves of $\mathcal{N}$ due to the requirement that the leaves of $\mathcal{D}$ bijectively map to the elements in X. As we will see in Section 4, for the purpose of computing the minimum number of edges in $\mathcal{N}$ that are not contained in an extension relative to a given phylogenetic digraph $\mathcal{D}$ of $\mathcal{N}$, where the minimum is taken over all embeddings of $\mathcal{D}$ in $\mathcal{N}$ and their extensions, it is sufficient to consider only a single extension of $\mathcal{D}$.

Phylogenetic digraphs. Let D be a connected directed acyclic graph, and let Y be a finite set. We say that D is a *leaf-labelled acyclic digraph on* Y if one of the following applies:

(i) $|Y| = 0$ and D is the isolated vertex ρ,

(ii) $|Y| = 1$ and D is the isolated vertex labelled with the element in Y, or

(iii) $|Y| \geqslant 1$, D has at most one vertex of in-degree zero and out-degree one, in which case this vertex is ρ, the leaves of D have in-degree one and out-degree zero and are bijectively labelled with the elements in Y, and all other vertices of D have in-degree zero and out-degree two, in-degree one and out-degree two, or in-degree two and out-degree one.

Similar to the leaf set of a phylogenetic network, we refer to Y as the *leaf set* of D. Furthermore, $\mathcal{L}(D)$ denotes the leaf set of D. In contrast to a phylogenetic network, a leaf-labelled acyclic digraph D may have more than one vertex with in-degree zero. Moreover, for a vertex w in D that has two parents v and v', there does not necessarily exist a vertex u such that there are edge-disjoint directed paths from u to v and from u to v' in D.

Let $\mathcal{N}$ be a phylogenetic network on X with root ρ, and let D be a leaf-labelled acyclic digraph on Y with $Y \subseteq X$. Recall that D may or may not contain a vertex ρ with in-degree zero and out-degree one. We say that $\mathcal{N}$ *displays* D if there exists a subgraph of $\mathcal{N}$ that is isomorphic to D up to suppressing vertices with in-degree one and out-degree one, in which case we call the subgraph an *embedding* M of D in $\mathcal{N}$ and view the edge set of M as a subset of the edge set of $\mathcal{N}$. More generally, for a collection $\mathcal{D} = \{D_1, D_2, \ldots, D_k\}$ of leaf-labelled acyclic digraphs, we say that $\mathcal{N}$ *displays* $\mathcal{D}$ if there exists an embedding M_i of D_i in $\mathcal{N}$ for each $i \in \{1, 2, \ldots, k\}$ such that M_j and $M_{j'}$ are vertex disjoint for all distinct $j, j' \in \{1, 2, \ldots, k\}$, in which case we refer to $\mathcal{M} = \{M_1, M_2, \ldots, M_k\}$ as an *embedding* of $\mathcal{D}$ in $\mathcal{N}$. Now let $\mathcal{M}$ be an embedding of $\mathcal{D}$ in $\mathcal{N}$. Recalling that we allow tree vertices in $\mathcal{M}$ to have in-degree and out-degree one, we say that $\mathcal{M}$ is *tree-child* if each non-leaf vertex of $\mathcal{M}$ has a child that is a tree vertex or a leaf.

Let $\mathcal{N}$ be a phylogenetic network on X with root ρ. Let $\mathcal{D} = \{D_\rho, D_1, D_2, \ldots, D_k\}$ be a collection of leaf-labelled acyclic digraphs. Then $\mathcal{D}$ is called a *phylogenetic digraph* of $\mathcal{N}$ if the following three properties are satisfied:

(i) the leaf sets $\mathcal{L}(D_\rho), \mathcal{L}(D_1), \mathcal{L}(D_2), \ldots, \mathcal{L}(D_k)$ partition X and D_ρ is the only element in $\mathcal{D}$ that contains ρ,

(ii) ρ is either an isolated vertex in $\mathcal{D}$ or the unique vertex in $\mathcal{D}$ with in-degree zero and out-degree one, and

(iii) there exists an embedding $\mathcal{M} = \{M_\rho, M_1, M_2, \ldots, M_k\}$ of $\mathcal{D}$ in $\mathcal{N}$.

Lastly, a phylogenetic digraph $\mathcal{D}$ of $\mathcal{N}$ is called a *tree-child digraph* of $\mathcal{N}$ if each non-leaf vertex of $\mathcal{D}$ has a child that is a leaf or a tree vertex.

Extensions and root extensions. Let $\mathcal{N}$ be a phylogenetic network on X. Furthermore, let $\mathcal{M} = \{M_\rho, M_1, M_2, \ldots, M_k\}$ be an embedding of a phylogenetic digraph $\mathcal{D} = \{D_\rho, D_1, D_2, \ldots, D_k\}$ of $\mathcal{N}$. We obtain an *extension* $\mathcal{R}$ of $\mathcal{D}$ in $\mathcal{N}$ from $\mathcal{M}$ by initially setting $\mathcal{R} = \mathcal{M}$ and then repeatedly applying one of the following two operations until no further such operation is possible:

(E1) For a vertex v of $\mathcal{R}$ with in-degree zero, add (u, v) to $\mathcal{R}$ if $u \notin \mathcal{R}$.

(E2) For a vertex v of $\mathcal{R}$ with in-degree one and out-degree one and v being a reticulation in $\mathcal{N}$, add (u, v) to $\mathcal{R}$ if $u \notin \mathcal{R}$.

By construction, observe that $\mathcal{R}$ contains exactly $k + 1$ connected components and that there is a natural bijection between these components and the components in $\mathcal{D}$. We therefore set $\mathcal{R} = \{R_\rho, R_1, R_2, \ldots, R_k\}$ and call R_i an *extension* of D_i in $\mathcal{N}$ for each $i \in \{\rho, 1, 2, \ldots, k\}$. It follows from the construction of $\mathcal{R}$ that there is no vertex of out-degree two in $\mathcal{R}$ that is not also a vertex of out-degree two in $\mathcal{M}$. Moreover, by construction, any underlying cycle in $\mathcal{R}$ is also an underlying cycle in $\mathcal{M}$. Lastly, we define $\mathcal{R}$ to be *tree-child* precisely if $\mathcal{M}$ is tree-child, and refer to $\mathcal{M}$ as the embedding of $\mathcal{D}$ that *underlies* $\mathcal{R}$.

We next introduce a special type of extension. We call an extension $\mathcal{R}$ of $\mathcal{D}$ in $\mathcal{N}$ a *root extension* of $\mathcal{D}$ in $\mathcal{N}$ if it can be obtained from $\mathcal{M}$ by initially setting $\mathcal{R} = \mathcal{M}$ and then repeatedly applying (E1) only until no further such operation is possible. Now let $\mathcal{R} = \{R_\rho, R_1, R_2, \ldots, R_k\}$ be a root extension of $\mathcal{D}$ in $\mathcal{N}$. Similar to the terminology for an extension, we call R_i a *root extension* of D_i in $\mathcal{N}$ for each $i \in \{\rho, 1, 2, \ldots, k\}$. Let r be a vertex of in-degree zero and out-degree zero or two in $\mathcal{D}$, and let P be the unique maximal length directed path in $\mathcal{R}$ that starts at a vertex u of in-degree zero and ends at r. We refer to P as the *root path* of r. Note that P may have no edge in which case $u = r$. If $u \neq r$, then u has in-degree zero and out-degree one in $\mathcal{R}$. Figure 2 illustrates the concepts of phylogenetic digraphs, extensions, and root extensions.

To ease reading throughout the remainder of the paper, we often consider a phylogenetic network $\mathcal{N}$, a phylogenetic digraph $\mathcal{D}$ of $\mathcal{N}$, and an extension $\mathcal{R}$ of $\mathcal{D}$ in $\mathcal{N}$. In this case, we view the vertex and edge set of $\mathcal{R}$ (as well as the vertex and edge set of any embedding of $\mathcal{D}$ in $\mathcal{N}$) as a subset of the vertex and edge set of $\mathcal{N}$, respectively. Furthermore, for clarity, the in-degree (resp. out-degree) of a vertex v in $\mathcal{R}$ refers to the number

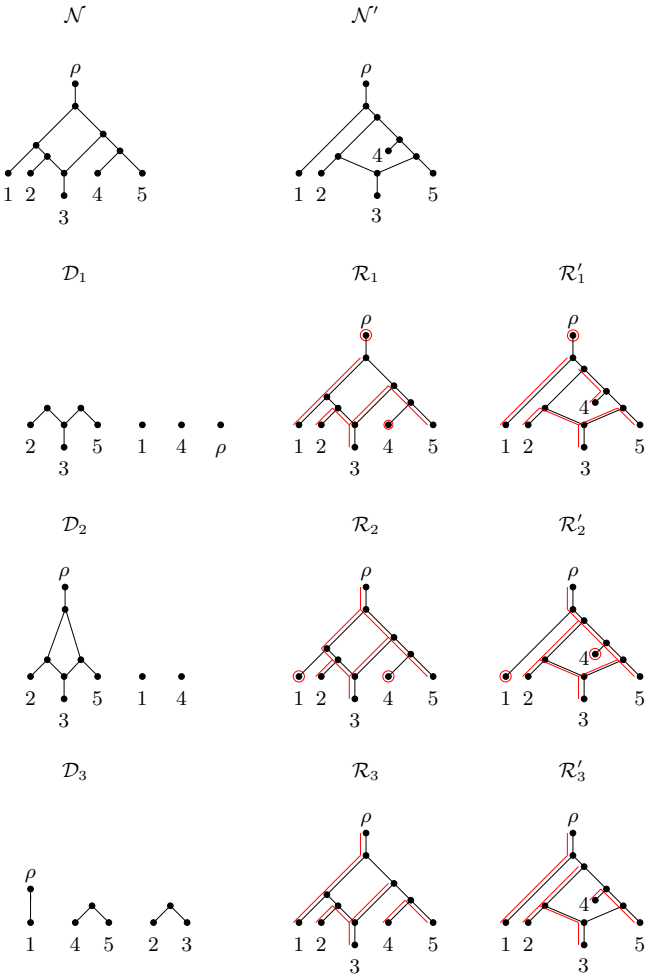

Figure 2: Two phylogenetic networks $\mathcal{N}$ and $\mathcal{N}'$ and three phylogenetic digraphs $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{D}_3$ for $\mathcal{N}$ and $\mathcal{N}'$. For each $i \in \{1, 2, 3\}$, $\mathcal{R}_i$ is an extension of $\mathcal{D}_i$ in $\mathcal{N}$ and $\mathcal{R}'_i$ is an extension of $\mathcal{D}_i$ in $\mathcal{N}'$, where the edges and vertices of the extensions are indicated in red. Note that $\mathcal{R}_3$ is not a root extension of $\mathcal{D}_3$ in $\mathcal{N}$.

of edges in $\mathcal{R}$ that are directed into (resp. out of) v. Lastly, let $\mathcal{D} = \{D_\rho, D_1, D_2, \ldots, D_k\}$ be a phylogenetic digraph of a phylogenetic network $\mathcal{N}$. Then there exists an embedding $\mathcal{M} = \{M_\rho, M_1, M_2, \ldots, M_k\}$ of $\mathcal{D}$ in $\mathcal{N}$ such that M_i and M_j are vertex disjoint for all distinct $i, j \in \{\rho, 1, 2, \ldots, k\}$ and an extension $\mathcal{R} = \{R_\rho, R_1, R_2, \ldots, R_k\}$ of $\mathcal{D}$ in $\mathcal{N}$ such that R_i and R_j are vertex disjoint for all distinct $i, j \in \{\rho, 1, 2, \ldots, k\}$. It follows that each edge in $\mathcal{D}$ corresponds to a unique directed path in $\mathcal{M}$ (resp. $\mathcal{R}$) whose non-terminal vertices all have in-degree one and out-degree one in $\mathcal{M}$ (resp. $\mathcal{R}$), and each vertex in $\mathcal{D}$ corresponds to a unique vertex in $\mathcal{M}$ (resp. $\mathcal{R}$). Reversely, each edge in $\mathcal{M}$ corresponds to a unique edge in $\mathcal{D}$. We will freely use this correspondence throughout the paper.

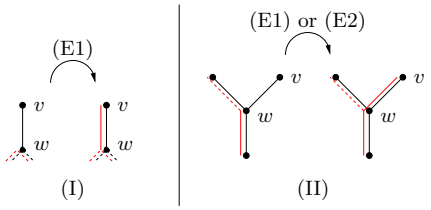

Figure 3: Assuming that $v \notin \mathcal{R}$, the setup as described in the proof of Lemma 5(ii) for when (I) w is a tree vertex or a leaf, and (II) w is a reticulation in $\mathcal{N}$. In both cases, an application of (E1) or (E2) can be used to extend R_i by an additional edge so that the resulting extension contains v. Red solid lines indicate vertices and edges of R_i. Black (resp. red) dashed lines indicate vertices and edges of $\mathcal{N}$ (resp. R_i) that may or may not be vertices and edges of $\mathcal{N}$ (resp. R_i).

4 Properties of extensions

In this section we establish several results for extensions that will be useful in the subsequent sections. Let $\mathcal{D}$ be a phylogenetic digraph of a phylogenetic network $\mathcal{N}$. Given the algorithmic definition of an extension, different orderings of the elements in $\mathcal{D}$ may result in different extensions even for a fixed underlying embedding. Also, if $\mathcal{R}$ and $\mathcal{R}'$ are extensions of $\mathcal{D}$ in $\mathcal{N}$ with distinct underlying embeddings, then $\mathcal{R}$ and $\mathcal{R}'$ are different.

Lemma 5. *Let $\mathcal{D}$ be a phylogenetic digraph of a phylogenetic network $\mathcal{N}$ on X with root ρ, and let $\mathcal{R}$ be an extension of $\mathcal{D}$ in $\mathcal{N}$. Then the following hold:*

(i) *If v is a vertex in $\mathcal{N}$ that is not in $X \cup \{\rho\}$, then there is an edge (v, w) in $\mathcal{N}$ that is in $\mathcal{R}$.*

(ii) *Each vertex in $\mathcal{N}$ is contained in $\mathcal{R}$.*

Proof. Let $\mathcal{D} = \{D_\rho, D_1, D_2, \ldots, D_k\}$, and let $\mathcal{R} = \{R_\rho, R_1, R_2, \ldots, R_k\}$. To see that (i) holds, recall that $\mathcal{R}$ does not contain any vertex with out-degree zero that is not in $X \cup \{\rho\}$. To complete the proof, we establish (ii). This part of the proof is illustrated in Figure 3. By definition of $\mathcal{D}$, the root and each leaf of $\mathcal{N}$ is contained in $\mathcal{R}$. Towards a contradiction, we may therefore assume that there is a tree vertex or a reticulation in $\mathcal{N}$ that is not contained in $\mathcal{R}$. Let v be a vertex of $\mathcal{N}$ that is not in $\mathcal{R}$ such that every vertex that is distinct from v and lies on a directed path from v to a leaf in $\mathcal{N}$ is in $\mathcal{R}$. Furthermore, let w be a child of v. If w is a tree vertex or leaf in $\mathcal{N}$, then there exists a component R_i with $i \in \{\rho, 1, 2, \ldots, k\}$ such that w has in-degree zero in R_i, thereby contradicting that R_i is an extension of D_i as we can apply (E1). Otherwise, if w is a reticulation in $\mathcal{N}$, then it follows from (i) that there exists a component R_i with $i \in \{\rho, 1, 2, \ldots, k\}$ such that w has in-degree zero, or in-degree one and out-degree one in R_i, thereby again contradicting that R_i is an extension of D_i as we can apply either (E1) or (E2), respectively. □

(i) (ii)

Figure 4: Setup as described in the proof of parts (i) and (ii) of Lemma 6. The red solid line in (i) (resp. (ii)) indicates that (v, w) is an edge of $\mathcal{M}$ (resp. $\mathcal{R}$), and the red dashed line in (i) indicates that (v, w') may or may not be an edge of $\mathcal{M}$.

Lemma 6. *Let $\mathcal{N}$ be a phylogenetic network on X. Furthermore, let $\mathcal{M}$ be an embedding of a phylogenetic digraph $\mathcal{D}$ of $\mathcal{N}$, and let $\mathcal{R}$ be an extension of $\mathcal{D}$ such that $\mathcal{M}$ underlies $\mathcal{R}$. Then the following hold for each tree vertex v in $\mathcal{N}$.*

(i) *If v is in $\mathcal{M}$, then no edge directed out of v is in $\mathcal{R}$ and not in $\mathcal{M}$.*

(ii) *If v is not in $\mathcal{M}$, then exactly one of the two edges directed out of v is contained in $\mathcal{R}$.*

Proof. Let u be the parent of v, and let w and w' be the two children of v in $\mathcal{N}$. Furthermore, let $\mathcal{R} = \{R_\rho, R_1, R_2, \ldots, R_k\}$. We first show that (i) holds. Since $\mathcal{M}$ does not contain any vertex of out-degree zero that is not in $X \cup \{\rho\}$, it follows that at least one of (v, w) and (v, w') is in $\mathcal{M}$. Without loss of generality, we may assume that $(v, w) \in \mathcal{M}$. The setup of the proof is shown in Figure 4(i). If $(v, w') \in \mathcal{M}$, then the result clearly holds. On the other hand, if $(v, w') \notin \mathcal{M}$, then it follows from the definition of an extension that $(v, w') \notin \mathcal{R}$, thereby establishing (i). We now turn to (ii) which is illustrated in Figure 4(ii). It follows from Lemma 5(ii) that $v \in \mathcal{R}$. Since $\mathcal{R}$ does not contain any vertex with out-degree zero that is not in $X \cup \{\rho\}$, at least one of (v, w) and (v, w') is contained in $\mathcal{R}$. Moreover, again by the definition of an extension, at most one of (v, w) and (v, w') is contained in $\mathcal{R}$. The lemma now follows. □

Let $\mathcal{D}$ be a phylogenetic digraph of a phylogenetic network $\mathcal{N}$ on X. The next proposition shows that the number of edges that are in $\mathcal{N}$ but not in an extension of $\mathcal{D}$ does not depend on the extension.

Proposition 7. *Let $\mathcal{N}$ be a phylogenetic network on X, and let $\mathcal{R}$ and $\mathcal{R}'$ be two extensions of a phylogenetic digraph $\mathcal{D}$ of $\mathcal{N}$. Then*

$$|E_\mathcal{N} - E_\mathcal{R}| = |E_\mathcal{N} - E_{\mathcal{R}'}|.$$

Proof. Let $\mathcal{M}$ and $\mathcal{M}'$ be the embeddings of $\mathcal{D}$ in $\mathcal{N}$ that underlie $\mathcal{R}$ and $\mathcal{R}'$, respectively. First assume that $\mathcal{M} = \mathcal{M}'$. Since $|E_\mathcal{N} - E_\mathcal{M}| = |E_\mathcal{N} - E_{\mathcal{M}'}|$ and each iteration of (E1) or (E2) in the construction of $\mathcal{R}$ and $\mathcal{R}'$ either adds a new edge (v, w) such that w is already in $\mathcal{R}$ and v is not already in $\mathcal{R}$ or a new edge (v', w') such that w' is already in $\mathcal{R}'$ and v' is not already in $\mathcal{R}'$, the result follows from Lemma 5(ii). Second assume that $\mathcal{M} \neq \mathcal{M}'$. Consider an edge $e = (v, w)$ of $\mathcal{N}$. If $v = \rho$, then, as ρ has out-degree one in $\mathcal{N}$, either $e \in \mathcal{R}$ and $e \in \mathcal{R}'$, or $e \notin \mathcal{R}$ and $e \notin \mathcal{R}'$. Furthermore, if v is a reticulation in

$\mathcal{N}$, then it follows from Lemma 5(i) that $e \in \mathcal{R}$ and $e \in \mathcal{R}'$. We next consider all edges in $\mathcal{N}$ that are directed out of a tree vertex. Let v be a tree vertex of $\mathcal{N}$, and let $e = (v, w)$ and $e' = (v, w')$ be the two edges directed out of v. We next consider four cases that will subsequently be used for a counting argument.

(1) Suppose that both of e and e' are contained in one of $\mathcal{M}$ and $\mathcal{M}'$, and neither e nor e' is contained in the other embedding. Without loss of generality, we may assume that both of e and e' are contained in $\mathcal{M}$ and, therefore, in $\mathcal{R}$. As $\mathcal{M}'$ does not contain a vertex with out-degree zero that is not in $X \cup \{\rho\}$, it follows that $v \notin \mathcal{M}'$. Thus, by Lemma 6(ii), exactly one of e and e' is in $\mathcal{R}'$.

(2) Suppose that exactly one of e and e' is contained in one of $\mathcal{M}$ and $\mathcal{M}'$ and neither e nor e' is contained in the other embedding. Without loss of generality, we may assume that $\mathcal{M}$ contains e and does not contain e', and that $\mathcal{M}'$ contains neither e nor e'. Evidently $e \in \mathcal{R}$ and, by the definition of an extension, $e' \notin \mathcal{R}$. Moreover, as $v \notin \mathcal{M}'$ it follows from Lemma 6(ii) that exactly one of e and e' is in $\mathcal{R}'$.

(3) Suppose that exactly one of e and e' is contained in each of $\mathcal{M}$ and $\mathcal{M}'$. Again by Lemma 6(i), exactly one of e and e' is contained in each of $\mathcal{R}$ and $\mathcal{R}'$.

(4) Suppose that both of e and e' are contained in one of $\mathcal{M}$ and $\mathcal{M}'$, and exactly one of e and e' is contained in the other embedding. Without loss of generality, we may assume that both of e and e' are contained in $\mathcal{M}$ and, therefore, in $\mathcal{R}$. Then, by Lemma 6(i), exactly one of e and e' is contained in $\mathcal{R}'$.

It follows that, if Case (2) or (3) applies to v, then exactly one of e and e' is an element in $E_{\mathcal{N}} - E_{\mathcal{R}}$ and exactly one of e and e' is an element in $E_{\mathcal{N}} - E_{\mathcal{R}'}$. We next turn to Cases (1) and (4). Let V (resp. V') denote the set that contains precisely each tree vertex of $\mathcal{N}$ whose two outgoing edges are both in $\mathcal{M}$ (resp. $\mathcal{M}'$). Since $\mathcal{M}$ and $\mathcal{M}'$ are embeddings of $\mathcal{D}$ in $\mathcal{N}$, we have $|V| = |V'|$ and, consequently, $|V - V'| = |V' - V|$. Hence, the number of tree vertices in $\mathcal{N}$ for which both outgoing edges are in $\mathcal{R}$ and exactly one outgoing edge is in $\mathcal{R}'$ is equal to the number of tree vertices in $\mathcal{N}$ for which both outgoing edges are in $\mathcal{R}'$ and exactly one outgoing edge is in $\mathcal{R}$. This completes the proof of the proposition. □

The last proposition motivates the following terminology. Let $\mathcal{D}$ be a phylogenetic digraph of a phylogenetic network $\mathcal{N}$ on X, and let $\mathcal{R}$ be an extension of $\mathcal{D}$ in $\mathcal{N}$. We set $c_{\mathcal{D}} = |E_{\mathcal{N}} - E_{\mathcal{R}}|$ and refer to $c_{\mathcal{D}}$ as the *cut size* of $\mathcal{D}$ in $\mathcal{N}$. By Proposition 7, $c_{\mathcal{D}}$ is well defined.

The next two results consider cut sizes of root extensions in tree-child networks.

Lemma 8. *Let $\mathcal{D}$ be a phylogenetic digraph of a tree-child network $\mathcal{N}$ on X, and let $\mathcal{R}$ be an extension of $\mathcal{D}$ in $\mathcal{N}$ with cut size $c_{\mathcal{D}}$. Then there also exists a root extension of $\mathcal{D}$ in $\mathcal{N}$ with cut size $c_{\mathcal{D}}$.*

Proof. Let $\mathcal{M}$ be the embedding of $\mathcal{D}$ in $\mathcal{N}$ that underlies $\mathcal{R}$. If $\mathcal{R}$ is not a root extension of $\mathcal{D}$ in $\mathcal{N}$, then there exists a reticulation edge (u, v) in $\mathcal{N}$ such that $(u, v) \in \mathcal{R}$ and

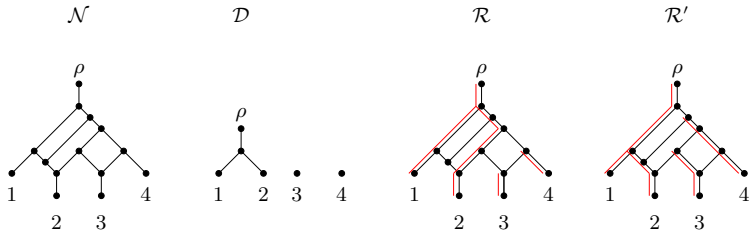

Figure 5: A phylogenetic network $\mathcal{N}$, a phylogenetic digraph $\mathcal{D}$ of $\mathcal{N}$, and two root extensions $\mathcal{R}$ and $\mathcal{R}'$ (indicated by red lines) of $\mathcal{D}$ in $\mathcal{N}$ with $|E_{\mathcal{N}} - E_{\mathcal{R}}| = 6 \neq 5 = |E_{\mathcal{N}} - E_{\mathcal{R}'}|$.

$(u, v) \notin \mathcal{M}$. Since $\mathcal{N}$ is tree-child, u is a tree vertex. Let v' be the child of u in $\mathcal{N}$ with $v \neq v'$. Clearly, v' is a tree vertex or a leaf in $\mathcal{N}$, and $(u, v') \notin \mathcal{R}$. Let R_v be the component in $\mathcal{R}$ that contains v, and let $R_{v'}$ be the component in $\mathcal{R}$ that contains v'. If R_v and $R_{v'}$ are distinct, obtain R'_v from R_v by deleting each vertex s for which there exists a directed path from s to u, and obtain $R'_{v'}$ from $R_{v'}$ by adding (u, v') and each edge of R_v that is contained in a directed path ending at u and, if R_v and $R_{v'}$ are not distinct, obtain R'_v from R_v by deleting (u, v) and adding (u, v'). Then $\mathcal{R}' = (\mathcal{R} - \{R_v, R_{v'}\}) \cup \{R'_v, R'_{v'}\}$ is an extension of $\mathcal{D}$ in $\mathcal{N}$ with cut size $c_{\mathcal{D}}$. It is straightforward to check that $\mathcal{M}$ is the embedding of $\mathcal{D}$ in $\mathcal{N}$ that underlies $\mathcal{R}'$ and that $\mathcal{R}'$ has one reticulation edge less than $\mathcal{R}$. If $\mathcal{R}'$ is not a root extension of $\mathcal{D}$ in $\mathcal{N}$, then repeat the construction for a reticulation edge in $\mathcal{N}$ that is in $\mathcal{R}'$ but not in $\mathcal{M}$ until no such edge remains. This completes the proof of the lemma. □

We end this section by noting that Proposition 7 does not hold for two root extensions of arbitrary phylogenetic networks (see Figure 5 for an example). Nevertheless, we have the following result for tree-child networks. Its proof is similar to that of Proposition 7 and is omitted.

Proposition 9. *Let $\mathcal{N}$ be a tree-child network on X, and let $\mathcal{R}$ and $\mathcal{R}'$ be two root extensions of a phylogenetic digraph $\mathcal{D}$ of $\mathcal{N}$. Then*

$$|E_{\mathcal{N}} - E_{\mathcal{R}}| = |E_{\mathcal{N}} - E_{\mathcal{R}'}|.$$

5 Measures of dissimilarity between two tree-child networks

In this short section, we formally define the tree-child SNPR distance between two tree-child networks and a measure that is associated with an extension of a phylogenetic digraph common to two tree-child networks. This measure bounds the SNPR distance between two tree-child networks from above and below.

Let $\mathcal{N}$ and $\mathcal{N}'$ be two tree-child networks on X. Let $\mathcal{D} = \{D_\rho, D_1, D_2, \ldots, D_k\}$ be a collection of leaf-labelled acyclic digraphs. If $\mathcal{D}$ is a phylogenetic digraph of $\mathcal{N}$ and $\mathcal{N}'$, we refer to $\mathcal{D}$ as an *agreement digraph* of $\mathcal{N}$ and $\mathcal{N}'$. Now let $\mathcal{D}$ be an agreement digraph for $\mathcal{N}$ and $\mathcal{N}'$. If $\mathcal{D}$ is tree-child, we say that $\mathcal{D}$ is an *agreement tree-child digraph* for

$\mathcal{N}$ and $\mathcal{N}'$. In the remainder of this paper, we are particularly interested in agreement digraphs of $\mathcal{N}$ and $\mathcal{N}'$ whose cut size is minimum. To this end, let $\mathcal{D}$ be an agreement tree-child digraph for $\mathcal{N}$ and $\mathcal{N}'$, and let $c_{\mathcal{D}}$ and $c'_{\mathcal{D}}$ be the cut size of $\mathcal{D}$ in $\mathcal{N}$ and $\mathcal{N}'$, respectively. Then $\mathcal{D}$ is called a *maximum agreement tree-child digraph* for $\mathcal{N}$ and $\mathcal{N}'$ if the sum $c_{\mathcal{D}}+c'_{\mathcal{D}}$ is minimum over all agreement tree-child digraphs for $\mathcal{N}$ and $\mathcal{N}'$, in which case $m_{\mathrm{tc}}(\mathcal{N},\mathcal{N}')$ denotes this minimum number. To calculate $m_{\mathrm{tc}}(\mathcal{N},\mathcal{N}')$, it follows from Proposition 7 that it is sufficient to consider a single extension of each agreement digraph for $\mathcal{N}$ and $\mathcal{N}'$. Referring back to Figure 2, observe that each of the three phylogenetic digraphs $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{D}_3$ is in fact an agreement digraph for the two tree-child networks $\mathcal{N}$ and $\mathcal{N}'$ that are shown in the same figure.

Let $\mathcal{N}$ and $\mathcal{N}'$ be two tree-child networks, let $\mathcal{D}$ be an agreement tree-child digraph for $\mathcal{N}$ and $\mathcal{N}'$, and let $\mathcal{R}$ and $\mathcal{R}'$ be an extension of $\mathcal{D}$ in $\mathcal{N}$ and $\mathcal{N}'$, respectively. We note that, similar to the elements of an agreement forest, the elements in $\mathcal{D}$ can be embedded in $\mathcal{N}$ and $\mathcal{N}'$. Intuitively, they can be thought of as subnetworks that are common to $\mathcal{N}$ and $\mathcal{N}'$. On the other hand, the digraphs induced by the edges in $E_{\mathcal{R}} - E_{\mathcal{M}}$ and $E_{\mathcal{R}'} - E_{\mathcal{M}'}$, where $\mathcal{M}$ and $\mathcal{M}'$ is the embedding that underlies $\mathcal{R}$ and $\mathcal{R}'$, respectively, are not necessarily the same. Although each connected component in such a digraph is a rooted tree whose (unique) root is a vertex of $\mathcal{M}$ and $\mathcal{M}'$, respectively, and whose edges are directed *towards* the root, one digraph may contain directed rooted trees with a small total number of unlabelled leaves and the other one may contain directed rooted trees with a much larger total number of unlabelled leaves.

Now, let $\mathcal{N}$ be a phylogenetic network on X, and let $e = (u, v)$ be an edge in $\mathcal{N}$. We consider the following three operations applied to $\mathcal{N}$:

SNPR$^{\pm}$ If u is a tree vertex, then delete e, suppress u, subdivide an edge that is not a descendant of v with a new vertex u', and add the new edge (u', v).

SNPR^{-} If u is a tree vertex and v is a reticulation, then delete e, and suppress u and v.

SNPR^{+} Subdivide e with a new vertex v', subdivide an edge in the resulting network that is not a descendant of v' with a new vertex u', and add the new edge (u', v').

By definition of a tree vertex, $u \neq \rho$ if we apply an SNPR$^{\pm}$. If it is not important which of SNPR^{-}, SNPR^{+}, and SNPR$^{\pm}$ has been applied to $\mathcal{N}$ we simply refer to it as an SNPR. As observed by Bordewich et al. [7, Proposition 3.1], the operation is *reversible*, i.e. if $\mathcal{N}'$ is a phylogenetic network on X that can be obtained from $\mathcal{N}$ by a single SNPR, then $\mathcal{N}$ can also be obtained from $\mathcal{N}'$ by a single SNPR. Lastly, we note that the well-known rSPR operation is an application of SNPR$^{\pm}$ to a phylogenetic tree.

Let $\mathcal{N}$ and $\mathcal{N}'$ be two phylogenetic networks on X. An SNPR *sequence* σ for $\mathcal{N}$ and $\mathcal{N}'$ is a sequence

$$\sigma = (\mathcal{N} = \mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_t = \mathcal{N}')$$

of phylogenetic networks on X such that, for all $i \in \{1, 2, \ldots, t\}$, we have $\mathcal{N}_i$ is obtained from $\mathcal{N}_{i-1}$ by a single SNPR in which case, we say that σ *connects* $\mathcal{N}$ and $\mathcal{N}'$. We refer to

t as the *length* of σ. Let $t^{\pm}$ be the number of phylogenetic networks in $\{\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_t\}$ that have been obtained by an SNPR$^{\pm}$ and, similarly, let t^- and t^+ be the number of phylogenetic networks in $\{\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_t\}$ that have been obtained by an SNPR$^-$ and SNPR$^+$, respectively. Clearly, $t^{\pm} + t^- + t^+ = t$. We set

$$w(\sigma) = 2t^{\pm} + t^- + t^+$$

and refer to $w(\sigma)$ as the *weight* of σ. Intuitively, each deletion or addition of an edge contributes one to $w(\sigma)$. Thus, since an SNPR$^{\pm}$ deletes an edge and, subsequently adds a new edge, such an operation adds two to $w(\sigma)$. It was shown by Bordewich et al. [7, Proposition 3.2] that any two phylogenetic networks $\mathcal{N}$ and $\mathcal{N}'$ on the same leaf set are connected by an SNPR sequence. If $\mathcal{N}$ and $\mathcal{N}'$ are tree-child, then, by the same proposition, there also exists an SNPR sequence that connects $\mathcal{N}$ and $\mathcal{N}'$ such that each network in the sequence is tree-child. We refer to such an SNPR sequence as a *tree-child* SNPR *sequence*. Moreover, we define the *tree-child* SNPR *distance* $d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$ between two tree-child networks $\mathcal{N}$ and $\mathcal{N}'$ as the minimum weight of any tree-child SNPR sequence connecting $\mathcal{N}$ and $\mathcal{N}'$. If $\mathcal{N}$ and $\mathcal{N}'$ are two phylogenetic X-trees, then $d_{\mathrm{rSPR}}(\mathcal{N}, \mathcal{N}')$ denotes the minimum number of rSPR operations needed to transform $\mathcal{N}$ into $\mathcal{N}'$.

Global assumption. Let $\sigma = (\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_t)$ be an SNPR sequence that connects the two phylogenetic networks $\mathcal{N}_0$ and $\mathcal{N}_t$ on X. Throughout the remainder of the paper, we assume that there exists no $i \in \{1, 2, \ldots, t\}$, such that $\mathcal{N}_i$ can be obtained from $\mathcal{N}_{i-1}$ by an SNPR$^{\pm}$ that deletes a reticulation edge in $\mathcal{N}_{i-1}$. Indeed, if such an i exists, then there exists an SNPR sequence

$$\sigma' = (\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_{i-1}, \mathcal{N}_i', \mathcal{N}_i, \mathcal{N}_{i+1}, \ldots, \mathcal{N}_t)$$

such that $\mathcal{N}_i'$ can be obtained from $\mathcal{N}_{i-1}$ by an SNPR$^-$ and $\mathcal{N}_i$ can be obtained from $\mathcal{N}_i'$ by an SNPR$^+$. Since we are interested in SNPR sequences of minimum weight and $w(\sigma') = w(\sigma)$, no generality is lost.

6 Bounding the tree-child SNPR distance

In this section, we establish Theorem 1 and show that the bounds are tight. Let $\mathcal{N}$ and $\mathcal{N}'$ be two tree-child networks, and let $\mathcal{R}$ be an extension of a tree-child digraph $\mathcal{D}$ of $\mathcal{N}$. We first show that, if $\mathcal{N}$, $\mathcal{N}'$, and $\mathcal{R}$ satisfy certain properties, then there exists an extension of $\mathcal{D}$ in $\mathcal{N}'$ such that the cut sizes of $\mathcal{D}$ in $\mathcal{N}$ and $\mathcal{N}'$ differ by at most one.

Lemma 10. *Let $\mathcal{N}$ and $\mathcal{N}'$ be two tree-child networks on X, let $\mathcal{D}$ be a tree-child digraph of $\mathcal{N}$, and let $\mathcal{R}$ be an extension of $\mathcal{D}$ in $\mathcal{N}$.*

(i) *Let (u, v) be a reticulation edge of $\mathcal{N}$ such that $(u, v) \notin \mathcal{R}$. If $\mathcal{N}'$ can be obtained from $\mathcal{N}$ by an* SNPR$^-$ *that deletes (u, v) and suppresses u and v, then $\mathcal{D}$ is a tree-child digraph of $\mathcal{N}'$ and there exists an extension $\mathcal{R}'$ of $\mathcal{D}$ in $\mathcal{N}'$ such that*

$$|E_{\mathcal{N}} - E_{\mathcal{R}}| - 1 = |E_{\mathcal{N}'} - E_{\mathcal{R}'}|.$$

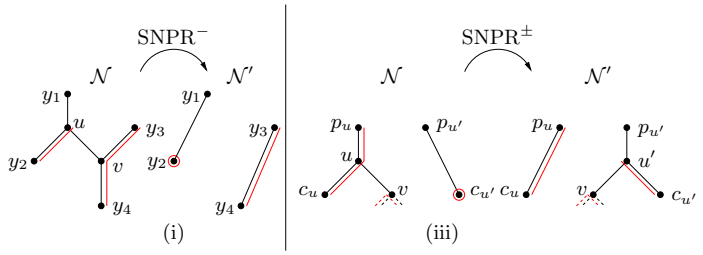

Figure 6: Setup of parts (i) and (iii) in the proof of Lemma 10. Red lines indicate vertices and edges in $\mathcal{R}$ and $\mathcal{R}'$. In (i), observe that u has in-degree zero and out-degree one, and v has in-degree one and out-degree one in $\mathcal{R}$ whereas, in (iii), observe that u has in-degree one and out-degree one, $(p_{u'}, c_{u'}) \notin \mathcal{R}$, and v may or may not be a leaf in $\mathcal{N}$. Vertex labels y_1, y_2, y_3, and y_4 are only used to clarify the figure.

(ii) *Let $e = (u, w)$ and $e' = (u', w')$ be two distinct tree edges of $\mathcal{N}$. If $\mathcal{N}'$ can be obtained from $\mathcal{N}$ by an* SNPR$^+$ *that subdivides e and e' with a new vertex v and v', respectively, and adds the new reticulation edge (v', v), then $\mathcal{D}$ is a tree-child digraph of $\mathcal{N}'$ and there exists an extension $\mathcal{R}'$ of $\mathcal{D}$ in $\mathcal{N}'$ such that*

$$|E_{\mathcal{N}} - E_{\mathcal{R}}| + 1 = |E_{\mathcal{N}'} - E_{\mathcal{R}'}|.$$

(iii) *Let $e = (u, v)$ be a tree edge of $\mathcal{N}$ such that $e \notin \mathcal{R}$, and let $f = (p_{u'}, c_{u'})$ be an edge in $\mathcal{N}$ that is distinct from e. If $\mathcal{N}'$ can be obtained from $\mathcal{N}$ by an* SNPR$^\pm$ *that deletes e, suppresses u, subdivides f with a new vertex u', and adds the new edge (u', v), then $\mathcal{D}$ is a tree-child digraph of $\mathcal{N}'$ and there exists an extension $\mathcal{R}'$ of $\mathcal{D}$ in $\mathcal{N}'$ such that*

$$|E_{\mathcal{N}} - E_{\mathcal{R}}| = |E_{\mathcal{N}'} - E_{\mathcal{R}'}|.$$

Proof. The setup used to establish parts (i) and (iii) is illustrated in Figure 6. We first establish (i). Since $(u, v) \notin \mathcal{R}$ and $\mathcal{D}$ is a tree-child digraph of $\mathcal{N}$, it follows that $\mathcal{D}$ is also such a digraph of $\mathcal{N}'$. Now, let $\mathcal{R}'$ be the digraph obtained from $\mathcal{R}$ by applying the following operation to each vertex $w \in \{u, v\}$. If w has in-degree zero and out-degree one in $\mathcal{R}$, then delete w and, if w has in-degree one and out-degree one in $\mathcal{R}$, then suppress w. As $(u, v) \notin \mathcal{R}$, it follows by Lemma 5 that each of u and v has either in-degree zero and out-degree one, or in-degree one and out-degree one in $\mathcal{R}$. Thus, $\mathcal{R}'$ is well defined. It now follows from the construction that $\mathcal{R}'$ is an extension of $\mathcal{D}$ in $\mathcal{N}'$ with $|E_{\mathcal{N}} - E_{\mathcal{R}}| - 1 = |E_{\mathcal{N}'} - E_{\mathcal{R}'}|$.

To see that (ii) holds, observe first that since $\mathcal{D}$ is a tree-child digraph of $\mathcal{N}$, it follows from the construction of $\mathcal{N}'$ from $\mathcal{N}$ that $\mathcal{D}$ is also such a digraph of $\mathcal{N}'$. Reversing the construction of $\mathcal{R}'$ in (i), let $\mathcal{R}'$ be the digraph obtained from $\mathcal{R}$ by applying the following operations. If $e \notin \mathcal{R}$, then add (v, w) and, if $e \in \mathcal{R}$, then subdivide e with v. Similarly, if $e' \notin \mathcal{R}$, then add (v', w') and, if $e' \in \mathcal{R}$, then subdivide e' with v'. It is now straightforward to check that $\mathcal{R}'$ is an extension of $\mathcal{D}$ in $\mathcal{N}'$ with $|E_{\mathcal{N}} - E_{\mathcal{R}}| + 1 = |E_{\mathcal{N}'} - E_{\mathcal{R}'}|$.

We now turn to (iii). Let p_u be the parent of u, and let c_u be the child of u that is not v. Since u is a tree vertex, p_u and c_u are well defined. Furthermore, let $\mathcal{M}$ be the embedding of $\mathcal{D}$ in $\mathcal{N}$ that underlies $\mathcal{R}$. By Lemma 5(i), $(u, c_u) \in \mathcal{R}$. If $(p_u, u) \notin \mathcal{R}$, then the degree constraints of the vertices in $\mathcal{D}$ imply that $(u, c_u) \notin \mathcal{M}$. On the other hand, if $(p_u, u) \in \mathcal{R}$, then $(u, c_u) \in \mathcal{R}$ and, in turn, either both of (p_u, u) and (u, c_u) are in $\mathcal{M}$ or neither. Now obtain $\mathcal{M}'$ from $\mathcal{M}$ by replacing (p_u, u) and (u, c_u) with (p_u, c_u) if (p_u, u) and (u, c_u) are in $\mathcal{M}$, and replacing f with $(p_{u'}, u')$ and $(u', c_{u'})$ if $f \in \mathcal{M}$. Thus, as $\mathcal{M}$ is the embedding of $\mathcal{D}$ in $\mathcal{N}$ that underlies $\mathcal{R}$, it follows from the construction of $\mathcal{N}'$ from $\mathcal{N}$ that $\mathcal{M}'$ is also an embedding of $\mathcal{D}$ in $\mathcal{N}'$. Hence, $\mathcal{D}$ is a tree-child digraph of $\mathcal{N}'$. To complete the proof, let $\mathcal{R}'$ be the digraph obtained from $\mathcal{R}$ by applying the following operations. If u has in-degree zero in $\mathcal{R}$, then delete u and, if u has in-degree one in $\mathcal{R}$, then suppress u. Moreover, if $f \notin \mathcal{R}$, then add $(u', c_{u'})$ and, if $f \in \mathcal{R}$, then subdivide $(p_{u'}, c_{u'})$ with u'. Again, by Lemma 5, u has either in-degree zero and out-degree one, or in-degree one and out-degree one in $\mathcal{R}$ and, so, $\mathcal{R}'$ is well defined. As $\mathcal{R}$ is an extension of $\mathcal{D}$ in $\mathcal{N}$ with $e \notin \mathcal{R}$ and $(u', v) \notin \mathcal{R}'$, it now follows that $\mathcal{R}'$ is an extension of $\mathcal{D}$ in $\mathcal{N}'$ with $|E_{\mathcal{N}} - E_{\mathcal{R}}| = |E_{\mathcal{N}'} - E_{\mathcal{R}'}|$. □

The next lemma and corollary show that there always exists a tree-child SNPR sequence connecting two tree-child networks with certain desirable properties. These properties will be leveraged later to establish one of the two inequalities of Theorem 1.

Lemma 11. *Let $(\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_t)$ be a tree-child* SNPR *sequence connecting two tree-child networks $\mathcal{N}_0$ and $\mathcal{N}_t$ on X. If there exists an $i \in \{0, 1, 2, \ldots, t-2\}$ such that $\mathcal{N}_{i+1}$ is obtained from $\mathcal{N}_i$ by an* SNPR^+ *(resp.* $\text{SNPR}^{\pm}$*) and $\mathcal{N}_{i+2}$ is obtained from $\mathcal{N}_{i+1}$ by an* SNPR^-*, then one of the following holds:*

(i) $(\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_i, \mathcal{N}_{i+3}, \mathcal{N}_{i+4}, \ldots, \mathcal{N}_{t-1}, \mathcal{N}_t)$ *is a tree-child* SNPR *sequence connecting $\mathcal{N}_0$ and $\mathcal{N}_t$ of length $t-2$,*

(ii) $(\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_i, \mathcal{N}_{i+2}, \mathcal{N}_{i+3}, \ldots, \mathcal{N}_{t-1}, \mathcal{N}_t)$ *is a tree-child* SNPR *sequence connecting $\mathcal{N}_0$ and $\mathcal{N}_t$ of length $t-1$ such that $\mathcal{N}_{i+2}$ is obtained from $\mathcal{N}_i$ by an* $\text{SNPR}^{\pm}$*,*

(iii) $(\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_i, \mathcal{N}_{i+2}, \mathcal{N}_{i+3}, \ldots, \mathcal{N}_{t-1}, \mathcal{N}_t)$ *is a tree-child* SNPR *sequence connecting $\mathcal{N}_0$ and $\mathcal{N}_t$ of length $t-1$ such that $\mathcal{N}_{i+2}$ is obtained from $\mathcal{N}_i$ by an* SNPR^-*, or*

(iv) *there exists a tree-child* SNPR *sequence*

$$(\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_i, \mathcal{N}'_{i+1}, \mathcal{N}_{i+2}, \mathcal{N}_{i+3}, \ldots, \mathcal{N}_{t-1}, \mathcal{N}_t)$$

connecting $\mathcal{N}_0$ and $\mathcal{N}_t$ of length t such that $\mathcal{N}'_{i+1}$ is obtained from $\mathcal{N}_i$ by an SNPR^- *and $\mathcal{N}_{i+2}$ is obtained from $\mathcal{N}'_{i+1}$ by an* SNPR^+ *(resp.* $\text{SNPR}^{\pm}$*).*

Proof. Suppose that there exists an element $i \in \{0, 1, 2, \ldots, t-2\}$ such that $\mathcal{N}_{i+1}$ is obtained from $\mathcal{N}_i$ by an SNPR^+ and $\mathcal{N}_{i+2}$ is obtained from $\mathcal{N}_{i+1}$ by an SNPR^-. Let $e = (u, v)$ be the reticulation edge of $\mathcal{N}_{i+1}$ that is added in obtaining $\mathcal{N}_{i+1}$ from $\mathcal{N}_i$.

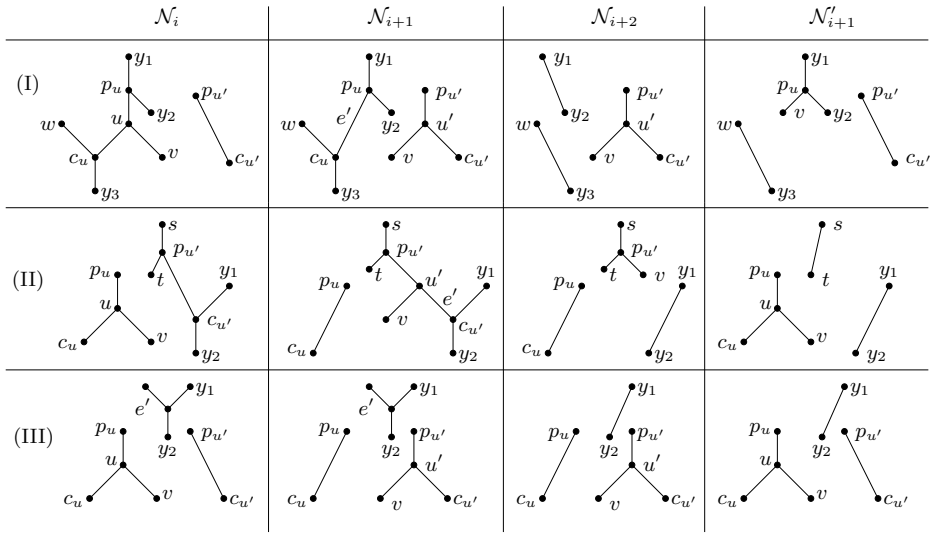

Figure 7: The three cases in the last case analysis in the proof of Lemma 11, where $\mathcal{N}_{i+1}$ is obtained from $\mathcal{N}_i$ by an SNPR$^{\pm}$, $\mathcal{N}_{i+2}$ is obtained from $\mathcal{N}_{i+1}$ by an SNPR^{-}, and $p_{u'} \neq p_u$ or $c_{u'} \neq c_u$. The three cases are (I) $e' = (p_u, c_u)$, (II) $e' = (u', c_{u'})$, and (III) $e' \neq (p_u, c_u)$ and $e' \neq (u', c_{u'})$. Vertex labels y_1, y_2, and y_3 are only used to clarify the figure.

Furthermore, let $e' = (u', v')$ be the reticulation edge in $\mathcal{N}_{i+1}$ that is deleted in obtaining $\mathcal{N}_{i+2}$ from $\mathcal{N}_{i+1}$. If $e = e'$, then $\mathcal{N}_i \cong \mathcal{N}_{i+2}$, and so (i) holds. We may therefore assume that $e \neq e'$. If $v = v'$, let w be the child of v in $\mathcal{N}_{i+1}$, and let c_u and p_u be the child and parent, respectively, of u in $\mathcal{N}_{i+1}$ such that $c_u \neq v$. Since u is a tree vertex, c_u and p_u are well defined. Observe that (u', w) and (p_u, c_u) are tree edges in $\mathcal{N}_i$. It now follows that $\mathcal{N}_{i+2}$ can be obtained from $\mathcal{N}_i$ by the SNPR$^{\pm}$ that deletes (u', w), suppresses u', subdivides (p_u, c_u) with a new vertex u, and adds the edge (u, w), and so (ii) holds. So assume that $e \neq e'$ and $v \neq v'$. As $\mathcal{N}_{i+1}$ is tree-child, it follows that e' is not incident with u or v. Thus, e' is a reticulation edge of $\mathcal{N}_i$. Let $\mathcal{N}'_{i+1}$ be the phylogenetic network obtained from $\mathcal{N}_i$ by deleting e' and suppressing u' and v'. By Lemma 4, $\mathcal{N}'_{i+1}$ is tree-child. It is now straightforward to check that $\mathcal{N}_{i+2}$ can be obtained from $\mathcal{N}'_{i+1}$ by an SNPR^{+}, and so (iv) holds.

Next suppose that there exists an element $i \in \{0, 1, 2, \ldots, t-2\}$ such that $\mathcal{N}_{i+1}$ is obtained from $\mathcal{N}_i$ by an SNPR$^{\pm}$ and $\mathcal{N}_{i+2}$ is obtained from $\mathcal{N}_{i+1}$ by an SNPR^{-}. Let $e = (u, v)$ be the edge of $\mathcal{N}_i$ that is deleted in the process of obtaining $\mathcal{N}_{i+1}$. Furthermore, let p_u be the parent of u and let c_u be the child of u that is not v in $\mathcal{N}_i$. Since u is a tree vertex, p_u and c_u are well defined. Let $f = (p_{u'}, c_{u'})$ be the edge of the digraph resulting from $\mathcal{N}_i$ by deleting e and suppressing u that is subdivided with a new vertex u' such that (u', v) is an edge in $\mathcal{N}_{i+1}$. If $p_{u'} = p_u$ and $c_{u'} = c_u$, then $\mathcal{N}_i \cong \mathcal{N}_{i+1}$, and so (iii) holds. Hence, we may assume that $p_{u'} \neq p_u$ or $c_{u'} \neq c_u$, and so (p_u, c_u) is an edge in $\mathcal{N}_{i+1}$. Let e' be the edge that is deleted in obtaining $\mathcal{N}_{i+2}$ from $\mathcal{N}_{i+1}$. There are three cases to consider, which are illustrated in Figure 7. First assume that $e' = (p_u, c_u)$. Then (u, c_u) is a reticulation edge in $\mathcal{N}_i$. Let w be the parent of c_u that is not u in $\mathcal{N}_i$.

Obtain $\mathcal{N}'_{i+1}$ by deleting (u, c_u) and suppressing the two resulting degree-two vertices. By Lemma 4, $\mathcal{N}'_{i+1}$ is tree-child. Moreover, noting that (p_u, v) is an edge in $\mathcal{N}'_{i+1}$, it follows that $\mathcal{N}_{i+2}$ can be obtained from $\mathcal{N}'_{i+1}$ by an SNPR$^{\pm}$ that deletes (p_u, v), suppresses p_u, subdivides $(p_{u'}, c_{u'})$ if $c_u \neq p_{u'}$ (resp. subdivides $(w, c_{u'})$ if $c_u = p_{u'}$) with a new vertex u', and adds the edge (u', v), and so (iv) holds. Second assume that $e' = (u', c_{u'})$. Then f is a reticulation edge in $\mathcal{N}_i$. Noting that $p_{u'}$ is a tree vertex in $\mathcal{N}_i$, let s be the parent of $p_{u'}$, and let t be the child of $p_{u'}$ that is not $c_{u'}$. Now obtain $\mathcal{N}'_{i+1}$ from $\mathcal{N}_i$ by deleting f and suppressing the two resulting degree-two vertices, one of which is $p_{u'}$. Again by Lemma 4, $\mathcal{N}'_{i+1}$ is tree-child. Furthermore, (s, t) is an edge in $\mathcal{N}'_{i+1}$. Then obtain $\mathcal{N}_{i+2}$ from $\mathcal{N}'_{i+1}$ by an SNPR$^{\pm}$ that deletes (u, v), suppresses u, subdivides (s, t) with a new vertex $p_{u'}$, and adds the edge $(p_{u'}, v)$. Again (iv) holds. Third assume that $e' \neq (p_u, c_u)$ and $e' \neq (u', c_{u'})$. Then e' is an edge of $\mathcal{N}_i$. Let $\mathcal{N}'_{i+1}$ be the tree-child network obtained from $\mathcal{N}_i$ by deleting e' and suppressing the two resulting degree-two vertices. Since e and f are edges of $\mathcal{N}'_{i+1}$, it now follows that $\mathcal{N}_{i+2}$ can be obtained from $\mathcal{N}'_{i+1}$ by an SNPR$^{\pm}$ that deletes e, suppresses u, subdivides f with a new vertex u', and adds the edge (u', v). Again, (iv) holds, thereby completing the proof of the lemma. □

Corollary 12. *Let $\mathcal{N}$ and $\mathcal{N}'$ be two tree-child networks on X. Then there exists a tree-child* SNPR *sequence $(\mathcal{N} = \mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_t = \mathcal{N}')$ that connects $\mathcal{N}$ and $\mathcal{N}'$ such that either $\mathcal{N}_t$ is obtained from $\mathcal{N}_{t-1}$ by an* SNPR^{+} *or* SNPR$^{\pm}$*, or $\mathcal{N}_i$ is obtained from $\mathcal{N}_{i-1}$ by an* SNPR^{-} *for each $i \in \{1, 2, \ldots, t\}$.*

Proof. The corollary follows from repeated applications of Lemma 11. □

The proof of Theorem 1 is an amalgamation of the next two lemmas.

Lemma 13. *Let $\mathcal{N}$ and $\mathcal{N}'$ be two tree-child networks on X. Then*

$$d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') \leqslant m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}').$$

Proof. Let $\mathcal{D} = \{D_\rho, D_1, D_2, \ldots, D_k\}$ be an agreement tree-child digraph for $\mathcal{N}$ and $\mathcal{N}'$, and let $\mathcal{R} = \{R_\rho, R_1, R_2, \ldots, R_k\}$ and $\mathcal{R}' = \{R'_\rho, R'_1, R'_2, \ldots, R'_k\}$ be an extension of $\mathcal{D}$ in $\mathcal{N}$ and $\mathcal{N}'$, respectively. Furthermore, let $c_{\mathcal{D}} = |E_{\mathcal{N}} - E_{\mathcal{R}}|$ and $c'_{\mathcal{D}} = |E_{\mathcal{N}'} - E_{\mathcal{R}'}|$ be the cut size of $\mathcal{D}$ in $\mathcal{N}$ and $\mathcal{N}'$, respectively. By Lemma 8, we may assume that $\mathcal{R}$ and $\mathcal{R}'$ is a root extension of $\mathcal{D}$ in $\mathcal{N}$ and $\mathcal{N}'$, respectively. We show by induction on $c_{\mathcal{D}} + c'_{\mathcal{D}}$ that $d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') \leqslant c_{\mathcal{D}} + c'_{\mathcal{D}}$. The lemma then follows by choosing $\mathcal{D}$ to be an agreement tree-child digraph for $\mathcal{N}$ and $\mathcal{N}'$ such that $c_{\mathcal{D}} + c'_{\mathcal{D}} = m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$.

If $c_{\mathcal{D}} + c'_{\mathcal{D}} = 0$, then $\mathcal{N} \cong \mathcal{N}'$ and consequently $d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') = 0$. Hence, the result follows. Now assume that $c_{\mathcal{D}} + c'_{\mathcal{D}} \geqslant 1$ and that the result holds for all pairs of tree-child networks $\mathcal{N}_1$ and $\mathcal{N}'_1$ on the same leaf set that have an agreement tree-child digraph $\mathcal{D}_1$ with cut size $c_{\mathcal{D}_1}$ and $c'_{\mathcal{D}_1}$ of $\mathcal{D}_1$ in $\mathcal{N}_1$ and $\mathcal{N}'_1$, respectively, such that $c_{\mathcal{D}_1} + c'_{\mathcal{D}_1} < c_{\mathcal{D}} + c'_{\mathcal{D}}$.

We first establish the lemma for a case that is easy to deal with. To this end, assume that there exists a reticulation edge $e = (u, v)$ in $\mathcal{N}$ or $\mathcal{N}'$ that is not contained in $\mathcal{R}$ or $\mathcal{R}'$, respectively. Without loss of generality, we may assume that $e \in \mathcal{N}$. As $\mathcal{N}$ is tree-child, u is a tree vertex. Let $\mathcal{N}''$ be the phylogenetic network obtained from $\mathcal{N}$ by an SNPR^{-}

that deletes e and suppresses the two resulting degree-two vertices. By Lemma 4, $\mathcal{N}''$ is tree-child. Furthermore, by Lemma 10(i), $\mathcal{D}$ is a tree-child digraph for $\mathcal{N}''$ and there exists a root extension $\mathcal{R}''$ of $\mathcal{D}$ in $\mathcal{N}''$ such that $c_{\mathcal{D}} - 1 = c''_{\mathcal{D}}$, where $c''_{\mathcal{D}} = |E_{\mathcal{N}''} - E_{\mathcal{R}''}|$. Since $c''_{\mathcal{D}} + c'_{\mathcal{D}} < c_{\mathcal{D}} + c'_{\mathcal{D}}$, it now follows from the induction assumption that

$$d_{\mathrm{tc}}(\mathcal{N}'', \mathcal{N}') \leqslant c''_{\mathcal{D}} + c'_{\mathcal{D}}.$$

Hence, there exists a tree-child SNPR sequence σ connecting $\mathcal{N}''$ and $\mathcal{N}'$ with $w(\sigma) \leqslant c''_{\mathcal{D}} + c'_{\mathcal{D}}$. Moreover, since $\mathcal{N}''$ can be obtained from $\mathcal{N}$ by a single SNPR$^-$, we have

$$d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') \leqslant 1 + c''_{\mathcal{D}} + c'_{\mathcal{D}} = c_{\mathcal{D}} + c'_{\mathcal{D}},$$

thereby establishing the lemma under the assumption that such an edge e exists.

To complete the proof, we may now assume that

(A) each reticulation edge of $\mathcal{N}$ and $\mathcal{N}'$ is contained in $\mathcal{R}$ and $\mathcal{R}'$, respectively.

Hence, by symmetry, there exists a tree edge $e = (u, w)$ in $\mathcal{N}$ that is not in $\mathcal{R}$. Choose e such that each directed path from w to a leaf in $\mathcal{N}$ only consists of edges in $\mathcal{R}$. Since $\mathcal{N}$ is acyclic such an edge exists. Let D_i be the element in $\mathcal{D}$ with $i \in \{\rho, 1, 2, \ldots, k\}$ such that the root extension R_i of D_i in $\mathcal{N}$ contains w. By the choice of e, R_i exists and each edge in $\mathcal{N}$ that lies on a directed path from w to a leaf is an element of R_i. Thus, if $u = \rho$, then $\mathcal{N} \cong \mathcal{N}'$ and the result follows as $d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') = 0$. We may therefore assume that $u \neq \rho$. The next statement follows from Lemma 5(i), the additional assumption that $u \neq \rho$, and assumption (A).

13.1. *In $\mathcal{N}$, the vertex u is a tree vertex, and w is either a tree vertex or a leaf.*

By the choice of e, observe that the root path of w consists only of w. In turn, w has in-degree zero and out-degree zero or two in R_i. Hence, w corresponds to a unique vertex, say $w_{\mathcal{D}}$, of D_i. Let $r_{w'}$ be the vertex in $\mathcal{N}'$ that $w_{\mathcal{D}}$ corresponds to. Now consider the root extension R'_i of D_i in $\mathcal{N}'$. Let w' be the first vertex of the root path of $r_{w'}$. In contrast to the root path of w in R_i, observe that the root path of $r_{w'}$ may consist of more than a single vertex in which case there is a directed path of length at least one from w' to $r_{w'}$. Since $\rho \in \mathcal{D}$ and $w_{\mathcal{D}} \neq \rho$, it follows that $w' \neq \rho$. Hence, by (A), we have that w' is a leaf, or has in-degree one and out-degree two in $\mathcal{N}'$. Let v' be the parent of w' in $\mathcal{N}'$. As $(v', w') \notin \mathcal{R}'$, the next statement follows from Lemma 5(i).

13.2. *In $\mathcal{N}'$, the edge $e' = (v', w')$ is a tree edge, and v' is either ρ or a tree vertex in $\mathcal{N}'$.*

Let D_j be the element in $\mathcal{D}$ with $j \in \{\rho, 1, 2, \ldots, k\}$ such that the root extension R'_j of D_j in $\mathcal{N}'$ contains v'. By Lemma 5(ii), R'_j exists. We may have $i = j$. We next construct a digraph $\mathcal{D}'$ from $\mathcal{D}$ and a network $\mathcal{N}''$ from $\mathcal{N}$. After detailing the construction, we show that $\mathcal{N}''$ is a tree-child network that can be obtained from $\mathcal{N}$ by a single SNPR$^{\pm}$, and that $\mathcal{D}'$ is an agreement tree-child digraph for $\mathcal{N}''$ and $\mathcal{N}'$. Guided by the second part of (13.2) and noting that, if $v' = \rho$, then ρ is a singleton component in $\mathcal{D}$, there are three cases to consider, which are illustrated in Figure 8:

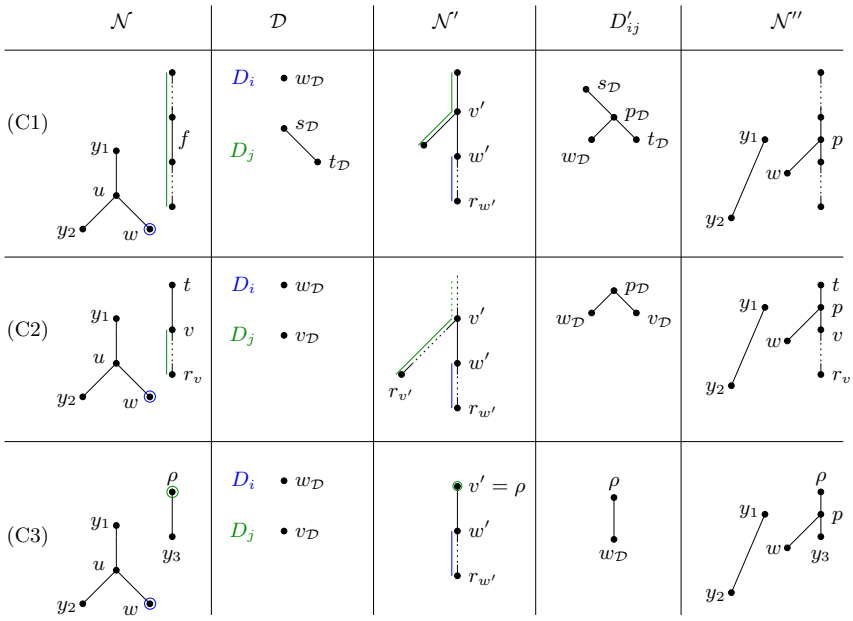

Figure 8: The three Cases (C1)–(C3) as described in the proof of Lemma 13. Blue in $\mathcal{N}$ and $\mathcal{N}'$ indicates edges and vertices in R_i and R'_i, and green in $\mathcal{N}$ and $\mathcal{N}'$ indicates edges and vertices in R_j and R'_j. Vertex labels y_1 and y_2 are only used to clarify the figure.

(C1) Suppose that v' is not a vertex of a root path of a vertex in $\mathcal{R}'$. Clearly $v' \neq \rho$. Then v' has in-degree one and out-degree one in R'_j. Recall that each edge in D_j corresponds to a unique directed path in R'_j that connects the two end vertices of that edge. Let $(s_\mathcal{D}, t_\mathcal{D})$ be the edge in D_j that corresponds to a directed path in R'_j that contains the two edges incident with v'. Obtain D'_{ij} from D_i and D_j by subdividing $(s_\mathcal{D}, t_\mathcal{D})$ with a new vertex $p_\mathcal{D}$ and adding the edge $(p_\mathcal{D}, w_\mathcal{D})$. Turning to $\mathcal{N}$, let f be an edge that lies on the directed path in R_j that corresponds to the edge $(s_\mathcal{D}, t_\mathcal{D})$ in D_j. Obtain $\mathcal{N}''$ from $\mathcal{N}$ by deleting e, suppressing u, subdividing f with a new vertex p, and adding the edge (p, w).

(C2) Suppose that $v' \neq \rho$ and that v' is a vertex of a root path P of $\mathcal{R}'$. Let $r_{v'}$ be the last vertex of P in $\mathcal{N}'$. Noting that $r_{v'}$ corresponds to a vertex $v_\mathcal{D}$ of in-degree zero in D_j, obtain D'_{ij} from D_i and D_j by adding the two edges $(p_\mathcal{D}, v_\mathcal{D})$ and $(p_\mathcal{D}, w_\mathcal{D})$, where $p_\mathcal{D}$ is a new vertex. Turning to $\mathcal{N}$, let r_v be the vertex in R_j that $v_\mathcal{D}$ corresponds to, and let v be the first vertex of the root path of r_v. As $v' \neq \rho$ and $\mathcal{D}$ is an agreement digraph for $\mathcal{N}$ and $\mathcal{N}'$, we have $v \neq \rho$. Moreover, it follows from (A) that v is not a reticulation in $\mathcal{N}$. Thus, v has a unique parent, say t, in $\mathcal{N}$. Then, obtain $\mathcal{N}''$ from $\mathcal{N}$ by deleting e, suppressing u, subdividing the edge $f = (t, v)$ with a new vertex p, and adding the edge (p, w).

(C3) Suppose that $v' = \rho$. As ρ has out-degree one in $\mathcal{N}$ and $\mathcal{N}'$, each of D_j and R'_j consist of the isolated vertex ρ only. Then obtain D'_{ij} from D_i and D_j by adding

a new edge $(\rho, w_{\mathcal{D}})$. Moreover, obtain $\mathcal{N}''$ from $\mathcal{N}$ by deleting e, suppressing u, subdividing the edge f that is directed out of ρ with a new vertex p, and adding the edge (p, w).

As $\mathcal{N}'$ does not contain a directed cycle, it follows from the construction that D'_{ij} is acyclic in all three cases. Hence, as $\mathcal{D}$ is a phylogenetic digraph for $\mathcal{N}'$,

$$\mathcal{D}' = (\mathcal{D} - \{D_i, D_j\}) \cup \{D'_{ij}\}$$

is a phylogenetic digraph of $\mathcal{N}'$. Let E'_i and E'_j be the edge set of R'_i and R'_j respectively, and let R'_{ij} be the subgraph of $\mathcal{N}'$ induced by the edge set $E'_i \cup E'_j \cup \{\{v', w'\}\}$. Since $\mathcal{R}'$ is a root extension of $\mathcal{D}$ in $\mathcal{N}'$, it again follows from the construction that

$$(\mathcal{R}' - \{R'_i, R'_j\}) \cup \{R'_{ij}\}$$

is a root extension of $\mathcal{D}'$ in $\mathcal{N}$.

We next turn to $\mathcal{D}'$ and show that $\mathcal{D}'$ is tree-child. Since $\mathcal{D}$ is tree-child, it follows from the definition that $\mathcal{R}'$ is tree-child. Moreover, since w' has in-degree zero in R'_i, it follows that w' has in-degree one in R'_{ij}. It is now straightforward to check that $(\mathcal{R}' - \{R'_i, R'_j\}) \cup \{R'_{ij}\}$ is tree-child. Hence, again by definition, $\mathcal{D}'$ is also tree-child.

The following statement is now an immediate consequence of the construction.

13.3. *The cut size of $\mathcal{D}'$ in $\mathcal{N}'$ is $c'_{\mathcal{D}} - 1$.*

Next, we establish that $\mathcal{N}''$ is a tree-child network on X.

13.4. *The network $\mathcal{N}''$ is acyclic.*

Proof. Using the same notation as in the construction of $\mathcal{N}''$ from $\mathcal{N}$, recall that $e = (u, w)$ is the edge in $\mathcal{N}$ that is deleted and that f is the edge in $\mathcal{N}$ that is subdivided with p in the process of obtaining $\mathcal{N}''$. To ease reading, let $f = (p_p, c_p)$ regardless of which of (C1)–(C3) applies. Since $\mathcal{N}$ is acyclic, any directed cycle in $\mathcal{N}''$ contains p. If $\mathcal{N}''$ has been obtained from $\mathcal{N}$ as described in (C3), then $\mathcal{N}''$ is acyclic because p has in-degree one and out-degree two in $\mathcal{N}''$ and is adjacent to ρ. Hence, we may assume that $\mathcal{N}''$ has been obtained as described in (C1) or (C2). Towards a contradiction, assume that $\mathcal{N}''$ contains a directed cycle. Then there exists a directed path P from w to c_p in $\mathcal{N}$ whose last edge is f. If $\mathcal{N}''$ has been obtained from $\mathcal{N}$ as described in (C2), then $f \in \mathcal{N}$ and $f \notin \mathcal{R}$, a contradiction because P contains an edge not in $\mathcal{R}$, which is not possible by the choice of e as described in the paragraph following the statement of assumption (A). On the other hand, if $\mathcal{N}''$ has been obtained from $\mathcal{N}$ as described in (C1), then, again by the choice of e and the existence of P, we have $R_i = R_j$. Since $\mathcal{D}$ is an agreement digraph for $\mathcal{N}$ and $\mathcal{N}'$, it follows that the edge $(s_{\mathcal{D}}, t_{\mathcal{D}})$ in D_i can be reached from $w_{\mathcal{D}}$, thereby contradicting that D'_{ij} is acyclic. □

It now follows from the construction of $\mathcal{N}''$ from $\mathcal{N}$ and (13.4) that $\mathcal{N}''$ is a phylogenetic network on X. For the remainder of the proof, let D_u be the element in $\mathcal{D}$ with $u \in \{\rho, 1, 2, \ldots, k\}$ such that the root extension R_u of D_u in $\mathcal{N}$ contains u. By Lemma 5(ii), R_u exists.

13.5. *The phylogenetic network $\mathcal{N}''$ is tree-child.*

Proof. Again using the same notation as in the construction of $\mathcal{N}''$ from $\mathcal{N}$, it follows from (13.1) that the newly added edge (p, w) in $\mathcal{N}''$ is a tree edge. Noting that u is a tree vertex by (13.1) in $\mathcal{N}$, let p_u be the parent of u, and let c_u be the child of u that is not w in $\mathcal{N}$. Observe that (p_u, c_u) is an edge in $\mathcal{N}''$. Now assume that $\mathcal{N}''$ is not tree-child.

First suppose that $\mathcal{N}''$ contains a pair of parallel edges. Then (p_u, c_u), (p_u, u), and (u, c_u) are edges of an underlying three-cycle in $\mathcal{N}$. Assumption (A) and Lemma 5(i) imply that all three edges incident with c_u are edges in R_u. If $(p_u, u) \notin R_u$, then u has in-degree zero and out-degree one in R_u. It follows that u is a vertex of $\mathcal{R}$ but not a vertex of the embedding of $\mathcal{D}$ in $\mathcal{N}$ that underlies $\mathcal{R}$. Hence, the unique child of u in R_u has in-degree one in R_u because $\mathcal{R}$ is a root extension of $\mathcal{D}$, a contradiction as c_u has in-degree two in R_u. Thus, $(p_u, u) \in R_u$. It follow that D_u contains a pair of parallel edges because $e \notin \mathcal{R}$, a contradiction to $\mathcal{D}$ being tree-child.

Second suppose that $\mathcal{N}''$ contains an edge that is incident with two reticulations. Then p_u and c_u are reticulations in $\mathcal{N}$. It follows from (A) and Lemma 5(i), that R_u contains the three edges incident with c_u and the three edges incident with p_u. Thus, D_u contains an edge that is incident with two reticulations because $e \notin \mathcal{R}$, another contradiction.

Third suppose that $\mathcal{N}''$ contains a pair of sibling reticulations. Then c_u is a reticulation and p_u is a tree vertex whose child that is not u, say s_u, is a reticulation in $\mathcal{N}$. Again by (A) and Lemma 5(i), R_u contains all three edges that are incident with c_u and there exists an element $R_{u'} \in \mathcal{R}$ with $u' \in \{\rho, 1, 2, \ldots, k\}$ such that $R_{u'}$ contains all three edges incident with s_u. If $u \neq u'$, then $(p_u, u) \notin \mathcal{R}$ and, thus, u has in-degree zero and out-degree one in R_u. It follows that u is a vertex of $\mathcal{R}$ but not a vertex of the embedding of $\mathcal{D}$ in $\mathcal{N}$ that underlies $\mathcal{R}$. Hence, the unique child of u in R_u has in-degree one in R_u, a contradiction as c_u has in-degree two in R_u. We may therefore assume that $u = u'$. But then R_u contains a pair of sibling reticulations s_u and c_u because $e \notin \mathcal{R}$, a final contradiction. □

It now follows from (13.4) and (13.5) and the construction as detailed in (C1)–(C3) that $\mathcal{N}''$ is a tree-child network on X that can be obtained from $\mathcal{N}$ by a single SNPR$^{\pm}$. We next show that $\mathcal{D}'$ is a phylogenetic digraph for $\mathcal{N}''$. To this end, we construct a root extension of $\mathcal{D}'$ in $\mathcal{N}''$.

If $\mathcal{N}''$ has been obtained from $\mathcal{N}$ as described in (C1), obtain a root extension R_{ij} of D'_{ij} from R_i and R_j by subdividing f in R_j with a new vertex p and adding the edge (p, w). Otherwise, if $\mathcal{N}''$ has been obtained from $\mathcal{N}$ as described in (C2) or (C3), obtain R_{ij} from R_i and R_j by adding the edge (p, v), where p is a new vertex, and adding the edges (p, v) and (p, w). Then, as $\mathcal{R}$ is a root extension of $\mathcal{D}$ in $\mathcal{N}$ and u is a tree vertex in $\mathcal{N}$ by (13.1), it follows that the digraph obtained from

$$\mathcal{R}'' = (\mathcal{R} - \{R_i, R_j\}) \cup \{R_{ij}\}$$

by suppressing (resp. deleting) u if it has in-degree one (resp. zero) in $\mathcal{R}''$ is a root extension of $\mathcal{D}'$ in $\mathcal{N}''$. Thus, $\mathcal{D}'$ is an agreement tree-child digraph for $\mathcal{N}''$ and $\mathcal{N}'$.

The next statement is again an immediate consequence of the construction of $\mathcal{R}''$.

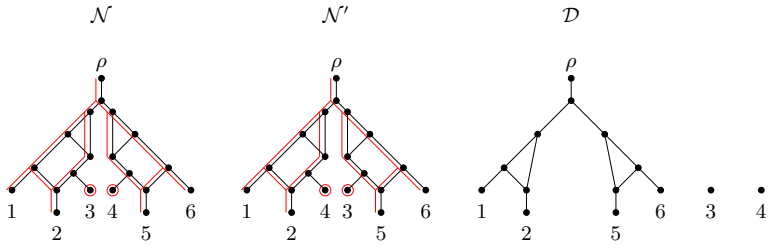

Figure 9: An example of two tree-child networks $\mathcal{N}$ and $\mathcal{N}'$ and an agreement tree-child digraph $\mathcal{D}$ for $\mathcal{N}$ and $\mathcal{N}'$ for which $6 = d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$ but $m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') = 8$. An extension of $\mathcal{D}$ in $\mathcal{N}$ and $\mathcal{N}'$ is indicated by red lines.

13.6. *The cut size of $\mathcal{D}'$ in $\mathcal{N}''$ is $c_{\mathcal{D}} - 1$.*

By combining (13.3) and (13.6), it now follows from the induction assumption that

$$d_{\mathrm{tc}}(\mathcal{N}'', \mathcal{N}') \leqslant c_{\mathcal{D}} - 1 + c'_{\mathcal{D}} - 1.$$

Hence, there exists a tree-child SNPR sequence σ connecting $\mathcal{N}''$ and $\mathcal{N}'$ with $w(\sigma) \leqslant c_{\mathcal{D}} - 1 + c'_{\mathcal{D}} - 1$. Since $\mathcal{N}''$ can be obtained from $\mathcal{N}$ by a single SNPR$^{\pm}$, we have

$$d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') \leqslant 2 + c_{\mathcal{D}} - 1 + c'_{\mathcal{D}} - 1 = c_{\mathcal{D}} + c'_{\mathcal{D}}.$$

The lemma now follows. □

Figure 9 shows two tree-child networks for which the inequality established in Lemma 13 is strict. However, the next lemma shows that, for two tree-child networks $\mathcal{N}$ and $\mathcal{N}'$, the difference $m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') - d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$ cannot be arbitrary large. In preparation for the lemma, we need an additional definition. Let $\mathcal{D}$ be a phylogenetic digraph of a phylogenetic network $\mathcal{N}$ on X. Furthermore, let $\mathcal{R}$ be an extension of $\mathcal{D}$ in $\mathcal{N}$, and let $\mathcal{M}$ be the embedding that underlies $\mathcal{R}$. Now consider a directed path P in $\mathcal{M}$. Let $V = \{v_1, v_2, \ldots, v_n\}$ be the subset of reticulations in $\mathcal{N}$ that lie on P. Then the *path extension* of P contains precisely all edges of P and, additionally, each edge of a maximal length directed path in $\mathcal{N}$ that only consists of edges in $E_{\mathcal{R}} - E_{\mathcal{M}}$ and ends at a vertex in V. Note that the path extension of P may contain each edge of P and no additional edge, even if $V \neq \emptyset$.

Lemma 14. *Let $\mathcal{N}$ and $\mathcal{N}'$ be two tree-child networks on X. Then*

$$\frac{1}{2} m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') \leqslant d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}').$$

Proof. Let $\sigma = (\mathcal{N} = \mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_t = \mathcal{N}')$ be a tree-child SNPR sequence connecting $\mathcal{N}$ and $\mathcal{N}'$ such that $\mathcal{N}_{i-1}$ and $\mathcal{N}_i$ are non-isomorphic for each $i \in \{1, 2, \ldots, t\}$. It follows from Bordewich et al. [7, Proposition 3.2] that σ exists. By Corollary 12, we may assume that $\mathcal{N}_t$ can be obtained from $\mathcal{N}_{t-1}$ by an SNPR^{+} or an SNPR$^{\pm}$, or $\mathcal{N}_i$ can be obtained

from $\mathcal{N}_{i-1}$ by an SNPR$^-$ for each $i \in \{1, 2, \ldots, t\}$. If the latter holds, then, by the reversibility of SNPR,

$$(\mathcal{N}' = \mathcal{N}_t, \ldots, \mathcal{N}_2, \mathcal{N}_1, \mathcal{N}_0 = \mathcal{N})$$

is a tree-child SNPR sequence connecting $\mathcal{N}$ and $\mathcal{N}'$ and $\mathcal{N}_i$ is obtained from $\mathcal{N}_{i+1}$ by an SNPR$^+$ for each $i \in \{t-1, t-2, \ldots, 0\}$. Hence, we may assume without loss of generality that $\mathcal{N}_t$ can be obtained from $\mathcal{N}_{t-1}$ by either an SNPR$^+$ or an SNPR$^\pm$. We show by induction on t that there exist an agreement tree-child digraph $\mathcal{D}$ for $\mathcal{N}$ and $\mathcal{N}'$ and extensions $\mathcal{R}$ and $\mathcal{R}'$ of $\mathcal{D}$ in $\mathcal{N}$ and $\mathcal{N}'$, respectively, such that $\frac{1}{2}m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') \leqslant w(\sigma)$. The lemma then follows by choosing σ such that $w(\sigma) = d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$.

If $t = 1$, then there are two cases to consider. First assume that $\mathcal{N}'$ can be obtained from $\mathcal{N}$ by an SNPR$^+$. Then $w(\sigma) = 1$, and $\mathcal{N}$ is an agreement tree-child digraph for $\mathcal{N}$ and $\mathcal{N}'$. Trivially, there is an extension $\mathcal{R}$ of $\mathcal{N}$ in $\mathcal{N}$ and an extension $\mathcal{R}'$ of $\mathcal{N}$ in $\mathcal{N}'$ such that $|E_{\mathcal{N}} - E_{\mathcal{R}}| + |E_{\mathcal{N}'} - E_{\mathcal{R}'}| = 1$ and, thus, $\frac{1}{2}m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') \leqslant \frac{1}{2} \cdot 1 < w(\sigma)$. Second assume that $\mathcal{N}'$ can be obtained from $\mathcal{N}$ by an SNPR$^\pm$ in which case $w(\sigma) = 2$. Recalling the global assumption stated at the end of Section 5, let $e = (u, v)$ be the tree edge in $\mathcal{N}$ that is deleted in the process of obtaining $\mathcal{N}'$ from $\mathcal{N}$. Let p_u be the parent of u, and let c_u be the child of u in $\mathcal{N}$ that is not v. Since u is a tree vertex, p_u and c_u are well defined. If p_u is a tree vertex, let s be the child of p_u in $\mathcal{N}$ that is not u. Furthermore, if c_u is a reticulation, let s' be the parent of c_u in $\mathcal{N}$ that is not u. If s and c_u are both reticulations, let $\mathcal{D}$ be the leaf-labelled acyclic digraph $\mathcal{D}$ obtained from $\mathcal{N}$ by deleting e and (s', c_u), and suppressing u, c_u, and s'. Otherwise, if at least one of s and c_u is not a reticulation, let $\mathcal{D}$ be the leaf-labelled acyclic digraph $\mathcal{D}$ obtained from $\mathcal{N}$ by deleting e and suppressing u. In both cases, $\mathcal{D}$ is an agreement digraph of $\mathcal{N}$ and $\mathcal{N}'$. We next show that $\mathcal{D}$ is tree-child. If $\mathcal{D}$ contains a pair of parallel edges or a stack, then $\mathcal{N} \cong \mathcal{N}'$, a contradiction to the choice of σ. On the other hand, if $\mathcal{D}$ contains a pair of sibling reticulations, then s and c_u are reticulations in $\mathcal{N}$. By construction, it follows that there is no reticulation in $\mathcal{D}$ that corresponds to c_u. Hence, $\mathcal{D}$ is tree-child. Moreover, there are extensions $\mathcal{R}$ of $\mathcal{D}$ in $\mathcal{N}$ and $\mathcal{R}'$ of $\mathcal{D}$ in $\mathcal{N}'$ such that $|E_{\mathcal{N}} - E_{\mathcal{R}}| + |E_{\mathcal{N}'} - E_{\mathcal{R}'}| \leqslant 2 + 2 = 4$, where the first inequality becomes an equality only if s and c_u are both reticulations in $\mathcal{N}$. It now follows that $\frac{1}{2}m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') \leqslant \frac{1}{2} \cdot 4 = w(\sigma)$. This completes the proof of the base case.

Now suppose that $t > 1$ and that the lemma holds for all pairs of tree-child networks for which there exists a tree-child SNPR sequence connecting the two networks of length less than t. Let

$$\sigma_1 = (\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_{t-1}) \text{ and } \sigma_2 = (\mathcal{N}_{t-1}, \mathcal{N}_t).$$

By Corollary 12, we may again assume that $\mathcal{N}_{t-1}$ can be obtained from $\mathcal{N}_{t-2}$ by an SNPR$^+$ or an SNPR$^\pm$, or $\mathcal{N}_i$ can be obtained from $\mathcal{N}_{i-1}$ by an SNPR$^-$ for each $i \in \{1, 2, \ldots, t-1\}$. Observe that $w(\sigma) = w(\sigma_1) + w(\sigma_2)$. By the induction assumption, we have

$$\frac{1}{2}m_{\mathrm{tc}}(\mathcal{N}_0, \mathcal{N}_{t-1}) \leqslant w(\sigma_1).$$

Hence, there exist a maximum agreement tree-child digraph $\mathcal{D}'$ for $\mathcal{N}_0$ and $\mathcal{N}_{t-1}$ and extensions $\mathcal{R}'_0$ and $\mathcal{R}'_{t-1}$ of $\mathcal{D}'$ in $\mathcal{N}_0$ and $\mathcal{N}_{t-1}$, respectively, such that

$$\frac{1}{2}m_{\mathrm{tc}}(\mathcal{N}_0, \mathcal{N}_{t-1}) = \frac{1}{2}(|E_{\mathcal{N}_0} - E_{\mathcal{R}'_0}| + |E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}|) \leqslant w(\sigma_1). \tag{1}$$

Let $\mathcal{M}'_0$ (resp. $\mathcal{M}'_{t-1}$) be the embedding of $\mathcal{D}'$ in $\mathcal{N}_0$ (resp. $\mathcal{N}_{t-1}$) that underlies $\mathcal{R}'_0$ (resp. $\mathcal{R}'_{t-1}$).

Assume that $\mathcal{N}_t$ can be obtained from $\mathcal{N}_{t-1}$ by an SNPR$^+$, in which case $w(\sigma_2) = 1$. Let (u, w) and (u', w') be the two edges in $\mathcal{N}_{t-1}$ that are subdivided with a new vertex v and v', respectively, in obtaining $\mathcal{N}_t$. Since $\mathcal{N}_t$ is tree-child, (u, w) and (u', w') are tree edges. Furthermore, either (v, v') or (v', v) is a reticulation edge in $\mathcal{N}_t$. Without loss of generality, we may assume that (v', v) is a reticulation edge in $\mathcal{N}_t$. It now follows from Lemma 10(ii) that $\mathcal{D}'$ is also a tree-child digraph for $\mathcal{N}_t$ and there exists an extension $\mathcal{R}'_t$ of $\mathcal{D}'$ in $\mathcal{N}_t$ such that

$$|E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}| + 1 = |E_{\mathcal{N}_t} - E_{\mathcal{R}'_t}|.$$

Hence, we have

$$\begin{aligned}
\frac{1}{2}m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') &\leqslant \frac{1}{2}(|E_{\mathcal{N}_0} - E_{\mathcal{R}'_0}| + |E_{\mathcal{N}_t} - E_{\mathcal{R}'_t}|) \\
&= \frac{1}{2}(|E_{\mathcal{N}_0} - E_{\mathcal{R}'_0}| + |E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}| + 1) \\
&< w(\sigma_1) + w(\sigma_2) = w(\sigma),
\end{aligned}$$

where the last inequality follows from Equation (1) and the fact that $w(\sigma_2) = 1$.

For the remainder of the proof, we may therefore assume that $\mathcal{N}_t$ is obtained from $\mathcal{N}_{t-1}$ by an SNPR$^\pm$ in which case $w(\sigma_2) = 2$. Let $e = (u, v)$ be the edge in $\mathcal{N}_{t-1}$ that is deleted in obtaining $\mathcal{N}_t$ from $\mathcal{N}_{t-1}$. By the definition of SNPR$^\pm$ and the global assumption, u and v are both tree vertices. Let p_u be the parent of u, and let c_u be the child of u with $c_u \neq v$ in $\mathcal{N}_{t-1}$. Observe that (p_u, c_u) is an edge in $\mathcal{N}_t$. Furthermore, let $(p_{u'}, c_{u'})$ be the edge in $\mathcal{N}_{t-1}$ that is subdivided with a new vertex u' in obtaining $\mathcal{N}_t$. Then (u', v), $(p_{u'}, u')$, and $(u', c_{u'})$ are edges in $\mathcal{N}_t$.

First assume that $e \notin \mathcal{R}'_{t-1}$. It follows from Lemma 10(iii) that $\mathcal{D}'$ is also a tree-child digraph for $\mathcal{N}_t$ and there exists an extension $\mathcal{R}'_t$ of $\mathcal{D}'$ in $\mathcal{N}_t$ such that

$$|E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}| = |E_{\mathcal{N}_t} - E_{\mathcal{R}'_t}|$$

and, therefore, again by Equation (1),

$$\begin{aligned}
\frac{1}{2}m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') &\leqslant \frac{1}{2}(|E_{\mathcal{N}_0} - E_{\mathcal{R}'_0}| + |E_{\mathcal{N}_t} - E_{\mathcal{R}'_t}|) \\
&= \frac{1}{2}(|E_{\mathcal{N}_0} - E_{\mathcal{R}'_0}| + |E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}|) \\
&< w(\sigma_1) + w(\sigma_2) = w(\sigma).
\end{aligned}$$

Hence, we may assume that $e \in \mathcal{R}'_{t-1}$. Let R'_u be the element in $\mathcal{R}'_{t-1}$ that contains e, let R'_{c_u} be the element in $\mathcal{R}'_{t-1}$ that contains c_u, and let $R'_{c_{u'}}$ be the element in $\mathcal{R}'_{t-1}$

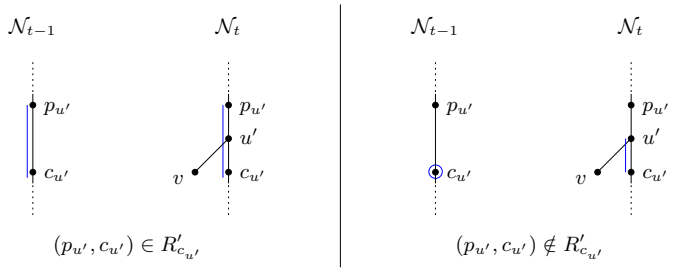

Figure 10: The two cases in the construction of $R_{c_{u'}}$ from $R'_{c_{u'}}$. Blue indicates vertices and edges of $R_{c_{u'}}$ and $R'_{c_{u'}}$.

that contains $c_{u'}$. Recall that R'_u, R'_{c_u}, and $R'_{c_{u'}}$ are not necessarily pairwise distinct. If $(p_{u'}, c_{u'}) \in R'_{c_{u'}}$, then set $R_{c_{u'}}$ to be the directed graph obtained from $R'_{c_{u'}}$ by subdividing $(p_{u'}, c_{u'})$ with a new vertex u'. Otherwise, if $(p_{u'}, c_{u'}) \notin R'_{c_{u'}}$, then set $R_{c_{u'}}$ to be the directed graph obtained from $R'_{c_{u'}}$ by adding $(u', c_{u'})$. The construction is shown in Figure 10. Intuitively, $R_{c_{u'}}$ is an extension of a component of a phylogenetic digraph in $\mathcal{N}_t$. Lastly, if $R'_u = R'_{c_{u'}}$, then set $R'_u = R_{c_{u'}}$ and, if $R'_{c_u} = R'_{c_{u'}}$, then set $R'_{c_u} = R_{c_{u'}}$ to account for the modification in obtaining $R_{c_{u'}}$ from $R'_{c_{u'}}$.

Assume that $e \notin \mathcal{M}'_{t-1}$. Then $(u, c_u) \notin \mathcal{R}'_{t-1}$. It again follows from the construction of $\mathcal{N}_t$ from $\mathcal{N}_{t-1}$ that $\mathcal{D}'$ is a phylogenetic digraph of $\mathcal{N}_t$. Guided by $\mathcal{R}'_{t-1}$, we next construct an extension of $\mathcal{D}'$ in $\mathcal{N}_t$. Let W be the subset of vertices of $\mathcal{N}_{t-1}$ that lie on a directed path from a vertex with in-degree zero to u in R'_u.

(R1) If u is the only element in W and $R'_u \neq R'_{c_u}$, then obtain R_u from R'_u by deleting u, and set $R_{c_u} = R'_{c_u}$.

(R2) If W contains u and $|W| \geqslant 2$, and $R'_u \neq R'_{c_u}$, then obtain R_u from R'_u by deleting each vertex in W, and obtain R_{c_u} from R'_{c_u} by adding (p_u, c_u) and each edge of R'_u that joins two vertices in $W - \{u\}$.

(R3) If u is the only element in W and $R'_u = R'_{c_u}$, then obtain R_u from R'_u by deleting u.

(R4) If W contains u and $|W| \geqslant 2$, and $R'_u = R'_{c_u}$, then obtain R_u from R'_u by deleting u and adding (p_u, c_u).

As an aside, recall that we are dealing with extensions (and not with the more restricted root extensions). Indeed, if c_u is a reticulation in $\mathcal{N}_{t-1}$ and a vertex with in-degree one and out-degree one in R'_{c_u}, then R_{c_u} is an extension and not a root extension. Now, regardless of which of (R1)–(R4) applies, let

$$\mathcal{R}'_t = (\mathcal{R}'_{t-1} - \{R'_u, R'_{c_u}, R'_{c_{u'}}\}) \cup \{R_u, R_{c_u}, R_{c_{u'}}\}.$$

It is easily checked that $\mathcal{R}'_t$ is an extension of $\mathcal{D}'$ in $\mathcal{N}_t$ with

$$|E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}| = |E_{\mathcal{N}_t} - E_{\mathcal{R}'_t}|,$$

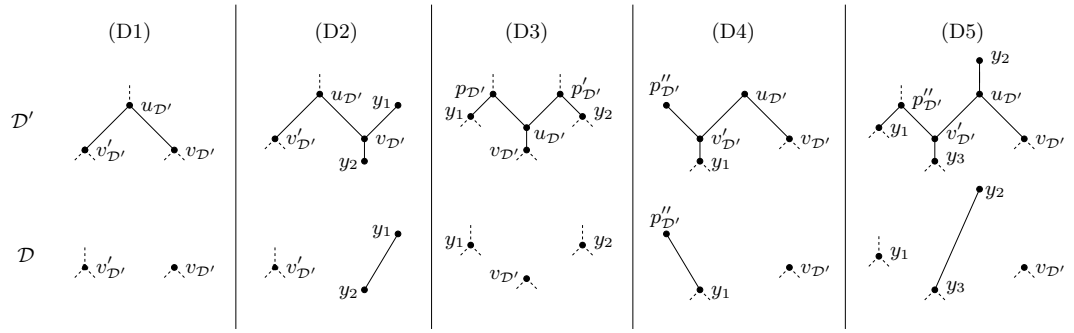

Figure 11: The setup of $\mathcal{D}'$ and $\mathcal{D}$ in all five cases (D1)–(D5) in the construction of $\mathcal{D}$ from $\mathcal{D}'$ as detailed in the proof of Lemma 14. Dashed edges may or may not exist in $\mathcal{D}'$ and $\mathcal{D}$. Vertex labels y_1, y_2, and y_3 are only used to clarify the figure.

and thus,

$$\begin{aligned}\frac{1}{2}m_{\mathrm{tc}}(\mathcal{N},\mathcal{N}') &\leqslant \frac{1}{2}(|E_{\mathcal{N}_0} - E_{\mathcal{R}'_0}| + |E_{\mathcal{N}_t} - E_{\mathcal{R}'_t}|)\\ &= \frac{1}{2}(|E_{\mathcal{N}_0} - E_{\mathcal{R}'_0}| + |E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}|)\\ &< w(\sigma_1) + w(\sigma_2) = w(\sigma).\end{aligned}$$

We complete the proof of the lemma by assuming that $e \in \mathcal{M}'_{t-1}$. Let $e_{\mathcal{D}'} = (u_{\mathcal{D}'}, v_{\mathcal{D}'})$ be the unique edge in $\mathcal{D}'$ that e corresponds to. If $u_{\mathcal{D}'}$ has in-degree two in $\mathcal{D}'$, let $p_{\mathcal{D}'}$ and $p'_{\mathcal{D}'}$ be the two parents of $u_{\mathcal{D}'}$. Furthermore, if $u_{\mathcal{D}'}$ has out-degree two, let $v'_{\mathcal{D}'}$ be the child of $u_{\mathcal{D}'}$ that is not $v_{\mathcal{D}'}$ and, if $v'_{\mathcal{D}'}$ is a reticulation, let $p''_{\mathcal{D}'}$ be the parent of $v'_{\mathcal{D}'}$ that is not $u_{\mathcal{D}'}$. Since $\mathcal{D}'$ is tree-child, observe that each of $p_{\mathcal{D}'}$, $p'_{\mathcal{D}'}$, and $p''_{\mathcal{D}'}$ has, if it exists, in-degree at most one, and that there exists a directed path from each of $p_{\mathcal{D}'}$, $p'_{\mathcal{D}'}$, and $p''_{\mathcal{D}'}$ to a leaf in $\mathcal{D}'$ that does not traverse a reticulation. Lastly, since $\mathcal{D}'$ is tree-child, at most one of $u_{\mathcal{D}'}$, $v_{\mathcal{D}'}$, and $v'_{\mathcal{D}'}$ is a reticulation. Noting that $u_{\mathcal{D}'} \neq \rho$, because $u \neq \rho$ by the definition of SNPR$^{\pm}$, we next obtain a digraph $\mathcal{D}$ from $\mathcal{D}'$ in one of the following five ways, which are illustrated in Figure 11.

(D1) Suppose that $u_{\mathcal{D}'}$ has in-degree zero and out-degree two (resp. in-degree one and out-degree two), and that neither $v_{\mathcal{D}'}$ nor $v'_{\mathcal{D}'}$ is a reticulation. Then obtain $\mathcal{D}$ from $\mathcal{D}'$ by deleting $e_{\mathcal{D}'}$ and deleting (resp. suppressing) $u_{\mathcal{D}'}$.

(D2) Suppose that $u_{\mathcal{D}'}$ has in-degree zero and out-degree two (resp. in-degree one and out-degree two), and that $v_{\mathcal{D}'}$ is a reticulation. Then obtain $\mathcal{D}$ from $\mathcal{D}'$ by deleting $e_{\mathcal{D}'}$, suppressing $v_{\mathcal{D}'}$, and deleting (resp. suppressing) $u_{\mathcal{D}'}$.

(D3) Suppose that $u_{\mathcal{D}'}$ is a reticulation. Then obtain $\mathcal{D}$ from $\mathcal{D}'$ by applying the following three steps in order. First, delete $(p_{\mathcal{D}'}, u_{\mathcal{D}'})$, suppress $u_{\mathcal{D}'}$, and delete the resulting edge $(p'_{\mathcal{D}'}, v_{\mathcal{D}'})$. Second, if $p_{\mathcal{D}'}$ has in-degree zero and out-degree two (resp. in-degree one and out-degree two) in $\mathcal{D}'$, delete (resp. suppress) $p_{\mathcal{D}'}$. Third, if $p'_{\mathcal{D}'}$ has in-degree

zero and out-degree two (resp. in-degree one and out-degree two) in $\mathcal{D}'$, delete (resp. suppress) $p'_{\mathcal{D}'}$.

(D4) Suppose that $u_{\mathcal{D}'}$ has in-degree zero and out-degree two, and that $v'_{\mathcal{D}'}$ is a reticulation. Then obtain $\mathcal{D}$ from $\mathcal{D}'$ by deleting $e_{\mathcal{D}'}$ and $u_{\mathcal{D}'}$, and suppressing $v'_{\mathcal{D}'}$.

(D5) Suppose that $u_{\mathcal{D}'}$ has in-degree one and out-degree two, and that $v'_{\mathcal{D}'}$ is a reticulation. Then obtain $\mathcal{D}$ from $\mathcal{D}'$ by deleting $e_{\mathcal{D}'}$, suppressing $u_{\mathcal{D}'}$, deleting $(p''_{\mathcal{D}'}, v'_{\mathcal{D}'})$, suppressing $v'_{\mathcal{D}'}$, and if $p''_{\mathcal{D}'}$ has in-degree zero and out-degree two (resp. in-degree one and out-degree two) in $\mathcal{D}'$, deleting (resp. suppressing) $p''_{\mathcal{D}'}$.

By construction, it follows that $\mathcal{D}$ neither contains any vertex with in-degree zero and out-degree one except for ρ nor a vertex with in-degree one and out-degree one. Hence, $\mathcal{D}$ is a collection of leaf-labelled acyclic digraphs whose union of leaf sets is X. We next show that $\mathcal{D}$ is an agreement tree-child digraph for $\mathcal{N}_0$ and $\mathcal{N}_t$.

14.1. *$\mathcal{D}$ is an agreement digraph for $\mathcal{N}_0$ and $\mathcal{N}_t$.*

Proof. Since $\mathcal{D}$ is a collection of leaf-labelled acyclic digraphs whose union of leaf sets is X, Properties (i) and (ii) in the definition of a phylogenetic digraph are satisfied. Observe that in each of (D1)–(D5), $\mathcal{D}$ is obtained from $\mathcal{D}'$ by an ordered sequence S of edge deletions, and vertex suppressions and deletions. Furthermore, by construction, a vertex is only suppressed (resp. deleted) if it has in-degree one and out-degree one (resp. in-degree zero and out-degree one) after an incident edge has been deleted. Following the order of operations in S, obtain an embedding $\mathcal{M}_0$ of $\mathcal{D}$ in $\mathcal{N}_0$ from $\mathcal{M}'_0$ as follows. For each edge $f_{\mathcal{D}'}$ that is deleted in obtaining $\mathcal{D}$ from $\mathcal{D}'$, delete each non-terminal vertex of the directed path in $\mathcal{M}'_0$ that corresponds to $f_{\mathcal{D}'}$ and, for each vertex that is deleted in obtaining $\mathcal{D}$ from $\mathcal{D}'$, delete the corresponding vertex in $\mathcal{M}'_0$ and each resulting vertex that has in-degree zero and out-degree one (relative to the embedding) until no such vertex exists. As $\mathcal{D}'$ is a phylogenetic digraph of $\mathcal{N}_0$, it follows from the construction that $\mathcal{M}_0$ is an embedding of $\mathcal{D}$ in $\mathcal{N}_0$ and that the elements in $\mathcal{M}_0$ are pairwise vertex disjoint in $\mathcal{N}_0$. Thus, Property (iii) in the definition of a phylogenetic digraph is satisfied and $\mathcal{D}$ is a phylogenetic digraph of $\mathcal{N}_0$.

We complete the proof by showing that there also exists an embedding $\mathcal{M}_t$ of $\mathcal{D}$ in $\mathcal{N}_t$. Obtain $\mathcal{M}_t$ from $\mathcal{M}'_{t-1}$ by applying the following two steps. First, if there exists an edge f in $\mathcal{M}'_{t-1}$ that corresponds to the edge $(p_{u'}, c_{u'})$ in $\mathcal{N}_{t-1}$, then subdivide f with a new vertex u'. Second, following again the order of operations in S, for each edge $f_{\mathcal{D}'}$ that is deleted in obtaining $\mathcal{D}$ from $\mathcal{D}'$, delete each non-terminal vertex of the directed path in $\mathcal{M}'_{t-1}$ that corresponds to $f_{\mathcal{D}'}$ and, for each vertex that is deleted in obtaining $\mathcal{D}$ from $\mathcal{D}'$, delete the corresponding vertex in $\mathcal{M}'_{t-1}$ and each resulting vertex that has in-degree zero and out-degree one (relative to the embedding) until no such vertex exists. To see that $\mathcal{M}_t$ is indeed an embedding of $\mathcal{D}$ in $\mathcal{N}_t$, recall that $\mathcal{N}_t$ can be obtained from $\mathcal{N}_{t-1}$ by deleting e, suppressing u, subdividing $(p_{u'}, c_{u'})$ with a new vertex u', and adding a new edge (u', v). Since $e_{\mathcal{D}'}$ is deleted and $u_{\mathcal{D}'}$ is either suppressed or deleted in each of (D1)–(D5), it now follows from the construction and the fact that $\mathcal{M}'_{t-1}$ satisfies Property

(iii) in the definition of a phylogenetic digraph that $\mathcal{M}_t$ is an embedding of $\mathcal{D}$ in $\mathcal{N}_t$ and that the elements of $\mathcal{M}_t$ are also pairwise vertex disjoint in $\mathcal{N}_t$. Thus, Property (iii) in the definition of a phylogenetic digraph is satisfied, and $\mathcal{D}$ is a phylogenetic digraph of $\mathcal{N}_t$. □

14.2. *$\mathcal{D}$ is tree-child.*

Proof. Assume $\mathcal{D}$ is not tree-child. Since $\mathcal{D}'$ is tree-child, it follows from the construction of $\mathcal{D}$ that $v'_{\mathcal{D}'}$ is a reticulation in $\mathcal{D}'$ and $\mathcal{D}$. However, if $v'_{\mathcal{D}'}$ is a reticulation in $\mathcal{D}'$, then (D4) or (D5) applies and in each case one of the reticulation edges that are directed into $v'_{\mathcal{D}'}$ is deleted. Thus, $v'_{\mathcal{D}'}$ is not a reticulation in $\mathcal{D}$, a contradiction. □

It now follows from (14.1) and (14.2) that $\mathcal{D}$ is an agreement tree-child digraph for $\mathcal{N}_0$ and $\mathcal{N}_t$.

14.3. *There exists an extension $\mathcal{R}_t$ of $\mathcal{D}$ in $\mathcal{N}_t$ such that*

$$|E_{\mathcal{N}_t} - E_{\mathcal{R}_t}| \leqslant |E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}| + 2.$$

Proof. To ease reading, we view $\mathcal{R}'_{t-1}$ as a collection of edges in $\mathcal{N}_{t-1}$ and describe the construction of $\mathcal{R}_t$ from $\mathcal{R}'_{t-1}$ by edge deletions and additions only. Let P be the directed path in $\mathcal{N}_{t-1}$ that $e_{\mathcal{D}'}$ corresponds to. Clearly e is an edge of P. Furthermore, let (s,t) be the first edge on P, and let E_s be the path extension of the subpath of P from s to v. Observe that, if $s = u$, then s is a tree vertex in $\mathcal{N}_{t-1}$. Hence, in this case, $R'_{c_u} = R'_u$ and $(u, c_u) \in R'_u$. On the other hand, if $s \neq u$, then (u, c_u) is an edge in $E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}$.

We next obtain $\mathcal{R}_t$ from $\mathcal{R}'_{t-1}$. Intuitively, we construct digraphs R_u and R_{c_u} from R'_u and R'_{c_u}, respectively, such that R_u and R_{c_u} are extensions of elements in $\mathcal{D}$ in $\mathcal{N}_t$. As we will see, some of the edges in R'_u and R'_{c_u} that are edges of $\mathcal{M}'_{t-1}$ become edges of R_u and R_{c_u}, respectively, that are not edges of the embedding that underlies $\mathcal{R}_t$. Now suppose that $\mathcal{D}$ has been obtained from $\mathcal{D}'$ by applying the construction as detailed in (D1) or (D2). Obtain R_u and R_{c_u} from R'_u and R'_{c_u}, respectively, in one of the following four ways:

(R1′) Suppose that $s = u$ and $(p_u, u) \in R'_u$. Obtain R_u from R'_u by deleting (p_u, u), (u, c_u), and e, and adding (p_u, c_u).

(R2′) Suppose that $s = u$ and $(p_u, u) \notin R'_u$. Obtain R_u from R'_u by deleting e and (u, c_u).

(R3′) Suppose that $s \neq u$ and $R'_u = R'_{c_u}$. Obtain R_u from R'_u by deleting (s,t), (p_u, u), and e, and if $t \neq u$, adding (p_u, c_u).

(R4′) Suppose that $s \neq u$ and $R'_u \neq R'_{c_u}$. First obtain R_u from R'_u by deleting each edge in E_s. Second if $t = u$, set $R_{c_u} = R'_{c_u}$. Otherwise if $t \neq u$, obtain R_{c_u} from R'_{c_u} by adding (p_u, c_u) and adding each edge in E_s except for (s,t), (p_u, u), and e.

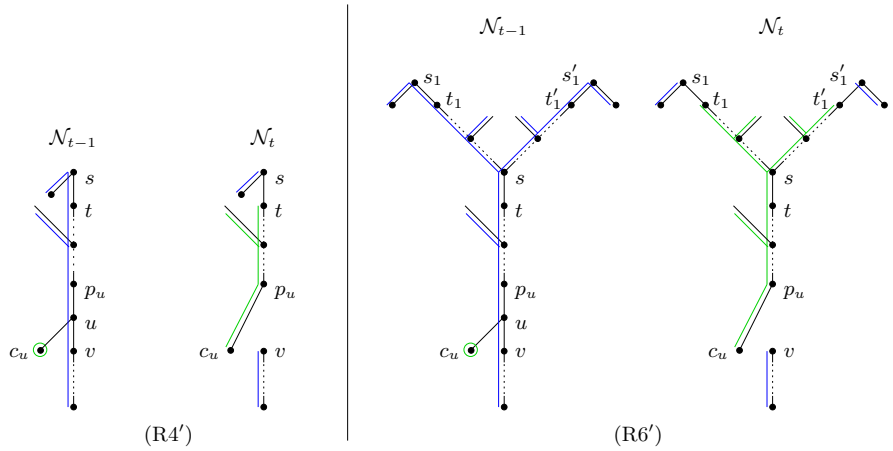

Figure 12: An illustration of (R4′) and (R6′) as described in the proof of (14.3). Blue indicates edges and vertices of R'_u and R_u, and green indicates edges and vertices of R'_{c_u} and R_{c_u}.

While (R3′) and (R4′) are similar in flavour, they are different in the sense that, if $R'_u \neq R'_{c_u}$, then certain edges in R'_u that lie on a path extension of a subpath of P are not edges of R_u and instead get added to R_{c_u}. For an illustration of (R4′), see the left-hand side of Figure 12.

Next, suppose that $\mathcal{D}$ has been obtained from $\mathcal{D}'$ by applying the construction as detailed in (D3). As $u_{\mathcal{D}'}$ has in-degree two and corresponds to s in $\mathcal{N}_{t-1}$, observe that $s \neq u$. Let P_1 (resp. P'_1) be the directed path in $\mathcal{N}_{t-1}$ that the edge $(p_{\mathcal{D}'}, u_{\mathcal{D}'})$ (resp. $(p'_{\mathcal{D}'}, u_{\mathcal{D}'})$) in $\mathcal{D}'$ corresponds to. Furthermore, let (s_1, t_1) (resp. (s'_1, t'_1)) be the first edge on P_1 (resp. P'_1). Since $\mathcal{D}'$ is tree-child, neither s_1 nor s'_1 is a reticulation in $\mathcal{N}_{t-1}$. Similar to the definition of E_s, let E_1 (resp. E'_1) be the path extension of P_1 (resp. P'_1). Finally, obtain R_u and R_{c_u} from R'_u and R'_{c_u}, respectively, in one of the following two ways:

(R5′) Suppose that $R'_u = R'_{c_u}$. Obtain R_u from R'_u by deleting (s_1, t_1), (s'_1, t'_1), (p_u, u), and e, and adding (p_u, c_u).

(R6′) Suppose that $R'_u \neq R'_{c_u}$. First obtain R_u from R'_u by deleting each edge in E_1, E'_1, and E_s. Second obtain R_{c_u} from R'_{c_u} by adding each edge in E_1 except for (s_1, t_1), adding each edge in E'_1 except for (s'_1, t'_1), adding each edge in E_s except for (p_u, u) and e, and adding (p_u, c_u). The construction is shown on the right-hand side of Figure 12.

Lastly, suppose that $\mathcal{D}$ has been obtained from $\mathcal{D}'$ by applying the construction as detailed in (D4) or (D5). Let Q_1 (resp. Q'_1) be the directed path in $\mathcal{N}_{t-1}$ that the edge $(u_{\mathcal{D}'}, v'_{\mathcal{D}'})$ (resp. $(p''_{\mathcal{D}'}, v'_{\mathcal{D}'})$) in $\mathcal{D}'$ corresponds to. Furthermore, let (s_1, t_1) (resp. (s'_1, t'_1)) be the first edge on Q_1 (resp. Q'_1). Note that $s_1 = s$ and, if $s = u$, then $c_u = t_1$. Say first that $\mathcal{D}$ has been obtained from $\mathcal{D}'$ by applying the construction as detailed in (D4). Let F be the subset of edges of $\mathcal{N}_{t-1}$ that lie on a directed path of R'_u that ends at s. Observe that each edge in F is contained in $E_{\mathcal{R}'_{t-1}} - E_{\mathcal{M}'_{t-1}}$. Now obtain R_u and R_{c_u} from R'_u and

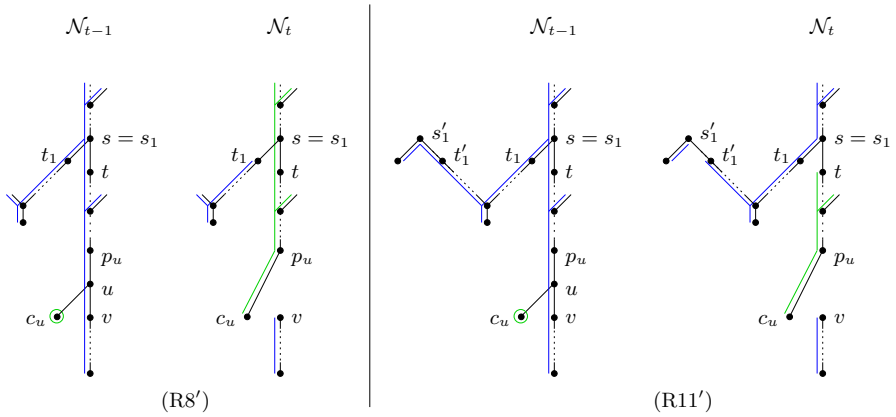

Figure 13: An illustration of (R8′) and (R11′) as described in the proof of (14.3). Blue indicates edges and vertices of R'_u and R_u, and green indicates edges and vertices of R'_{c_u} and R_{c_u}.

R'_{c_u}, respectively, by applying one of (R1′) and (R2′) if $s = u$, or by applying one of the following two ways if $s \neq u$:

(R7′) Suppose that $s \neq u$ and $R'_u = R'_{c_u}$. Obtain R_u from R'_u by deleting (s_1, t_1), (p_u, u), and e, and adding the edge (p_u, c_u).

(R8′) Suppose that $s \neq u$ and $R'_u \neq R'_{c_u}$. First obtain R_u from R'_u by deleting each edge in E_s and F, and deleting (s_1, t_1). Second obtain R_{c_u} from R'_{c_u} by adding each edge in E_s except for (p_u, u) and e, adding (p_u, c_u), and adding each edge in F. See the left-hand side of Figure 13 for an illustration.

On the other hand, if $\mathcal{D}$ has been obtained from $\mathcal{D}'$ by applying the construction as detailed in (D5), then obtain R_u and R_{c_u} from R'_u and R'_{c_u}, respectively, in one of the following three ways:

(R9′) Suppose that $s = u$. Obtain R_u from R'_u by deleting (s'_1, t'_1), (p_u, u), (u, c_u), and e, and adding (p_u, c_u).

(R10′) Suppose that $s \neq u$ and $R'_u = R'_{c_u}$. Obtain R_u from R'_u by deleting (s'_1, t'_1), (s, t), (p_u, u), and e, and, if $t \neq u$, then adding (p_u, c_u).

(R11′) Suppose that $s \neq u$ and $R'_u \neq R'_{c_u}$. First obtain R_u from R'_u by deleting (s'_1, t'_1) and each edge in E_s. Second, if $t = u$, set $R_{c_u} = R'_{c_u}$. Otherwise, if $t \neq u$ obtain R_{c_u} from R'_{c_u} by adding each edge in E_s except for (s, t), (p_u, u), and e, and adding (p_u, c_u). See the right-hand side of Figure 13 for an illustration.

Finally, let $\mathcal{R}_t = (\mathcal{R}'_{t-1} - \{R'_u, R'_{c_u}, R'_{c_{u'}}\}) \cup \{R_u, R_{c_u}, R_{c_{u'}}\}$. Since $\mathcal{R}'_{t-1}$ is an extension of $\mathcal{D}'$ in $\mathcal{N}_{t-1}$, a careful check shows that $\mathcal{R}_t$ is an extension of $\mathcal{D}$ in $\mathcal{N}_t$.

Now, let $C' = E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}$, and let $C = E_{\mathcal{N}_t} - E_{\mathcal{R}_t}$. To complete the proof of (14.3), we compare the number of edges in C' with the number of edges in C. First, observe

that, if $(p_{u'}, c_{u'}) \in C'$, then $(p_{u'}, u') \in C$ and $(u', c_{u'}) \notin C$. Furthermore, if $(p_{u'}, c_{u'}) \notin C'$, then neither $(p_{u'}, u')$ nor $(u', c_{u'})$ is in C. We next list the edges that are in C' but not in C and vice versa for each of (R1′)–(R11′). While $C' - C$ contains edges in $\mathcal{N}_{t-1}$ that are not edges in $\mathcal{N}_t$, the set $C - C'$ contains edges in $\mathcal{N}_t$ that are not edges in $\mathcal{N}_{t-1}$. Thus, edges that are common to $\mathcal{N}_{t-1}$ and $\mathcal{N}_t$ and common to C' and C are not considered in the following table. Moreover, regardless of which of (R1′)–(R11′) applies, we note that $C' - C$ may or may not contain $(p_{u'}, c_{u'})$ and $C - C'$ may or may not contain $(p_{u'}, u')$. However, $C' - C$ contains $(p_{u'}, c_{u'})$ if and only if $C - C'$ contains $(p_{u'}, u')$, and so we have also omitted in the table the possibility that $C' - C$ may contain $(p_{u'}, c_{u'})$ and the possibility that $C - C'$ may contain $(p_{u'}, u')$.

	$(C' - C) - \{(p_{u'}, c_{u'})\}$	$(C - C') - \{(p_{u'}, u')\}$
(R1′)	empty	(u', v)
(R2′)	(p_u, u)	(p_u, c_u), (u', v)
(R3′) and $t = u$	(u, c_u)	(p_u, c_u), (u', v)
(R3′) and $t \neq u$	(u, c_u)	(s, t), (u', v)
(R4′) and $t = u$	(u, c_u)	(p_u, c_u), (u', v)
(R4′) and $t \neq u$	(u, c_u)	(s, t), (u', v)
(R5′) and (R6′)	(u, c_u)	(s_1, t_1), (s_1', t_1'), (u', v)
(R7′) and (R8′)	(u, c_u)	(s_1, t_1), (u', v)
(R9′)	empty	(s_1', t_1'), (u', v)
(R10′) and $t = u$	(u, c_u)	(s_1', t_1'), (p_u, c_u), (u', v)
(R10′) and $t \neq u$	(u, c_u)	(s_1', t_1'), (s, t), (u', v)
(R11′) and $t = u$	(u, c_u)	(s_1', t_1'), (p_u, c_u), (u', v)
(R11′) and $t \neq u$	(u, c_u)	(s_1', t_1'), (s, t), (u', v)

Since $|C - C'| \leqslant |C' - C| + 2$ in all cases, this completes the proof of (14.3). □

14.4. *There exists an extension $\mathcal{R}_0$ of $\mathcal{D}$ in $\mathcal{N}_0$ such that*

$$|E_{\mathcal{N}_0} - E_{\mathcal{R}_0}| \leqslant |E_{\mathcal{N}_0} - E_{\mathcal{R}_0'}| + 2.$$

Proof. Again, to ease reading, we view $\mathcal{R}_0'$ simply as a collection of edges in $\mathcal{N}_0$ and describe the construction of $\mathcal{R}_0$ from $\mathcal{R}_0'$ by edge deletions only. Let P be the directed path in $\mathcal{N}_0$ that $e_{\mathcal{D}'}$ corresponds to, and let (s, t) be the first edge on P. Let R_s' be the element in $\mathcal{R}_0'$ that contains s. We next construct an extension $\mathcal{R}_0$ of $\mathcal{D}$ in $\mathcal{N}_0$ by modifying R_s'. This construction is similar to the constructions described in proof of (14.3), but much less involved. First, suppose that $\mathcal{D}$ has been obtained from $\mathcal{D}'$ by applying (D1), (D2), or (D4). Then

(R1″) obtain R_s from R_s' by deleting (s, t).

Second, suppose that $\mathcal{D}$ has been obtained from $\mathcal{D}'$ by applying (D3). Recall that s is a reticulation. Let P_1 (resp. P_1') be the directed path in $\mathcal{N}_0$ that the edge $(p_{\mathcal{D}'}, u_{\mathcal{D}'})$ (resp. $(p'_{\mathcal{D}'}, u_{\mathcal{D}'})$) in $\mathcal{D}'$ corresponds to. Furthermore, let (s_1, t_1) (resp. (s_1', t_1')) be the first edge on P_1 (resp. P_1'). Then

(R2″) obtain R_s from R'_s by deleting (s_1, t_1) and (s_1', t_1').

Third, suppose that $\mathcal{D}$ has been obtained from $\mathcal{D}'$ by applying (D5). Let P_1 be the directed path in $\mathcal{N}_0$ that the edge $(p''_{\mathcal{D}'}, v'_{\mathcal{D}'})$ in $\mathcal{D}'$ corresponds to. Furthermore, let (s_1, t_1) be the first edge on P_1. Then

(R3″) obtain R_s from R'_s by deleting (s, t) and (s_1, t_1).

Now, let $\mathcal{R}_0 = (\mathcal{R}'_0 - \{R'_s\}) \cup \{R_s\}$. Since $\mathcal{R}'_0$ is an extension of $\mathcal{D}'$ in $\mathcal{N}_0$, a careful check shows that $\mathcal{R}_0$ is an extension of $\mathcal{D}$ in $\mathcal{N}_0$. Moreover, since each of (R1″)–(R3″) deletes at most two edges in obtaining R_s from R'_s, it follows that (14.4) holds. □

Finally, by combining Equation (1) with (14.1)and (14.4), we get

$$\begin{aligned}
\frac{1}{2}m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') &\leqslant \frac{1}{2}(|E_{\mathcal{N}_0} - E_{\mathcal{R}_0}| + |E_{\mathcal{N}_t} - E_{\mathcal{R}_t}|) \\
&\leqslant \frac{1}{2}(|E_{\mathcal{N}_0} - E_{\mathcal{R}'_0}| + 2 + |E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}| + 2) \\
&\leqslant \frac{1}{2}(|E_{\mathcal{N}_0} - E_{\mathcal{R}'_0}| + |E_{\mathcal{N}_{t-1}} - E_{\mathcal{R}'_{t-1}}|) + \frac{1}{2} \cdot 4 \\
&\leqslant w(\sigma_1) + w(\sigma_2) \\
&= w(\sigma).
\end{aligned}$$

This completes the proof of the lemma. □

Proof of Theorem 1. The theorem follows from Lemmas 13 and 14. □

The following result shows how the rSPR distance between two phylogenetic trees can be computed exactly within the framework of agreement digraphs. In particular, it shows that agreement digraphs generalise agreement forest.

Proposition 15. *Let $\mathcal{T}$ and $\mathcal{T}'$ be two phylogenetic X-trees. Then*

$$d_{\mathrm{rSPR}}(\mathcal{T}, \mathcal{T}') = \frac{1}{2}d_{\mathrm{tc}}(\mathcal{T}, \mathcal{T}') = \frac{1}{2}m_{\mathrm{tc}}(\mathcal{T}, \mathcal{T}').$$

Proof. The first equality follows from Bordewich et al. [7, Proposition 7.1] and the fact that each SNPR$^{\pm}$ contributes two to the weight of any SNPR sequence connecting $\mathcal{T}$ and $\mathcal{T}'$. Moreover, to establish the second equality, let

$$\sigma = (\mathcal{T} = \mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_t = \mathcal{T}')$$

be an SNPR sequence connecting $\mathcal{T}$ and $\mathcal{T}'$. Then it follows from Lemma 13 and a careful inspection of the proof of Lemma 14 when applied to two phylogenetic X-trees that (D1)

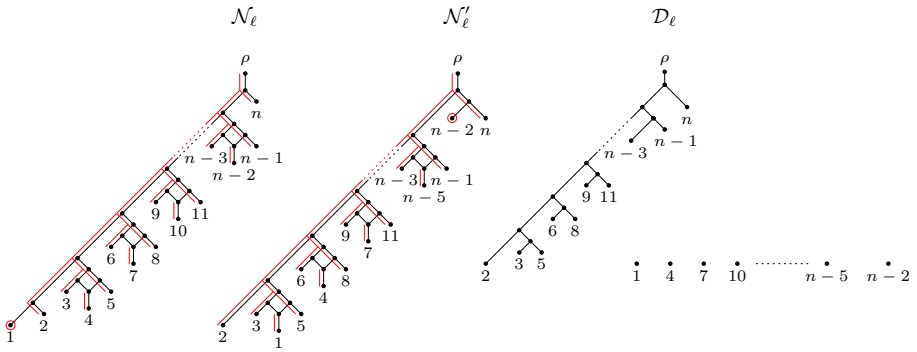

Figure 14: Two phylogenetic networks $\mathcal{N}_\ell$ and $\mathcal{N}'_\ell$ on $n = 3\ell$ leaves with $\ell > 1$, an agreement tree-child digraph $\mathcal{D}_\ell$ for $\mathcal{N}_\ell$ and $\mathcal{N}'_\ell$, an extension of $\mathcal{D}_\ell$ in $\mathcal{N}_\ell$ and an extension of $\mathcal{D}_\ell$ in $\mathcal{N}'_\ell$ indicated in red. This example shows that the bound given in Lemma 14 is essentially tight. For details, see the proof of Proposition 16.

and, consequently, (R1′)–(R4′) and (R1′) always apply. Hence, the last set of inequalities in the proof of Lemma 14 can be replaced with

$$\begin{aligned} m_{\mathrm{tc}}(\mathcal{T}, \mathcal{T}') &\leqslant |E_{\mathcal{T}_0} - E_{\mathcal{R}_0}| + |E_{\mathcal{T}_t} - E_{\mathcal{R}_t}| \\ &\leqslant |E_{\mathcal{T}_0} - E_{\mathcal{R}'_0}| + 1 + |E_{\mathcal{T}_{t-1}} - E_{\mathcal{R}'_{t-1}}| + 1 \\ &\leqslant w(\sigma_1) + w(\sigma_2) \\ &= w(\sigma), \end{aligned}$$

where $\sigma_1 = (\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_{t-1})$ and $\sigma_2 = (\mathcal{T}_{t-1}, \mathcal{T}_t)$. □

The next proposition shows that the bound given in Lemma 14 is essentially tight.

Proposition 16. *For any integer ℓ with $\ell > 1$, there exist two tree-child networks $\mathcal{N}_\ell$ and $\mathcal{N}'_\ell$ on 3ℓ leaves such that $\frac{1}{2}m_{\mathrm{tc}}(\mathcal{N}_\ell, \mathcal{N}'_\ell) + 1 = d_{\mathrm{tc}}(\mathcal{N}_\ell, \mathcal{N}'_\ell)$.*

Proof. Let ℓ be an integer with $\ell > 1$. Consider the two tree-child networks $\mathcal{N}_\ell$ and $\mathcal{N}'_\ell$ that are shown in Figure 14. Each of $\mathcal{N}_\ell$ and $\mathcal{N}'_\ell$ has 3ℓ leaves. Moreover, the agreement tree-child digraph $\mathcal{D}_\ell$ for $\mathcal{N}_\ell$ and $\mathcal{N}'_\ell$ that is also shown in Figure 14 has cut size $2\ell - 1$ in each of $\mathcal{N}_\ell$ and $\mathcal{N}'_\ell$. Thus, $m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') \leqslant 4\ell - 2$. We now show that $m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') = 4\ell - 2$. Assume that $m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') < 4\ell - 2$. Then there exists an agreement tree-child digraph $\mathcal{D}^*_\ell = \{D_\rho, D_1, D_2, \ldots, D_k\}$ whose cut size in $\mathcal{N}_\ell$ or $\mathcal{N}'_\ell$ is strictly less than $2\ell-1$. Since, for each $j \in \{1, 2, \ldots, \ell-1\}$, $\mathcal{N}_\ell$ displays the two triples $(3j, 3j+1, 3j+2)$ and $(3j+1, 3j+2, 3j)$ whereas $\mathcal{N}'_\ell$ only displays the triple $(3j, 3j + 2, 3j + 1)$ and no other triple involving $3j$, $3j + 1$, and $3j + 2$, a careful check shows that $\mathcal{D}^*_\ell$ contains an element that is not a phylogenetic tree. To see this, note that if $\mathcal{D}^*_\ell$ only consists of phylogenetic trees, then each $j \in \{1, 2, \ldots, \ell-1\}$, $\mathcal{N}_\ell$ contributes two to the cut size of $\mathcal{D}^*_\ell$ in $\mathcal{N}_\ell$ and two to the cut size of $\mathcal{D}^*_\ell$ in $\mathcal{N}'_\ell$. Thus, there exists an element D_i in $\mathcal{D}^*_\ell$ for some $i \in \{\rho, 1, 2, \ldots, k\}$ that contains a vertex v of in-degree two and out-degree one. Moreover, as $\mathcal{D}^*_\ell$ is an agreement digraph of $\mathcal{N}_\ell$ and $\mathcal{N}'_\ell$, the child of v is j for some $j \in \{4, 7, \ldots, 3\ell - 5\}$. First, assume

that there exists a vertex u in D_i and two edge-disjoint directed paths P and P' from u to v. Since $\mathcal{D}^*_\ell$ is tree-child, at least one of P and P' contains a vertex w such that the edge (w, v) with $w \neq u$ exists. Furthermore, as D_i can be embedded in $\mathcal{N}_\ell$ and P and P' are edge disjoint, we may assume, that the child of w that is not v is $j-1$ or $j+1$. In either case, it is easily seen that there is no embedding of D_i in $\mathcal{N}'_\ell$, a contradiction. Thus, we may assume that there exist two vertices u and u' with in-degree zero in D_i and directed paths from each of u and u' to v whose only common vertex is v. As D_i can be embedded in $\mathcal{N}_\ell$, we may assume without loss of generality that the child of u is $j+1$ or $j-1$ which leads us to the same contradiction as in the previous case because there is no such embedding of D_i in $\mathcal{N}'_\ell$. Hence, there exists no agreement tree-child digraph whose cut size in $\mathcal{N}_\ell$ or $\mathcal{N}'_\ell$ is strictly less than $2\ell - 1$. It now follows that

$$m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}') = 4\ell - 2. \tag{2}$$

Turning to $d_{\mathrm{tc}}(\mathcal{N}_\ell, \mathcal{N}'_\ell)$, observe that there exists a tree-child SNPR sequence connecting $\mathcal{N}_\ell$ and $\mathcal{N}'_\ell$ that prunes and regrafts the leaves $1, 4, 7, \ldots, 3\ell - 2$ in order. Hence,

$$d_{\mathrm{tc}}(\mathcal{N}_\ell, \mathcal{N}'_\ell) \leqslant 2\ell. \tag{3}$$

By combining Lemma 14 with Equations (2) and (3), we have

$$2\ell - 1 = \frac{1}{2} m_{\mathrm{tc}}(\mathcal{N}_\ell, \mathcal{N}'_\ell) \leqslant d_{\mathrm{tc}}(\mathcal{N}_\ell, \mathcal{N}'_\ell) \leqslant 2\ell$$

which, in turn, implies that $d_{\mathrm{tc}}(\mathcal{N}_\ell, \mathcal{N}'_\ell) \in \{2\ell - 1, 2\ell\}$. Since each of $\mathcal{N}_\ell$ and $\mathcal{N}'_\ell$ has $\ell - 1$ reticulations, the weight of any SNPR sequence connecting $\mathcal{N}_\ell$ and $\mathcal{N}'_\ell$ is even. Thus, $d_{\mathrm{tc}}(\mathcal{N}_\ell, \mathcal{N}'_\ell) = 2\ell$, thereby establishing the proposition. □

7 Concluding remarks

In this paper, we have taken a step towards approximating the tree-child SNPR distance $d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$ between two tree-child networks $\mathcal{N}$ and $\mathcal{N}'$. By introducing phylogenetic digraphs and their extensions, thereby generalising agreement forests for two phylogenetic trees to two phylogenetic networks, we have shown that $d_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$ is tightly bounded from above and below within small constant factors of $m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$. A possible next step is the development of an algorithm to compute $m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$. Due to the intricacies of phylogenetic digraphs and their embeddings, this is a major challenge. In addition, it immediately follows from Proposition 15 and the NP-hardness of computing the rSPR distance between two phylogenetic trees $\mathcal{T}$ and $\mathcal{T}'$ [6] that computing $m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$ is also NP-hard. Since it seems natural to assume that any algorithm for computing $m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$ needs to repeatedly compute cut sizes, it would be interesting to investigate if the cut size of a given phylogenetic digraph for a phylogenetic network $\mathcal{N}$ can be computed efficiently. In a different direction, the development of reductions and divide-and-conquer strategies for computing $m_{\mathrm{tc}}(\mathcal{N}, \mathcal{N}')$ is another avenue for future research. For example, in the context of phylogenetic trees, the introduction of the subtree and chain reduction has led

to fixed-parameter tractable algorithms for computing the rSPR distance [6]. Do similar reductions exist for phylogenetic networks?

As mentioned in the introduction, we use two different weights for SNPR operations to make our approach work. Specifically, each SNPR^+ and SNPR^- is weighted one and each $\text{SNPR}^\pm$ is weighted two. While these weights differ from the uniform weights that are used for computing the rSPR distance between two phylogenetic trees, they are a consequence of how our generalisation from agreement forests to agreement digraphs and their embeddings works. Without going into detail, given two phylogenetic trees $\mathcal{T}$ and $\mathcal{T}'$ with $d_{\text{rSPR}}(\mathcal{T}, \mathcal{T}') = k$, there exists an agreement forest $\mathcal{F}$ for $\mathcal{T}$ and $\mathcal{T}'$ with $k+1$ components. Moreover, one can obtain $\mathcal{F}$ from $\mathcal{T}$ (resp. $\mathcal{T}'$) by deleting k edges in $\mathcal{T}$ (resp. $\mathcal{T}'$) and suppressing vertices with in-degree one and out-degree one after each edge deletion. Intuitively, each rSPR operation is witnessed by an edge in $\mathcal{T}$ and by an edge in $\mathcal{T}'$. In the language of this paper, any agreement forest $\mathcal{F}$ for $\mathcal{T}$ and $\mathcal{T}'$ has the property that its cut size in $\mathcal{T}$ is equal to its cut size in $\mathcal{T}'$ and, thus, $d_{\text{rSPR}}(\mathcal{T}, \mathcal{T}')$ simply equates to the cut size of $\mathcal{F}$ in one of the two trees. Now consider two tree-child networks $\mathcal{N}$ and $\mathcal{N}'$ such that $\mathcal{N}'$ can be obtained from $\mathcal{N}$ by a sequence σ of only SNPR^+ operations. Suppose that the length of σ is k. In this case, $\mathcal{N}$ is an agreement tree-child digraph $\mathcal{D}$ for $\mathcal{N}$ and $\mathcal{N}'$, the cut size of $\mathcal{D}$ in $\mathcal{N}$ is zero, and the cut size of $\mathcal{D}$ in $\mathcal{N}'$ is k. More generally, for an arbitrary tree-child SNPR sequence that connects two tree-child networks $\mathcal{N}$ and $\mathcal{N}'$, each $\text{SNPR}^\pm$ contributes to the cut sizes of both trees, whereas each SNPR^+ and SNPR^- only contributes to the cut size of one tree. Thus any approach for computing $d_{\text{tc}}(\mathcal{N}, \mathcal{N}')$ that is based on cut sizes as defined in this paper (probably) needs to apply non-uniform weights to the different types of SNPR operations. Ultimately, it would be interesting to investigate whether or not an approach exists for computing $d_{\text{tc}}(\mathcal{N}, \mathcal{N}')$ that allows for any combination of weights.

Acknowledgements

We thank the two anonymous referees for their careful reading and constructive comments. All authors thank the New Zealand Marsden Fund for their financial support. Part of this paper is based upon work supported by the National Science Foundation under Grant No. DMS-1929284 while the second and third authors were in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, Rhode Island, US, during the *Theory, Methods, and Applications of Quantitative Phylogenomics* program.

References

[1] B. L. Allen and M. Steel (2001). Subtree transfer operations and their induced metrics on evolutionary trees. *Annals of Combinatorics*, 5:1–15.

[2] C. Allen-Savietta (2020). *Estimating phylogenetic networks from concatenated sequence alignments.* PhD thesis, University of Wisconsin-Madison.

[3] V. Ardévol Martínez, S. Chaplick, S. Kelk, R. Meuwese, M. Mihalák, and G. Stamoulis. Relaxed agreement forests. In: H. Fernau, S. Gaspers, and R. Klasing (Eds.), *SOFSEM 2024: Theory and Practice of Computer Science*, pp. 40–54, Springer.

[4] R. Atkins and C. McDiarmid (2019). Extremal distances for subtree transfer operations in binary trees. *Annals of Combinatorics*, 23:1–26.

[5] M. Baroni, S. Grünewald, V. Moulton, and C. Semple (2005). Bounding the number of hybridisation events for a consistent evolutionary history. *Journal of Mathematical Biology*, 51:171–182.

[6] M. Bordewich and C. Semple (2005). On the computational complexity of the rooted subtree prune and regraft distance. *Annals of Combinatorics*, 8:409–423.

[7] M. Bordewich, S. Linz, and C. Semple (2017). Lost in space? Generalising subtree prune and regraft to spaces of phylogenetic networks. *Journal of Theoretical Biology*, 423:1–12.

[8] G. Cardona, F. Rosselló, and G. Valiente (2009). Comparison of tree-child phylogenetic networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 6:552–569.

[9] J. Chen, J-H. Fan, and S-H. Sze (2015). Parameterized and approximation algorithms for maximum agreement forest in multifurcating trees. *Theoretical Computer Science*, 562:496–512.

[10] C. Choy, J. Jansson, K. Sadakane, and W.-K. Sung (2005). Computing the maximum agreement of phylogenetic networks. *Theoretical Computer Science*, 335:93–107.

[11] Y. Ding, S. Grünewald, P. J. Humphries (2011) On agreement forests. *Journal of Combinatorial Theory Series A*, 118:2059–2065.

[12] J. Döcker, S. Linz, and C. Semple (2021). The display sets of binary normal and tree-child networks. *The Electronic Journal of Combinatorics*, 28, #P1.8.

[13] P. L. Erdős, A. Francis, and T. R. Mezei (2021). Rooted NNI moves and distance-1 tail moves on tree-based phylogenetic networks. *Discrete Applied Mathematics*, 294:205–213.

[14] A. Francis, K. T. Huber, V. Moulton, and T. Wu (2017). Bounds for phylogenetic network space metrics. *Journal of Mathematical Biology*, 76:1229–1248.

[15] P. Gambette, L. van Iersel, M. Jones. M. Lafond, F. Pardi, and C. Scornavacca (2017). Rearrangement moves on rooted phylogenetic networks. *PLoS Computational Biology*, 13:e1005611.

[16] J. Hein, T. Jiang, L. Wang, and K. Zhang (1996). On the complexity of comparing evolutionary trees. *Discrete Applied Mathematics*, 71:153–169.

[17] K. T. Huber. S. Linz, V. Moulton, and T. Wu (2015). Spaces of phylogenetic networks from generalized nearest-neighbor interchange operations. *Journal of Mathematical Biology*, 72:699–725.

[18] K. T. Huber, V. Moulton, and T. Wu (2016). Transforming phylogenetic networks: Moving beyond tree space. *Journal of Theoretical Biology*, 404:30–39.

[19] R. Janssen (2021). Heading in the right direction? Using head moves to traverse phylogenetic network space. *Journal of Graph Algorithms and Applications*, 25:263–310.

[20] R. Janssen (2024). PhyloX: A Python package for complete phylogenetic network workflows. *Journal of Open Source Software*, 9:6427.

[21] R. Janssen, M. Jones, P. L. Erdős, L. van Iersel, and C. Scornavacca (2018). Exploring the tiers of rooted phylogenetic network space using tail moves. *Bulletin of Mathematical Biology*, 80:2177–2208.

[22] R. Janssen and J. Klawitter (2019). Rearrangement operations on unrooted phylogenetic networks. *Theory and Applications of Graphs*, 22:1–31.

[23] R. Janssen (2021). *Rearranging phylogenetic networks*. PhD thesis, Delft University of Technology.

[24] J. Jansson and W.-K. Sung (2004). The maximum agreement of two nested phylogenetic networks. In: R. Fleischer and G. Trippen (Eds.), *15th International Symposium on Algorithms and Computation*, Lecture Notes in Computer Science, Volume 3341, pp. 581–593.

[25] S. Kelk and S. Linz (2020). New reduction rules for the tree bisection and reconnection distance. *Annals of Combinatorics*, 24:475–502.

[26] S. Kelk, S. Linz, and R. Meuwese (2024). Deep kernelization for the tree bisection and reconnection (TBR) distance in phylogenetics, *Journal of Computer and System Sciences*, 142:103519.

[27] J. Klawitter (2018). The SNPR neighbourhood of tree-child networks. *Journal of Graph Algorithms and Applications*, 22:329–355.

[28] J. Klawitter (2019). The agreement distance of rooted phylogenetic networks. *Discrete Mathematics and Theoretical Computer Science*, 21:19.

[29] J. Klawitter (2020). The agreement distance of unrooted phylogenetic networks. *Discrete Mathematics and Theoretical Computer Science*, 22:1 #22.

[30] J. Klawitter (2020). *Spaces of phylogenetic networks*. PhD thesis, University of Auckland.

[31] J. Klawitter and S. Linz (2019). On the subnet prune and regraft distance. *The Electronic Journal of Combinatorics*, 23, #P2.3.

[32] S. Kong, J. C. Pons, L. Kubatko, and K. Wicke (2022). Classes of explicit phylogenetic networks and their biological and mathematical significance. *Journal of Mathematical Biology*, 84:47.

[33] S. Linz and C. Semple (2009). Hybridization in nonbinary trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 6:30–45.

[34] S. Linz and K. Wicke (2023). Exploring spaces of semi-directed level-1 networks. *Journal of Mathematical Biology*, 87:70.

[35] A. Markin, S. Wagle, T. K. Anderson, and O. Eulenstein (2022). RF-Net 2: Fast inference of virus reassortment and hybridization networks. *Bioinformatics*, 38: 2144–2152.

[36] N. F. Müller, K. E. Kistler, and T. Bedford (2022). A Bayesian approach to infer recombination patterns in coronaviruses. *Nature Communication*, 13:4186.

[37] N. F. Müller, U. Stolz, G. Dudas, T. Stadler, and T. G. Vaughan (2020). Bayesian inference of reassortment networks reveals fitness benefits of reassortment in human influenza viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 117:17104–17111.

[38] N. Olver, F. Schalekamp, S. van der Ster, L. Stougie, and A. van Zuylen (2022). A duality based 2-approximation algorithm for maximum agreement forest. *Mathematical Programming*, 198:811–853.

[39] K. St. John (2017). Review paper: The shape of phylogenetic treespace. *Systematic Biology*, 66:e83–e94.

[40] C. Semple and M. Steel (2003). *Phylogenetics*. Oxford University Press.

[41] F. Shi, J. Chen, Q. Feng, and J. Wang (2018). A parameterized algorithm for the maximum agreement forest problem on multiple rooted multifurcating trees. *Journal of Computer and System Sciences*, 97:28–44.

[42] G. Valiente (2009). *Combinatorial Pattern Matching Algorithms in Computational Biology Using Perl and R*. Chapman & Hall.

[43] L. van Iersel, R. Janssen, M. Jones, Y. Murakami, N. Zeh (2022). A practical fixed-parameter algorithm for constructing tree-child networks from multiple binary trees. *Algorithmica*, 84, 917–960.

[44] R. van Wersch, S. Kelk, S. Linz, and G. Stamoulis (2022). Reflections on kernelizing and computing unrooted agreement forests. *Annals of Operations Research*, 309:425–451.

[45] C. Whidden, R. G. Beiko, and N. Zeh (2013). Fixed-parameter algorithms for maximum agreement forests. *SIAM Journal on Computing*, 42:1431–1466.