Scornavacca and Weller
Algorithms for Molecular Biology (2022) 17:15
https://doi.org/10.1186/s13015-022-00216-w

Algorithms for Molecular Biology

RESEARCH

Open Access

Treewidth-based algorithms for the small parsimony problem on networks

Celine Scornavacca[1] and Mathias Weller[2*]

Abstract

Background: Phylogenetic reconstruction is one of the paramount challenges of contemporary bioinformatics. A subtask of existing tree reconstruction algorithms is modeled by the Small Parsimony problem: given a tree T and an assignment of character-states to its leaves, assign states to the internal nodes of T such as to minimize the *parsimony score*, that is, the number of edges of T connecting nodes with different states. While this problem is polynomial-time solvable on trees, the matter is more complicated if T contains reticulate events such as hybridizations or recombinations, i.e. when T is a network. Indeed, three different versions of the parsimony score on networks have been proposed and each of them is NP-hard to decide. Existing parameterized algorithms focus on combining the number c of possible character-states with the number of reticulate events (per biconnected component).

Results: We consider the parameter treewidth t of the underlying undirected graph of the input network, presenting dynamic programming algorithms for (slight generalizations of) all three versions of the parsimony problem on size-n networks running in times $c^t n^{O(1)}$, $(3c)^t n^{O(1)}$, and $6^{tc} n^{O(1)}$, respectively. Our algorithms use a formulation of the treewidth that may facilitate formalizing treewidth-based dynamic programming algorithms on phylogenetic networks for other problems.

Conclusions: Our algorithms allow the computation of the three popular parsimony scores, modeling the evolutionary development of a (multistate) character on a given phylogenetic network of low treewidth. Our results subsume and improve previously known algorithm for all three variants. While our results rely on being given a "good" tree-decomposition of the input, encouraging theoretical results as well as practical implementations producing them are publicly available. We present a reformulation of tree decompositions in terms of "agreeing trees" on the same set of nodes. As this formulation may come more natural to researchers and engineers developing algorithms for phylogenetic networks, we hope to render exploiting the input network's treewidth as parameter more accessible to this audience.

Keywords: Phylogenetics, Parsimony, Phylogenetic networks, Parameterized complexity, Dynamic programming, Treewidth

Introduction

Molecular phylogenetic reconstruction consists in inferring a well-founded evolutionary scenario of a set of species from molecular data [1]. An evolutionary scenario, also called a *phylogeny*, is usually represented by a directed tree with a unique source called *root*. In a phylogeny, the tips of the tree are associated to extant species for which we have data, and each internal node represents an extinct species giving rise to new species—a *speciation*. Therefore, each internal node represents the hypothetical ancestor of all species below it, and the root models the lowest common ancestor of all the species at the tips.

*Correspondence: mathias.weller@u-pem.fr

[2] LIGM, Université Gustave Eiffel, CNRS, Paris, France
Full list of author information is available at the end of the article

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Parsimony on trees

In this paper, molecular data consists of a set of molecular sequences (e.g. DNA or protein sequences) of the same length (one sequence per species). This kind of data can be seen as a matrix M of n sequences, each having m characters (exhibiting one of c possible states) where $M_{i,j}$ corresponds to the state of the jth character exhibited by the ith species. There are several methods to reconstruct well-founded phylogenies from matrices of characters [1]. They are all based on the idea of retrieving similarities among species by comparing the states taken by these species at the different characters of M. Here, we will focus on *parsimony methods*. The main hypothesis of these methods is that character changes are not frequent. Thus, the phylogenies that best explain the data are those requiring the fewest evolutionary changes, i.e. the ones having the optimal *parsimony score*, formally defined in "Parsimony". The problem of finding the optimal parsimony score for a given phylogeny T with respect to an $n \times m$ matrix on a finite set of c character states is called the Small Parsimony problem and can be solved in $O(n \cdot m \cdot c)$ time [2] since each column in the matrix can be analyzed independently in linear time. When T is unknown, the problem of finding the phylogeny minimizing the parsimony score is called the Big Parsimony problem. This latter is known to be NP-hard and numerous heuristic techniques for it are known [1].

Parsimony on networks

When the evolution of the species of interest include, in addition to speciations, reticulate events such as *hybridizations* or *recombinations*, a single species may inherit from multiple direct ancestors. In this case, the phylogenies are no longer represented by rooted trees but by rooted DAGs [3] called *networks*. When scoring a given network, three very different definitions of the parsimony score have been proposed: the *hardwired* [4], the *softwired* [5, 6], and the *parental* parsimony score [7]. Roughly, the hardwired score takes into account all edges of the given network (characters are inherited from all parents), the softwired score takes only the edges of any "switching" (each character is inherited from one parent), and the parental score allows embedding lineages into the network (each allele of a character is inherited from one parent). See "Parsimony" for details and Fig. 1 for an example. While these definitions coincide for trees, they give rise to three different small parsimony problems for networks.

When tracing mutually dependent characters (e.g. different genomic locations in a same non-recombinant region) on networks, we also have to make sure that dependent characters are inherited from the same parent (some columns of the matrix have to use the same

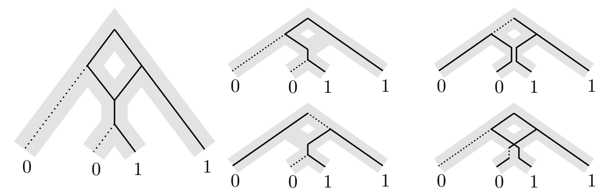

Fig. 1 Example for parsimony scores of a network (in gray). Black edges participate in the score (solid = score 0, dotted = score 1). For the hardwired score (left), all edges of the network are considered. For the softwired score (2 possible trees: middle), only edges of any switching are considered. For the parental score (4 possible trees: middle & right), a tree is inscribed in the network

"switching"/"embedding"). To avoid dealing with this problem, the small parsimony problems on networks have been studied predominantly under the assumption of independent genomic locations. This boils down to having $m = 1$ since each column of the matrix can be analyzed independently (as is the case for the small parsimony problem on trees). Another popular restriction is to consider *binary* networks, in which the root has outdegree 2, tips have indegree 1, and internal nodes have either indegree 1 and outdegree 2 (speciations) or indegree 2 and outdegree 1 (reticulations).

The hardwired small parsimony problem has been proven NP-hard and APX-hard whenever the number of states that a character can take, denoted c, is strictly greater than 2, and polynomial-time solvable for binary characters [8]. A polynomial-time 1.35-approximation for all c and a $\frac{12}{11}$-approximation for $c = 3$ have been proposed [8]. Additionally, the problem has been shown fixed-parameter tractable (FPT) in the parsimony score [8, $2^p \cdot O(\min(q^{\frac{2}{3}}, \sqrt{z}) \cdot q)$ time], and in $c + r$ [9, $O(n \cdot c^{r+2})$ time], where n, q, z are the number of leaves, vertices and edges in the phylogenetic network and p and r are the hardwired parsimony score and the number of reticulate events in the network.

The softwired small parsimony problem is also NP-hard and APX-hard [8, 10] for binary characters, and not FPT in the parsimony score (it is NP-hard to decide if the softwired parsimony score is 1). Also, it has been shown that, for any constant $\epsilon > 0$, no $n^{1-\epsilon}$ approximation can be computed in polynomial time, unless P = NP. On the positive side, the problem is FPT in $c + r$ [6, 8, $O(2^r \cdot n \cdot c)$ time] and $c + \ell$ [8, 11, $O(2^\ell \cdot c^2 \cdot q \cdot z)$ time], where ℓ is the maximum number of reticulations over all biconnected components of the network (also called the *level* of the network).

Unsurprisingly, the parental small parsimony problem has also been proven NP-hard, even for very restricted classes of networks, but it is FPT both with respect to $c + r$ and with respect to $c + \ell$ [12, $O((2^c)^{r+2} \cdot q)$ and $O((2^c)^{\ell+3} \cdot q)$ time].

In this paper, we consider the case of independent characters, showing that the three variants of the small parsimony problem on networks are fixed-parameter tractable with respect to $c + t$ (running in time $O(T + c^{t+1} \cdot z)$, $O(T + c^t \cdot (3^t \cdot c \cdot q + z))$, and $O(T + 6^{t \cdot c} \cdot 4^{t \cdot \log(c)} \cdot z)$), provided that a width-t tree-decomposition of the input network N can be computed in T time (this is the case for t equaling the treewidth of N and $T \in 2^{O(k^2)}$ [13]). Our proofs are constructive in the sense that a dynamic programming algorithm is provided for each version of the problem. The main strength of our algorithms lies in their parameterization, since the treewidth can be arbitrarily small, even for growing values of ℓ. An implication of parameterizing by the treewidth is that our algorithms run in polynomial time even on classes of networks on which previously known algorithms require exponential time[1] while our algorithms run in polynomial time on all classes of networks that were previously known to allow for polynomial-time algorithms. Hence, our algorithms can potentially be orders of magnitude faster than the state-of-the-art solutions. Moreover, our formulations are not limited to binary networks and they can take into account polymorphism as well as external information controlling the states that ancestral species may take.

Treewidth for phylogenetic networks

The treewidth of a graph can roughly be described as a measure of "tree-likeness" and it ranks among the smallest of such parameters [14] (in particular, the treewidth can be seen to be smaller than the level ℓ on any network). Together with the fact that it facilitates the design of dynamic programming algorithms, this explains the enormous popularity the treewidth received in the parameterized complexity community [15, 16]. Starting with the groundbreaking work of Bryant and Lagergren [17] (using the celebrated result of Courcelle [18]), treewidth also gained traction with researchers studying algorithms for phylogenetics-related problems (surveyed in [19]). While this yielded some algorithms parameterized by the treewidth *of the display graph* of multiple trees (the result of "gluing" all trees at their leaves), we are not aware of any algorithms parameterized by the treewidth of the input network. In an attempt to facilitate the use of this parameter in future work, we dedicate Sect. "An alternative formulation of treewidth" to presenting a "phylogenetics-friendly" formulation by representing tree-decompositions of the input network as a rooted tree Γ on the same vertex set as the network. In particular, this formulation generalizes our previously considered parameter "scanwidth" [20], which can be seen as a variant of treewidth that takes directness into account. While we expected scanwidth-based dynamic programming formulations to be easier and more straight-forward than their treewidth-counterparts, this comes at the cost of the scanwidth being potentially arbitrarily larger than the treewidth. Intuitively speaking, we expect scanwidth dynamic programming to be easier since phylogenetic networks exhibit a "natural flow of information": most often, we know everything about the leaves, but the more we approach the root, the more information has to be inferred from the lower parts. In contrast to the scanwidth-layout, tree-decompositions disregard edge directions and, thereby, this "natural flow". Thus, while using the scanwidth allows for more naïve and intuitive dynamic programming formulations, using the treewidth requires more care and ingenuity.

Since we will suppose that a (not necessarily optimal) tree-decomposition of the input network is given in the input, let us discuss the current state-of-the-art for computing good decompositions. Optimal decompositions are indeed very hard to compute, with even the best-known parameterized algorithm being considered impractical (see survey [15]). This gloomy cloud has, however, two silver linings: First, if we do not insist on optimality, then we can use a recently published algorithm to compute 2-approximated tree-decompositions in $2^{O(k)} n^{O(1)}$ time [21]. We will state our results in a way that allows plugging-in any algorithm that computes or approximates tree decompositions. Second, with development driven by recent instances of the PACE challenge [22], more practical exact algorithms to compute tree decompositions are now available as well [23]. Herein, the running times of Tamaki's implementation [23] are hard to predict and show erratic behavior even for fixed graph size. As expected, however, examples for high running times occur only for instances with high treewidth, that is, for "highly tangled" networks (see Fig. 2 for two select examples). This hints towards some hidden properties of the input networks that govern the complexity of treewidth computations As we expect "natural networks" to be only moderately tangled, we think that existing algorithms, exact and approximative, are currently well-enough developed to deal with real world phylogenetic networks in reasonable timeframes. Indeed, we would welcome efforts similar to those made for the treewidth to also be made for the previously discussed scanwidth, which is also hard to compute [20].

For ease of presentation, the three main proofs (correctness of the dynamic programming formulations) are

[1] For example, networks whose "worst" biconnected component is equal to the result of glueing two copies of the same n-leaf tree at corresponding leaves are known to have treewidth two, but level at least $n - 1$.

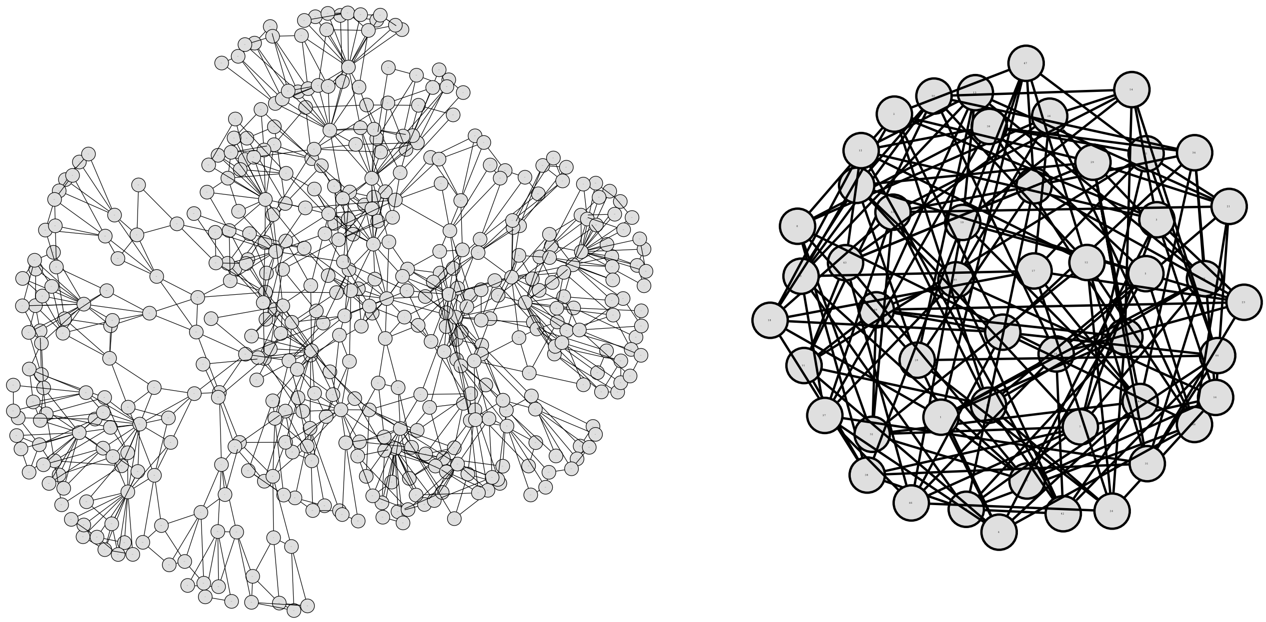

Fig. 2 Tamaki's tree-decomposer [23] has a harder time with the right, more tangled instance (50 nodes, 175 edges, treewidth 25 computed in 79s) than with the larger instance on the left (465 nodes, 1004 edges, treewidth 9 computed in 0.5s), illustrating that tangledness is a more important factor than size. Indeed, both instances display a tangledness that already exceeds what we expect to see in real-world phylogenetic networks. The instances are `ex065` (right) and `ex011` (left) of the PACE2017 challenge [22]

given as high-level sketches and their more detailed and formal versions can be found in the appendix.

Preliminaries

Mappings

For any x and y, we define $\delta(x, y)$ to be 0 if $x = y$ and 1, otherwise, and we abbreviate $1 - \delta(x, y) =: \overline{\delta}(x, y)$. We further abbreviate $\delta(\phi(x), \phi(y))$ as $\delta_\phi(x, y)$ for any function ϕ. We may denote a pair (x, y) as $x \to y$ if it is referring to an assignment of y to x by some function and as xy if it refers to an arc in a network. We sometimes use the name of a function $\phi : X \to Y$ to refer to its set of pairs $\{x \to y \mid \phi(x) = y\}$ and we let $\phi\mid_Z := \{(x \to y) \in \phi \mid x \in Z\}$ denote the *restriction* of ϕ to Z. We say $\phi(x) = \bot$ to indicate that ϕ is not defined for x. We denote the result of forcing $\phi(x) = y$ (whether or not x is mapped by ϕ) as

$$\phi[x \to y] := \begin{cases} \phi \cup \{x \to y\} & \text{if } \phi(x) = \bot \\ (\phi \setminus \{x \to \phi(x)\})[x \to y] & \text{otherwise} \end{cases}$$

Finally, for sets Z, X and $Y \subseteq X$ and functions ϕ and ψ, we write $\psi \trianglelefteq \phi$ (and say that ψ is a *subfunction* of ϕ) if (a) $\phi : X \to Z$ and $\psi : Y \to Z$ and $\psi(x) \leq \phi(x)$ for all $x \in Y$, or (b) $\phi : X \to 2^Z$ and $\psi : Y \to Z$ and $\psi(x) \in \phi(x)$ for all $x \in Y$, or (c) $\phi : X \to 2^Z$ and $\psi : Y \to 2^Z$ and $\psi(x) \subseteq \phi(x)$ for all $x \in Y$.

Graphs and phylogenetic networks

In this work, we consider directed acyclic graphs (DAGs) N that may have a unique source ρ_N called *root*. If the sinks (aka *leaves*) of N are labeled, we call N a *phylogenetic network*. We refer to the nodes and directed edges (arcs) of N by $V(N)$ and $A(N)$, respectively. The *underlying undirected graph* of N is the undirected graph on node-set $V(N)$ that contains an edge $\{u, v\}$ if and only if N contains the arc (u, v). As we do not deal with mixed graphs, we use the term uv to refer to the arc from u to v or the undirected edge between u and v, depending on the context. We refer to the edge-set of an undirected graph G as $E(G)$.

We denote the set of nodes of a DAG N with in-degree at least two by $R(N)$ and we call such nodes *reticulations*. If $R(N) = \emptyset$, then N is called a *tree*. The result of, for each $v \in R(N)$ removing all but one of its incoming arcs is called a *switching* of N and $\mathcal{S}(N)$ denotes the set of all switchings of N (observe that all switchings are spanning trees). For each $v \in V(N)$, we denote the successors (or "children") of v in N by $\mathrm{Succ}_N(v)$ and its predecessors (or "parents") by $\mathrm{Pred}_N(v)$. If N contains a directed u-w-path, then we say that w is a *descendant* of u and u is an *ancestor* of w (denoted as $w \leq_N u$ and $w <_N u$ if $u \neq w$). A set $Z \subseteq V(N)$ such that $u \not\leq_N w$ and $w \not\leq_N u$ for all $u, w \in Z$ is called an *anti-chain* in N. The *induced subgraph* $N[Z]$ of a set $Z \subseteq V(N)$ is the result of

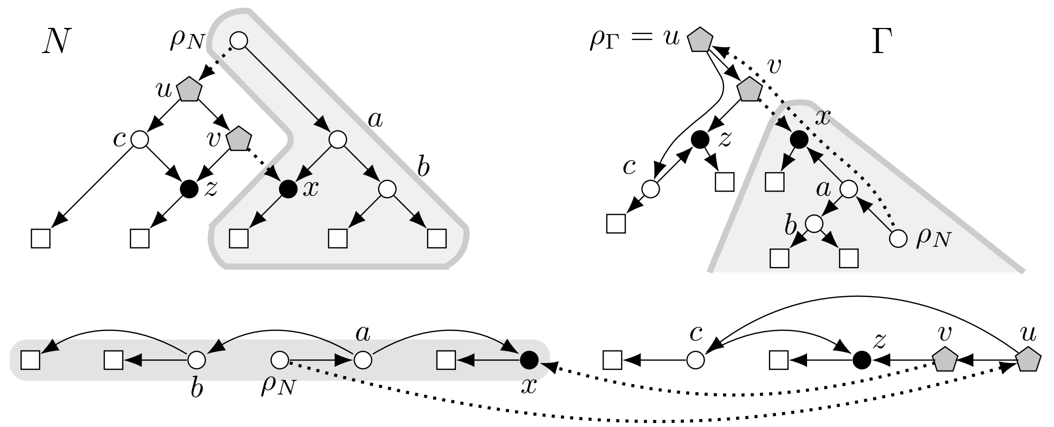

Fig. 3 Example of a network N (left) with a linear order σ of its nodes (below) as well as their canonical tree Γ^σ (right) whose arcs are not drawn (the arcs of N are drawn in their stead). Reticulations are black, leaves are boxes. For the first (wrt. σ) reticulation x, the set $V(\Gamma^\sigma_x)$ is marked (gray area) and equals $\sigma[1..x]$ in this example. Further, the arcs in $A_x(N)$ are dotted and the nodes in $\mathrm{YW}^\Gamma_x = \mathrm{ZW}^\sigma_x$ are gray pentagons. Note that $x \overset{N,\sigma}{\rightsquigarrow} \rho_N$ but neither $\rho_N \overset{N,\sigma}{\rightsquigarrow} x$ (since $x \notin \sigma[1..\rho_N]$) nor $z \overset{N,\sigma}{\rightsquigarrow} x$ (since x is not weakly connected to z in $N[\sigma[1..z]]$)

removing all nodes $x \in V(N) \setminus Z$ from N (together with their incident arcs) and, for any $v \in V(N)$, the network $N_v := N[\{w \mid w \leq_N v\}]$ is called the subnetwork *rooted at v*.

An alternative formulation of treewidth

In this section, we give an alternative definition of the *treewidth*, which allows to tackle the small parsimony problem for networks in a simpler and more intuitive way. Note that this alternative definition is known in the FPT community (Dendris et al. [24] call it the "support" of a vertex with respect to an ordering while, when referring to Arnborg [25]) and Mescoff et al. [26], call it "tree vertex separation"). However, since in these works its connection to treewidth is mostly touched in passing, we felt the need to prove it explicitly here.

Since tree decompositions are agnostic to edge directions, all results in this section are stated for undirected graphs G instead of networks N,. Keeping in mind that the framework is to be applied to phylogenetic networks, all examples will be made with DAGs while, for the sake of versatility, all results are stated for undirected graphs. The reader may simply ignore the edge directions in the examples as all undirected graphs will be underlying undirected graphs of some DAGs.

For a linear ordering σ of the nodes of an undirected graph G and any $x \in V(G)$, we write $y \leq_\sigma x$ for all nodes y preceeding x in σ (including x itself) and let $\sigma[1..x]$ denote the restriction of σ to these nodes. We write $x \overset{G,\sigma}{\rightsquigarrow} y$ if x and y are connected in $G[\sigma[1..x]]$ (see Fig. 3 for an example). Note that $\overset{G,\sigma}{\rightsquigarrow}$ is a partial order on $V(G)$. We consider nodes outside $\sigma[1..v]$ that have an edge to the parts of $\sigma[1..v]$ that are connected to v in $G[\sigma[1..v]]$. We denote these nodes by ZW^σ_v and their number by zw^σ_v.

Definition 1 Let σ be a linear order of the nodes of an undirected graph G and let $v \in V(G)$. Then,

$$\mathrm{ZW}^\sigma_v := \{u >_\sigma v \mid \exists_{w \in \sigma[1..v]} uw \in E(G) \wedge v \overset{G,\sigma}{\rightsquigarrow} w\}$$
$$\text{and } \mathrm{zw}^\sigma_v := |\mathrm{ZW}^\sigma_v|.$$

We abbreviate $\mathrm{zw}(\sigma) := \max_v \mathrm{zw}^\sigma_v$ and $\mathrm{zw}(G) := \min_\sigma \mathrm{zw}(\sigma)$ and we refer to the transitive reduction of the directed graph $(V(G), \{uv \in V(G)^2 \mid u \overset{G,\sigma}{\rightsquigarrow} v\})$ as the *canonical tree* Γ^σ of σ for G (we will see below that Γ^σ is a rooted tree; see Fig. 3).

In the following, we say that a rooted tree Γ on $V(G)$ *agrees* with an undirected graph G if, for all $uv \in E(G)$ either $u <_\Gamma v$ or $v <_\Gamma u$. We also extend the definition of $\overset{G,\sigma}{\rightsquigarrow}$ to such trees by writing $u \overset{G,\Gamma}{\rightsquigarrow} v$ if u and v are connected in $G[\Gamma_u]$. In analogy to Definition 1, $\overset{G,\Gamma}{\rightsquigarrow}$ gives rise to a set YW^Γ_v containing the nodes "above" v in Γ that have a edge in G to a node "below" v in Γ.

Definition 2 (see Fig. 3) Let G be an undirected graph and let Γ agree with G. For each $v \in V(G)$, we define

$$\mathrm{YW}^\Gamma_v := \{u >_\Gamma v \mid \exists_{w \leq_\Gamma v} uw \in E(G)\} \quad \text{and} \quad \mathrm{yw}^\Gamma_v := |\mathrm{YW}^\Gamma_v|.$$

Then, we abbreviate $\mathrm{yw}(\Gamma) := \max_v \mathrm{yw}^\Gamma_v$ and $\mathrm{yw}(G) := \min_\Gamma \mathrm{yw}(\Gamma)$.

Note that the path P resulting from traversing σ from right to left is a rooted tree agreeing with G. However, $\mathrm{yw}(P)$ is expected to be large for this choice. Indeed, we can show that the most "refined" trees Γ have the smallest $\mathrm{yw}(\Gamma)$.

Lemma 1 *Let Γ and Γ' be rooted trees agreeing with an undirected graph G and let $\leq_{\Gamma'}$ be a subset of $\leq_\Gamma$, that is, $x \leq_{\Gamma'} y \Rightarrow x \leq_\Gamma y$ for all $x, y \in V(G)$. Then, $\mathrm{yw}(\Gamma') \leq \mathrm{yw}(\Gamma)$.*

Proof Let $x \in V(G)$ and let $y \in \mathrm{YW}_x^{\Gamma'}$, that is, $y >_{\Gamma'} x$ and there is some $z \leq_{\Gamma'} x$ with $yz \in E(G)$. Since $\leq_\Gamma$ is a superset of $\leq_{\Gamma'}$, we have $y >_\Gamma x \geq z$, implying $y \in \mathrm{YW}_x^\Gamma$. □

The following lemma proves a number of interesting properties relating σ and Γ^σ such as Γ^σ being a rooted tree whose descendant relation is a refinement of $\leq_\sigma$, culminating in the equality of ZW_x^σ and $\mathrm{YW}_x^{\Gamma^\sigma}$ for all x.

Lemma 2 *Let σ be a linear order of the nodes of a connected undirected graph G and let Γ^σ be its canonical tree. Then,*

(*a*) *for each u and v with $v \leq_{\Gamma^\sigma} u$, we have $v \leq_\sigma u$,*
(*b*) *for each $u, v \in V(G)$, we have $v \leq_{\Gamma^\sigma} u$ if and only if $u \overset{G,\sigma}{\rightsquigarrow} v$,*
(*c*) *Γ^σ is connected,*
(*d*) *Γ^σ is rooted at the last vertex r of σ,*
(*e*) *Γ^σ is a tree,*
(*f*) *for all $uv \in E(G)$ with $v <_\sigma u$, we have $v <_{\Gamma^\sigma} u$,*
(*g*) *Γ^σ agrees with G, and*
(*h*) *$\mathrm{YW}_x^{\Gamma^\sigma} = \mathrm{ZW}_x^\sigma$ for all $x \in V(G)$.*
(*i*) *For each arc $xy \in A(\Gamma^\sigma)$, Γ_y^σ contains a neighbor of x in G.*
(*j*) *Each $x \in V(G)$ has at most as many children in Γ^σ as it has neighbors in G.*

Proof (a), (b): We show for all vertices w on a u-v-path p in Γ^σ that $w \leq_\sigma u$ and $u \overset{G,\sigma}{\rightsquigarrow} w$. The base case $w = u$ holds trivially. For the induction step, let q preceed w in p. Since Γ^σ contains the arc qw, Definition 1 implies $q \overset{G,\sigma}{\rightsquigarrow} w$ and, since $q \leq_\sigma u$ by induction hypothesis, $w \leq_\sigma q \leq_\sigma u$ and $u \overset{G,\sigma}{\rightsquigarrow} w$. For the reverse direction of (b), note that, by Definition 1, uv is an arc of the DAG of which Γ^σ is the transitive reduction.

(c),(d): Since G is connected, each $x \in V(G)$ has an r-x-path in $G = G[\sigma[1..r]]$, implying $r \overset{G,\sigma}{\rightsquigarrow} x$. Thus, (b) implies that Γ^σ is connected and rooted at r.

(e): To prove that Γ^σ is a tree, assume there is a vertex $x \in V(G)$ with two distinct parents y and z in Γ^σ. Without loss of generality, let $y <_\sigma z$. By (b), $y \overset{G,\sigma}{\rightsquigarrow} x$ and $z \overset{G,\sigma}{\rightsquigarrow} x$, implying that $\sigma[1..y]$ contains a y-x-path p_y in G and $\sigma[1..z]$ contains a z-x-path p_z in G. Since $\sigma[1..y] \subsetneq \sigma[1..z]$ the concatenation of p_z with (the reverse) of p_y is a path in G whose nodes are in $\sigma[1..z]$. Thus, $z \overset{G,\sigma}{\rightsquigarrow} y$, implying $y \leq_{\Gamma^\sigma} z$ and, since $zx \in A(\Gamma^\sigma)$, this contradicts Γ^σ being a transitive reduction.

(f): Note that $u \overset{G,\sigma}{\rightsquigarrow} v$, implying $v \leq_{\Gamma^\sigma} u$ by (b).

(g): For each $uv \in E(G)$, either $u <_\sigma v$, implying $u \leq_{\Gamma^\sigma} v$, or $v <_\sigma u$, implying $v \leq_{\Gamma^\sigma} u$ (both by (f)).

(h) "$\subseteq$": Let $x \in V(G)$ and let $y \in \mathrm{YW}_x^{\Gamma^\sigma}$. By Definition 2, $y >_{\Gamma^\sigma} x$ (implying $y >_\sigma x$ by (a)) and there is some $z \leq_{\Gamma^\sigma} x$ (implying $z \leq_\sigma x$ by (a)) with $yz \in E(G)$. Then, by (b), $x \overset{G,\sigma}{\rightsquigarrow} z$. But then, $y \in \mathrm{ZW}_x^\sigma$ by Definition 1.

(h) "$\supseteq$": Let $x \in V(G)$ and let $y \in \mathrm{ZW}_x^{\Gamma^\sigma}$, that is, $x <_\sigma y$ and there is some $z \in \sigma[1..x]$ with $x \overset{G,\sigma}{\rightsquigarrow} z$ and $yz \in E(G)$. Then, $z \leq_\sigma x <_\sigma y$. By (b), $z \leq_{\Gamma^\sigma} x$ and, by (f), $z \leq_{\Gamma^\sigma} y$. Thus, as Γ^σ is a tree (by (e)), x and y are not unrelated in Γ^σ. Moreover, $y \nleq_\sigma x$ implies $y \nleq_{\Gamma^\sigma} x$ by (b) and, thus, $x <_{\Gamma^\sigma} y$. Together with $z \leq_{\Gamma^\sigma} x$ and $yz \in E(G)$, this implies $y \in \mathrm{YW}_x^{\Gamma^\sigma}$.

(i) By (b), G contains an x-y-path p whose vertices are in $\sigma[1..x]$ and, thus, $x \overset{G,\sigma}{\rightsquigarrow} v$ for all vertices v on p. We show $u \leq_{\Gamma^\sigma} y$ for all u on p except x, starting with the obvious $y \leq_{\Gamma^\sigma} y$. Then, this implies that the second vertex on p, which is a neighbor of x in G, is in Γ_y^σ. Let $v \leq_{\Gamma^\sigma} y$ be a vertex on p and let u be the predecessor of v in p. If $u = x$ then we are done, so suppose $u \neq x$. Further, by (f), either $u <_\Gamma^\sigma v \leq_\Gamma^\sigma y$, implying the claim directly, or $v <_\Gamma^\sigma u$, implying that u is on an x-v-path in Γ^σ. By (e) there is only one such path and it starts with $(x, y, \ldots)$ and, since $u \neq x$, this implies $u \leq_\Gamma^\sigma y$.

(j) is immediate from (i) combined with (e). □

In order to show that $\mathrm{zw}(G)$ and $\mathrm{yw}(G)$ coincide, we need to "normalize" some aspects of the structure of agreeing trees. To this end, we use the following operation on rooted trees which can be interpreted as contracting a set of unwanted nodes upwards. Formally, for a rooted tree T and for $X \subset V(T)$ that does not contain the root r of T, we let $T \uparrow X$ denote the result of (1) replacing each arc uv with $uv \cap X = \{u\}$ with the arc wv where w is the lowest ancestor of u that is not in X, and (2) removing all nodes in X from T. Note that $T \uparrow X$ may have strictly larger out-degree than T, but does not create new ancestor-descendant relations.

Observation 1 Let T be a tree, let $X \subseteq V(T)$ not contain its root, and let $u, v \in V(T \uparrow X)$ with $u \leq_{T\uparrow X} v$. Then, $u \leq_T v$.

Lemma 3 *Let Γ be a rooted tree agreeing with an undirected graph G. Then, there is some rooted tree Γ^* agreeing with G such that $\mathrm{yw}(\Gamma^*) \leq \mathrm{yw}(\Gamma)$ and, for all $u, v \in V(G)$ with $v \leq_{\Gamma^*} u$, we have $u \overset{G,\Gamma^*}{\rightsquigarrow} v$.*

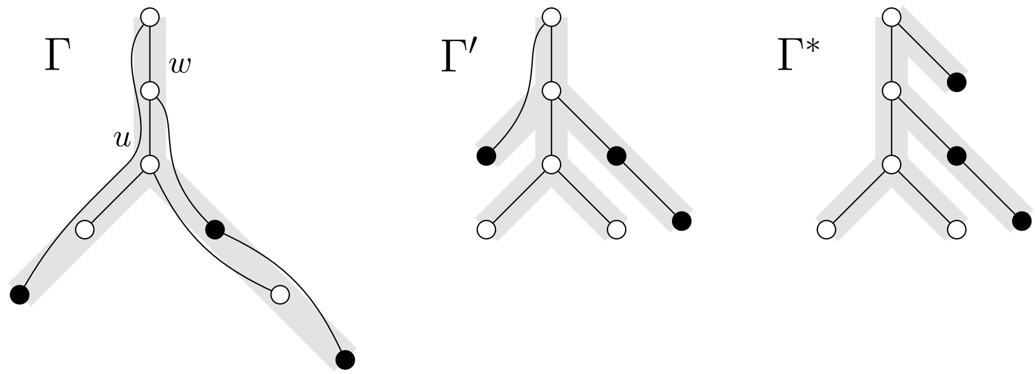

Fig. 4 Example for the construction of Γ' (middle) from Γ (left) in Lemma 3. Repeated application yields Γ^* (right), for which $v \le_{\Gamma^*} u \Rightarrow u \overset{G,\Gamma^*}{\rightsquigarrow} v$. The rooted trees Γ, Γ', and Γ^* are drawn with thick, gray lines. Thin, black lines are edges of G. For the indicated node u, the black nodes are in X, that is, they are below u in Γ but not connected to u in $G[\Gamma_u]$

Proof Let $u \in V(G)$ such that $X := \{v <_\Gamma u \mid u \overset{G,\Gamma}{\not\rightsquigarrow} v\} \neq \emptyset$. We will modify Γ into Γ' with $\text{yw}(\Gamma') \le \text{yw}(\Gamma)$ such that Γ' agrees with G and the relation $\le_{\Gamma'}$ is a strict subset of $\le_\Gamma$. To this end, note that u has a parent w in Γ as, otherwise, $G[\Gamma_u] = G$, implying $X = \emptyset$. Then, Γ' results from Γ by (see Fig. 4)

1. replacing Γ by $\Gamma \uparrow (\Gamma_u \setminus X)$ and
2. dangling $\Gamma_u \uparrow X$ from w.

First, we show that Γ' agrees with G. To this end, let $xy \in E(G)$ and let x and y be unrelated in Γ'. If neither x nor y are in Γ_u then, by construction of Γ', they are also unrelated in Γ, contradicting that Γ agrees with G. So, without loss of generality, suppose $x \le_\Gamma u$. Since $xy \in E(G)$ and Γ is a tree agreeing with G, we thus know that u and y are not unrelated in Γ. If $u <_\Gamma y$, then $w \le_\Gamma y$ and, thus, $x \le_{\Gamma'} y$. Thus, suppose $y \le_\Gamma u$. Clearly, if $x, y \in X$ or $x, y \notin X$, then x and y are also unrelated in Γ, contradicting its agreement with G. Thus, without loss of generality, suppose $x \in X$ and $y \notin X$, that is, $u \overset{G,\Gamma}{\not\rightsquigarrow} x$ and $u \overset{G,\Gamma}{\rightsquigarrow} y$, contradicting $xy \in E(G)$.

Second, we show that $\le_{\Gamma'}$ is a strict subset of $\le_\Gamma$. To this end, let $xy \in A(\Gamma')$ and assume towards a contradiction that $y \not<_\Gamma x$. Clearly, if $x \not\le_{\Gamma'} w$, then $xy \in A(\Gamma)$ contradicting $y \not<_\Gamma x$. Further, if $x = w$, then either $y \in X$ or y is a child of w in Γ, all of which imply $y <_\Gamma x$. Thus, $x <_{\Gamma'} w$. Since $xy \cap X = \{x\}$ or $xy \cap X = \{y\}$ contradicts $xy \in A(\Gamma')$, we have $x, y \in X$ or $x, y \notin X$. But then, $y <_\Gamma x$ by Observation 1. Thus, $\le_{\Gamma'}$ is a subset of $\le_\Gamma$ and it is strict since we have $v \le_\Gamma u$ and $v \not\le_{\Gamma'} u$ for all $v \in X \neq \emptyset$.

Third, $\text{yw}(\Gamma') \le \text{yw}(\Gamma)$ follows by Lemma 1. □

Lemma 4 *Let Γ be a tree agreeing with a graph G and let p be a non-empty path in G. Then, p contains a unique maximum u with respect to Γ, that is, $v \le_\Gamma u$ for all vertices v of p.*

Proof Let x on p be maximal with respect to Γ (that is, for all z on p, we have $x \not<_\Gamma z$) and assume towards a contradiction that there is another vertex $y \neq x$ on p that is maximal w.r.t. Γ. Without loss of generality, let x precede y in p and let p_{xy} denote the unique x-y-subpath of p. Since $y \not\le_\Gamma x$, there is an edge $st \in E(G)$ on p_{xy} with $s \le_\Gamma x$ and $t \not\le_\Gamma x$. Hence, $t \not\le_\Gamma s$. Further, $s \not\le_\Gamma t$ since, otherwise, the unique t-s-path in Γ contains x, contradicting its maximality. But then Γ does not agree with G. □

Lemma 5 *Let G be a graph. Then,* $\text{zw}(G) = \text{yw}(G)$.

Proof "$\ge$": Let σ be an ordering of $V(G)$ such that $\text{zw}(\sigma) = \text{zw}(G)$. By Lemma 2(h), we have $\text{zw}(\sigma) = \text{yw}(\Gamma\sigma)$ for the canonical extension tree Γ^σ of σ. Thus, $\text{zw}(G) = \text{zw}(\sigma) = \text{yw}(\Gamma\sigma) \ge \text{yw}(G)$.

"$\le$": Let Γ be some rooted tree agreeing with G such that $\text{yw}(\Gamma) = \text{yw}(G)$. By Lemma 3, we may assume

$$\forall_{u,v \in V(G)} u \le_\Gamma v \Rightarrow v \overset{G,\Gamma}{\rightsquigarrow} u. \tag{1}$$

Let σ be any ordering of $V(G)$ obtained by repeatedly picking and removing any leaf of Γ. □

Claim 1 *For each $u, v \in V(G)$, we have $u \le_\Gamma v$ if and only if $v \overset{G,\sigma}{\rightsquigarrow} u$.*

Proof First, note that all nodes below v in Γ are chosen before v, so $\Gamma_v \subseteq \sigma[1..v]$.

"$\Rightarrow$": Let $u \leq_\Gamma v$, that is, $u \in \Gamma_v$, implying $u \leq_\sigma v$. By (1), v is connected to u in $G[\Gamma_v]$ and, as $\Gamma_v \subseteq \sigma[1..v]$, also in $G[\sigma[1..v]]$.
"$\Leftarrow$": Let p be a v-u-path in $G[\sigma[1..v]]$. By Lemma 4, p has a unique maximum w in Γ. Hence, $v \leq_\Gamma w$ and, by "$\Rightarrow$", we have $v \leq_\sigma w$. Since p lives entirely in $G[\sigma[1..v]]$, that is, $V(p) \subseteq \sigma[1..v]$, we also have $w \leq_\sigma v$. Thus, $v = w$ and, since $u \in V(p)$, we have $u \leq_\Gamma w = v$ by maximality of w. □

To prove the lemma, we show $\mathrm{YW}_x^\Gamma \supseteq \mathrm{ZW}_x^\sigma$ for each $x \in V(G)$. Let $y \in \mathrm{ZW}_x^\sigma$, that is $y >_\sigma x$ and there is some $z \in \sigma[1..x]$ with $yz \in E(G)$ and $x \overset{G,\sigma}{\rightsquigarrow} z$. By Claim 1, $z \leq_\Gamma x$. Further, as $yz \in E(G)$ and Γ agrees with G, y and z are not unrelated in Γ and, since $z \leq_\Gamma x$, neither are x and y. Since $y <_\Gamma x$ implies $y <_\sigma x$ by Claim 1, contradicting $y >_\sigma x$, we conclude $x <_\Gamma y$. Together with $z \leq_\Gamma x$ and $yz \in E(G)$, this implies $y \in \mathrm{YW}_x^\Gamma$.

Having shown that the notion of $\mathrm{zw}(G)$ and $\mathrm{yw}(G)$ are equivalent, we can now turn our attention to the treewidth. In particular, we introduce (nice) tree-decompositions and use their properties to show that the treewidth of any undirected graph G equals $\mathrm{yw}(G)$.

Definition 3 (see Fig. 5) Let G be an undirected graph and let T be a rooted tree whose vertices are associated to subsets of $V(G)$ by a function $B : V(T) \to 2^{V(G)}$ such that

(a) for each $uv \in E(G)$, there is some $x \in V(T)$ with $u, v \in B(x)$ and
(b) for each $v \in V(G)$, the nodes $x \in V(T)$ with $v \in B(x)$ are weakly connected in T.

We call (T, B) a *tree decomposition* of G and its *width* is $\mathrm{tw}(T,B) := \max_{x \in V(T)} \mathrm{tw}_x(T,B)$ with $\mathrm{tw}_x(T,B) := |B(x)| - 1$. We call $\mathrm{tw}(G) := \min_{T,B} \mathrm{tw}(T,B)$ the *treewidth* of G. We call (T, B) *nice* if T is binary and all $x \in V(T)$ fall into one of the following categories

"leaf": x is a leaf of T and $B(x) = \varnothing$,
"root": x is the root of T and $B(x) = \varnothing$,
"introduce v**"**: x has a single child y in T and $B(y) = B(x) - v$,
"forget v**"**: x has a single child y in T and $B(x) = B(y) - v$,
"join": x has two children y and z and $B(x) = B(y) = B(z)$.

As stated at the beginning of the section, recall that, while tree decompositions are defined for undirected

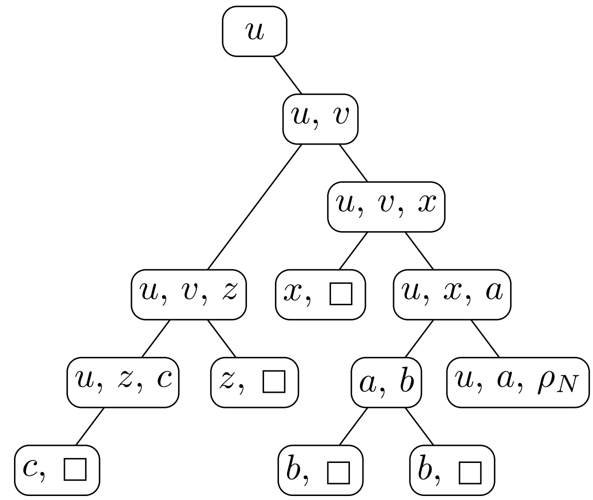

Fig. 5 The tree decomposition (Γ, B) for the network N given in Fig. 3 constructed in the "$\geq$"-part of Lemma 1. Leaves are represented by boxes instead of their names. Note that (Γ, B) is *not* a nice tree decomposition

graphs, we may talk about tree decompositions of DAGs, meaning tree decompositions of their underlying undirected graphs. Note that all graphs G have a nice tree decomposition with $|V(T)| \in O(\mathrm{tw}(G) \cdot |G|)$ and width $\mathrm{tw}(G)$ [27]. Further, since all bags of (T, B) containing a vertex v of G are connected, we can observe the following.

Observation 2 Let (T, B) be a nice tree decomposition for an undirected graph G and let $v \in V(G)$. Then, T contains a single "forget v"-node x and $y <_T x$ for all y with $v \in B(y)$.

Proposition 1 *Let G be an undirected graph. Then,* $\mathrm{yw}(G) = \mathrm{tw}(G)$. *Further, given a tree decomposition* (T, B) *for G, we can compute a tree* Γ *agreeing with G such that* $\mathrm{yw}(\Gamma) = \mathrm{tw}(T, B)$ *in linear time.*

Proof "$\leq$": Let (T, B) be a nice tree decomposition for G of width $\mathrm{tw}(G)$ and let $F \subset V(T)$ denote the set of all "forget"-nodes in T (noting that F contains the root of T). We define Γ as the transitive reduction of $(F, >_T \cap (F \times F))$.[2] Note that $u \leq_\Gamma v \iff u \leq_T v$ for all $u, v \in F$ and, by Observation 2, $V(\Gamma) = F = V(G)$.

First, we show that Γ agrees with G. To this end, let $uv \in E(G)$ and let $f_u, f_v \in F$ denote the unique "forget u" and "forget v"-nodes in T, which are distinct since T is nice. By Definition 3(a), there is a node $q \in V(T)$ with $u, v \in B(q)$ and, by Observation 2,

[2] Intuitively, Γ can be obtained from T by contracting all nodes in $V(T) \setminus F$ onto their respective parents and identifying all nodes $x \in F$ with the vertex $v \in V(G) \setminus B(x)$ of G that is forgotten in x.

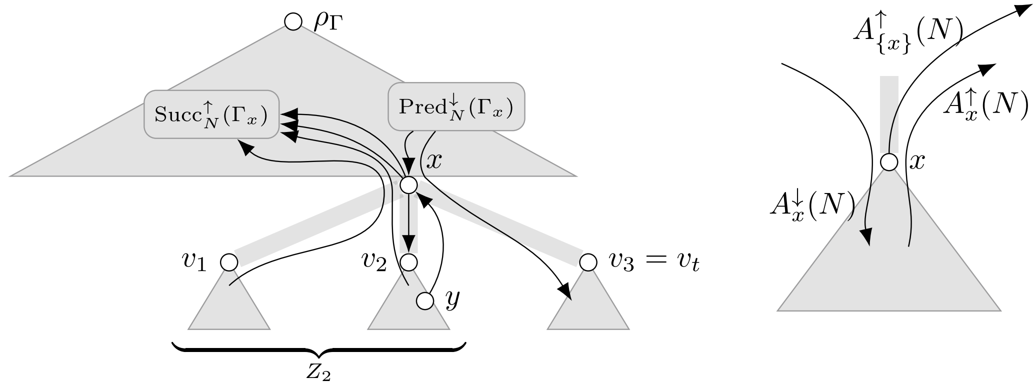

Fig. 6 A tree Γ is depicted in gray and some arcs of *N* are depicted in black. Recall that *t* is the number of children of *x* and $Z_i := \bigcup_{1\le j\le i} \Gamma_{v_j}$. Note that $x \in \mathrm{Succ}_N^{\uparrow}(Z_2) \setminus \mathrm{Succ}_N^{\uparrow}(\Gamma_x)$ since *x* is an ancestor of a node of Γ_{v_2} in *N*. Note that *x* is a reticulation of *N* with parents *y* (drawn) and *z* (not drawn) with $y <_\Gamma v_2 <_\Gamma x <_\Gamma z$. Thus, $z \in \mathrm{Pred}_N^{\downarrow}(x)$ but $y \in \mathrm{Pred}_N^{\uparrow v_2}(x) \subseteq \mathrm{Pred}_N^{\uparrow}(x)$. Finally, note that $\mathrm{YW}_x^\Gamma = \mathrm{Pred}_N^{\downarrow}(\Gamma_x) \cup \mathrm{Succ}_N^{\uparrow}(\Gamma_x)$ and $\bigcup_{i\le t} \mathrm{YW}_{v_i}^\Gamma \subseteq \mathrm{YW}_x^\Gamma \uplus \{x\}$

$q <_T f_u, f_v$. Thus, f_u and f_v are not unrelated in T and, thus, neither in Γ.

Second, we show for all $v \in \Gamma$ and the unique "forget v"-node f_v in T that $\mathrm{YW}_v^\Gamma \subseteq B(f_v)$. Let $u \in \mathrm{YW}_v^\Gamma$, that is, $u >_\Gamma v$ and there is some $w \le_\Gamma v$ such that $uw \in E(G)$ (note that $w \neq u$ but $w = v$ is possible). Let f_u and f_w be the unique "forget u" and "forget w"-nodes in T, which are distinct since T is nice. Then, $w \le_\Gamma v <_\Gamma u$ and, since $f_u, f_w \in F$, we also have $f_w \le_T f_v <_T f_u$. Since $uw \in E(G)$, Definition 3(a) implies that there is a node q of T with $u, w \in B(q)$ and, by Observation 2, $q <_T f_u, f_w$. Then, by Definition 3(b), $u \in B(x)$ for all x with $q \le_T x <_T f_u$ and, since $q <_T f_w \le_T f_v <_T f_u$, we have $u \in B(f_v)$. As u was chosen arbitrary, we conclude $\mathrm{YW}_v^\Gamma \subseteq B(f_v)$. Hence, $\mathrm{yw}(G) \le |\mathrm{YW}_v^\Gamma| \le |B(f_v)|$ and, since f_v has a child x with $B(x) = B(f_v) \cup \{v\}$, we know $|B(f_v)| = |B(x)| - 1 \le \mathrm{tw}(T, B) = \mathrm{tw}(G)$.

"≥": Let Γ be a tree with $\mathrm{yw}(\Gamma) = \mathrm{yw}(G)$ that agrees with G. For all $u \in V(G)$, we define $B(u) := \mathrm{YW}_u^\Gamma \cup \{u\}$ and show that (Γ, B) is a tree-decomposition for G noting that its width is $\mathrm{yw}(\Gamma) = \mathrm{yw}(G)$ (see example in Fig. 5).

First, to prove Definition 3(a), let $uv \in E(G)$. Since Γ agrees with G, either $u <_\Gamma v$ or $v <_\Gamma u$. Without loss of generality, suppose the latter. Then, $u \in \mathrm{YW}_v^\Gamma$ by Definition 2 (using $w = v$), implying that $uv \in B(v)$.

Second, let $u, v \in V(G)$ be distinct such that $u \in B(v) = \mathrm{YW}_v^\Gamma \cup \{v\}$, implying $u \in \mathrm{YW}_v^\Gamma$ since $u \neq v$. By Definition 2, there is some $w \le_\Gamma v$ such that $uw \in E(G)$ and $v <_\Gamma u$, implying that Γ contains a unique u-v-path p. To show Definition 3(b), it suffices to prove $u \in B(x)$ for all $x \in V(p)$ (since v has been chosen arbitrarily, a path with these properties exists for all v' with $u \in B(v')$, so they all contain the node u and are, thus, connected). For $x = u$ this follows by definition of $B(u)$. Otherwise, $x <_\Gamma u$ since $x \in V(p)$. But then, $w \le_\Gamma v \le_\Gamma x <_\Gamma u$ and $uw \in E(G)$, implying $u \in \mathrm{YW}_x^\Gamma \subseteq B(x)$. □

Parsimony

Notation Large parts of this work are in context of a rooted tree Γ on the node set $V(N)$ of a given phylogenetic network N (see Fig. 6). Specifically for the tree Γ, we permit ourselves to abbreviate $V(\Gamma_x)$ to Γ_x to increase readability. In such context, we additionally define the following sets for any nodes $y, z \in V(N)$: $\mathrm{Pred}_N^{\uparrow y}(z) := \mathrm{Pred}_N(z) \cap \Gamma_y$ and $\mathrm{Pred}_N^{\downarrow y}(z) := \mathrm{Pred}_N(z) \setminus \Gamma_y$ denote the respective *predecessors* of z in N that are or are not in Γ_y. Likewise, $\mathrm{Succ}_N^{\uparrow y}(z) := \mathrm{Succ}_N(z) \cap \Gamma_y$ and $\mathrm{Succ}_N^{\uparrow y}(z) := \mathrm{Succ}_N(z) \setminus \Gamma_y$ denote the respective *successors* of z in N that are or are not in Γ_y – note that the arrow in the notation indicates the direction of the arc between z and the members of the set when drawing Γ top-down. If $z = y$, we drop y and simply write $\mathrm{Pred}_N^{\downarrow}(z)$, $\mathrm{Pred}_N^{\uparrow}(z)$, $\mathrm{Succ}_N^{\downarrow}(z)$, and $\mathrm{Succ}_N^{\uparrow}(z)$. We also abbreviate $\mathrm{Pred}_N^{\downarrow}(z) \cap R(G) =: \mathrm{Pred}_N^{R\downarrow}(z)$ and $\mathrm{Succ}_N^{\uparrow}(z) \cap R(G) =: \mathrm{Succ}_N^{R\uparrow}(z)$ as well as $\mathrm{Pred}_N^{\downarrow}(z) \setminus R(G) =: \mathrm{Pred}_N^{T\downarrow}(z)$ and $\mathrm{Succ}_N^{\uparrow}(z) \setminus R(G) =: \mathrm{Succ}_N^{T\uparrow}(z)$. All these functions generalize to sets $Z \subseteq V(N)$ (for example, $\mathrm{Pred}_N(Z) := \bigcup_{z \in Z} \mathrm{Pred}_N(z) \setminus Z$). Further, for any $X \subseteq V(N)$, we define the sets of arcs of N

(a) from a node $u \in X$ to any ancestor of u in Γ as $A_X^{\uparrow}(N) := \{uw \in A(N) \mid u \in X \wedge u <_\Gamma w\}$ and
(b) to a node $u \in X$ from any ancestor of u in Γ as $A_X^{\downarrow}(N) := \{uw \in A(N) \mid w \in X \wedge w <_\Gamma u\}$.

For brevity, we abbreviate $A_X(N) := A^{\uparrow}_X(N) \cup A^{\downarrow}_X(N)$, $A^{\uparrow}_v(N) := A^{\uparrow}_{\Gamma_v}(N)$, $A^{\downarrow}_v(N) := A^{\downarrow}_{\Gamma_v}(N)$, and $A_v(N) := A_{\Gamma_v}(N)$.

Introduction to Parsimony Given states of a character, observed in extant species, as well as a species phylogeny, the small parsimony problem asks to infer states of the same character for all ancestral species such as to minimize the "parsimony score" of this assignment. This problem comes in three flavors called "hardwired", "softwired", and "parental" parsimony. Throughout this section, let C be a fixed finite set (a "character"). For convenient use of the $\trianglelefteq$-relation, let C be an anti-chain (that is, for each $x, y \in C$, we have $x \leq y$ only if $x = y$). Formally, for a phylogeny N and a function $\phi : V(N) \to 2^C$, we define the hardwired and softwired parsimony score as

$$\mathrm{par}^H_N(\phi) := \min_{\psi: V(N)\to C,\ \psi \trianglelefteq \phi} \sum_{uv\in A(N)} \delta_\psi(u, v)$$

$$\mathrm{par}^S_N(\phi) := \min_{\substack{\psi\,:\,V(N)\to C,\ \psi \trianglelefteq \phi\\ T\in\mathcal{S}(N)}} \sum_{uv\in A(T)} \delta_\psi(u, v).$$

The "parental parsimony" is defined using "parental trees" but, in this work, we use the equivalent formulation using lineage functions [12].

Definition 4 A *lineage function* for a phylogeny N is any function $f : V(N) \to 2^C$. The *cost* of f is $\mathrm{cost}_{(f)} := \sum_{v\in V(N)} \mathrm{cost}_f(v)$ where

$$\mathrm{cost}_f(v) := |f(v) \setminus \bigcup_{u\in\mathrm{Pred}(v)} f(u)| + \begin{cases} -1 & \text{if } v = \rho_N \text{ and } |f(v)| = 1 \\ 0 & \text{if } v \neq \rho_N \text{ and } |f(v)| \leq \sum_{u\in\mathrm{Pred}(v)} |f(u)| \\ \infty & \text{otherwise} \end{cases}$$

Given N and a function $\phi : V(N) \to 2^C$, we denote the set of all lineage functions f on N with $f \trianglelefteq \phi$ as $\mathcal{LF}_{N,\phi}$. Finally, the *parental parsimony score* is

$$\mathrm{par}^P_N(\phi) := \min_{f\in\mathcal{LF}_{N,\phi}} \mathrm{cost}_{(f)} \tag{2}$$

For each of the presented variants, we give a dynamic programming formulation using a given tree Γ that agrees with the undirected graph G underlying the input network and corresponds to Lemma 3, that is, each non-leaf x of Γ has a child v with $x \in \mathrm{YW}^{\Gamma}_v$. The running time of the resulting algorithm will depend on the width $\mathrm{yw}(\Gamma)$ of Γ (recalling that $\mathrm{yw}(\Gamma)$ coincides with the treewidth of G for optimal Γ).

As stated in the introduction, in this paper we focus on the case of analyzing a specific position in the genome. Since the function ϕ can associate several states to a same leaf, our definition permits to describe polymorphism in a population. While in our current formulation the algorithms "choose" an optimal state to associate to each leaf, the parental parsimony can be easily modified to explain *all* states of each leaf at the end of the run. This allows keeping the information on polymorphism in all steps of the algorithm (see "Parental parsimony"). Note also that ϕ can associate information to internal nodes, thus permitting the user to impose restrictions on the states associated to ancestral species.

In the presentation of the dynamic programming, a table entry $Q^y_x[z]$ means that x and y are considered fix for this table and z is a variable index. Further, tables $Q^{y_1}_{x_1}$ and $Q^{y_2}_{x_2}$ are independent of one another, allowing an implementation to forget $Q^{y_1}_{x_1}$ if it is no longer needed, even if $Q^{y_2}_{x_2}$ still is. In the following, for an anti-chain Y in Γ and a class $\mathcal{G}$ of subnetworks of N, a *Y-substitution system* of $\mathcal{G}$ is a series of subnetworks $(N^y)_{y\in Y}$ of N such that, for all $N' \in \mathcal{G}$, the digraph $(V(N), (A(N') \setminus \bigcup_{y\in Y} A_y(N')) \cup \bigcup_{y\in Y} A_y(N^y))$ is also in $\mathcal{G}$. Roughly, we can "swap out" the arcs in $A_y(N')$ for $A_y(N^y)$ for each $y \in Y$ without loosing membership in $\mathcal{G}$. Note that the N^y are not necessarily distinct, so a trivial Y-substitution system for $\{N'\}$ would be $(N')_{y\in Y}$. The formulations are based on the following lemma about independent sub-solutions, showing that an optimal solution (S, ψ) for a sub-network (of G) "below" an anti-chain Z in Γ is also optimal on any sub-network "below" an anti-chain Y in Γ that is itself "below" Z (among all solutions with ψ's behavior on $\bigcup_{y\in Y} \mathrm{YW}^{\Gamma}_y$).

Lemma 6 (see Fig. 7) *Let $Y, Z \subseteq V(N)$ be anti-chains in Γ such that $Y \subseteq \bigcup_{z\in Z} \Gamma_z$. Let $\mathcal{G}$ be a class of subnetworks of N and let $S \in \mathcal{G}$ and $\psi : V(N) \to C$ such that (a) $\sum_{z\in Z} \sum_{uw\in A_z(S)} \delta_\psi(u, w)$ is minimum among all such S and ψ. Let $(S^y)_{y\in Y}$ be a Y-substitution system for $\mathcal{G}$ and let $\psi_y : V(N) \to C$ for each $y \in Y$ such that (b) ψ_y and ψ coincide on YW^{Γ}_y. Then,*

$$\sum_{y\in Y} \sum_{uw\in A_y(S^y)} \delta_{\psi_y}(u, w) \geq \sum_{y\in Y} \sum_{uw\in A_y(S)} \delta_\psi(u, w).$$

Proof Towards a contradiction, assume that the lemma is false. We construct $\psi^* : V(N) \to C$ with

$$\psi^*(u) = \begin{cases} \psi_y(u) & \text{if } u \in \Gamma_y \text{ for any } y \in Y \\ \psi(u) & \text{otherwise} \end{cases}$$

Note that ψ^* and ψ coincide with ψ_y on YW^{Γ}_y for all $y \in Y$. Thus, $\delta_{\psi^*}(u, w) = \delta_{\psi_y}(u, w)$ if $uw \in A_y(S^*)$ for any $y \in Y$ and $\delta_{\psi^*}(u, w) = \delta_\psi(u, w)$, otherwise. Further, we construct a digraph $S^* := (V(N), (A(S)\setminus \bigcup_{y\in Y} A_y(S)) \cup \bigcup_{y\in Y} A_y(S^y))$ which is in $\mathcal{G}$ since $(S^y)_{y\in Y}$ is a Y-substitution system for $\mathcal{G}$. Since all S^y are subnetworks of N, we know that Γ agrees with S^*. Furthermore,

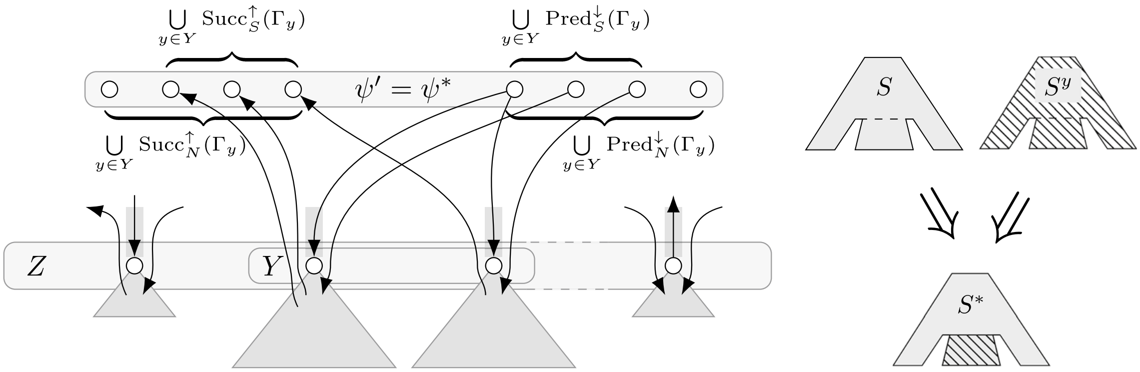

Fig. 7 Lemma 6 proves that any solution (S, ψ) that is optimal on sub-trees rooted at Z in Γ must also be optimal (among all solutions with ψ's behavior on $\bigcup_{y\in Y} \mathrm{YW}_y^\Gamma$ (gray box on top)) on all sub-trees of Γ that are rooted below Z (at Y). That is, no solution (S^y, ψ_y) can be better than (S, ψ) on the sub-network induced by Γ_y for any $y \in Y$. To prove this, a new solution (S^*, ψ^*) is constructed by replacing the sub-solution of (S, ψ) below Y by the sub-solutions (S^y, ψ_y) below Y

since $Y \subseteq \bigcup_{z\in Z} \Gamma_z$, we know that each $y \in Y$ has a $z \in Z$ with $y \leq_\Gamma z$. Thus,

$$
\begin{aligned}
\sum_{z\in Z} \sum_{uw\in A_z(S^*)} \delta_{\psi^*}(u,w) &= \sum_{z\in Z} \sum_{v\in\Gamma_z} \sum_{uw\in A_{\{v\}}(S^*)} \delta_{\psi^*}(u,w) \\
&= \sum_{z\in Z} \sum_{\substack{v\in\Gamma_z \\ v\notin\bigcup_{y\in Y}\Gamma_y}} \sum_{uw\in A_{\{v\}}(S^*)} \delta_{\psi^*}(u,w) + \sum_{y\in Y} \sum_{uw\in A_{\{v\}}(S^*)} \delta_{\psi^*}(u,w) \\
&= \sum_{z\in Z} \sum_{\substack{v\in\Gamma_z \\ v\notin\bigcup_{y\in Y}\Gamma_y}} \sum_{uw\in A_y(S)} \delta_{\psi}(u,w) + \sum_{y\in Y} \sum_{uw\in A_y(S^y)} \delta_{\psi_y}(u,w) \\
&\overset{\text{assumption}}{<} \sum_{z\in Z} \sum_{\substack{v\in\Gamma_z \\ v\notin\bigcup_{y\in Y}\Gamma_y}} \sum_{uw\in A_{\{v\}}(S)} \delta_{\psi}(u,w) + \sum_{y\in Y} \sum_{uw\in A_y(S)} \delta_{\psi}(u,w) \\
&= \sum_{z\in Z} \sum_{uw\in A_z(S)} \delta_{\psi}(u,w)
\end{aligned}
$$

contradicting optimality of S and ψ (that is, Lemma 6(a)) since $S^* \in \mathcal{G}$. □

Hardwired parsimony

To compute the hardwired parsimony score at a node v of N, we require knowledge of the character assigned to v and its neighbors. For all $u \in \mathrm{YW}_v^\Gamma$, we thus "guess" the character $\psi(u)$ assigned to u by an optimal assignment. In our dynamic programming, we scan Γ bottom-up, computing a table entry $T^{\mathcal{HW}}[x, \psi]$ for each $x \in V(\Gamma) = V(N)$ and each $\psi : \mathrm{YW}_x^\Gamma \to C$, containing the parsimony cost incurred by all arcs in $A_x(N)$, assuming that all nodes in YW_x^Γ receive their characters according to ψ. Note that $A_x(N) = \bigcup_i A_{v_i}(N) \cup A_{\{x\}}(N)$, where the v_i are the children of x in Γ. Thus, $T^{\mathcal{HW}}[x, \psi]$ can be calculated as follows.

Definition 5 Let Γ be a tree that agrees with N, let $x \in V(N)$ and let $\psi_x : \mathrm{YW}_x^\Gamma \to C$ with $\psi_x \trianglelefteq \phi$. Let $v_1, v_2, \ldots, v_t$ denote the children of x in Γ ($t = 0$ if x is a leaf). Then, we define a table entry

$$T^{\mathcal{HW}}[x,\psi_x] := \min_{c_x\in\phi(x)}\left(\sum_{1\le i\le t} T^{\mathcal{HW}}[v_i,\psi_x[x\to c_x]\mid_{\mathrm{YW}^\Gamma_{v_i}}] + \sum_{z\in\mathrm{Pred}^{\downarrow}_N(x)\cup\mathrm{Succ}^{\uparrow}_N(x)} \delta\,(c_x,\psi_x(z))\right) \tag{3}$$

Lemma 7 *Let $x\in V(N)$ and let $\psi_x:\mathrm{YW}^\Gamma_x\to C$ with $\psi_x\trianglelefteq\phi$. Let $\psi:V(N)\to C$ with $\psi_x\trianglelefteq\psi\trianglelefteq\phi$ such that ψ minimizes $\sum_{uw\in A_x(N)}\delta_{\,\psi}(u,w)$. Then,*

$$T^{\mathcal{HW}}[x,\psi_x] = \sum_{uw\in A_x(N)} \delta_{\,\psi}(u,w)$$

Proof Sketch. For "≥", we construct a mapping ψ' from mappings ψ_i that are optimal on $A_{v_i}(N)$ among all mappings with $\psi_i(x):=c_x$. This is possible since all such ψ_i coincide with ψ' and ψ_x on YW^Γ_x. By induction hypothesis, the cost of ψ' on $A_x(N)$ is $\sum_{1\le i\le t} T^{\mathcal{HW}}[v_i,\psi'\mid_{\mathrm{YW}^\Gamma_{v_i}}]+\sum_{uw\in A_{\{x\}}(N)}\delta_{\,\psi'}(u,w)$. Then, "≥" follows from optimality of ψ on $A_x(N)$.

For "≤", it suffices to show that the cost of ψ on $A_x(N)$ is equal to the result of setting $c_x:=\psi(x)$ in the right hand side of (3) (which is a valid choice for the minimum since $\psi(x)\in\phi(x)$). First, the cost of ψ on $A_{v_i}(N)$ is $T^{\mathcal{HW}}[v_i,\psi\mid_{\mathrm{YW}^\Gamma_{v_i}}]$ by independence of sub-solutions and the induction hypothesis. Second, the cost of ψ on $A^{\downarrow}_{\{x\}}(N)$ is $\sum_{z\in\mathrm{Pred}^{\downarrow}_N(x)}\delta\,(c_x,\psi_x(z))$ and the cost of ψ on $A^{\uparrow}_{\{x\}}(N)$ is $\sum_{z\in\mathrm{Succ}^{\uparrow}_N(x)}\delta\,(c_x,\psi_x(z))$ since ψ and ψ_x coincide on YW^Γ_x. □

In order to solve the hardwired parsimony problem given N, ϕ and Γ, all we have to do is compute $T^{\mathcal{HW}}[x,\psi_x]$ for each x bottom-up in Γ and each of the (at most) $|C|^{|\mathrm{YW}^\Gamma_x|}$ many choices of $\psi_x:\mathrm{YW}^\Gamma_x\to C$ with $\psi_x\trianglelefteq\phi$. Then, by Lemma 7, the hardwired parsimony score of N with respect to ϕ can be read from $T^{\mathcal{HW}}[\rho_\Gamma,\varnothing]$. To compute $T^{\mathcal{HW}}$, the sum over the children of x for all $x\in V(N)$ in (3) can be computed in amortized $O(|A(N)|)$ time and, with a bit of bookkeeping, it is possible to maintain the value of the second sum in (3) in $O(|A(N)|)$ amortized time per choice of ψ. Then the following holds:

Theorem 1 *Given a network N, some $\phi:V(N)\to 2^C$ and a tree Γ agreeing with N, the hardwired parsimony score of (N,ϕ) can be computed in $O(|C|^{\mathrm{yw}(\Gamma)+1}\cdot|A(N)|)$ time.*

Proposition 1 lets us turn tree decompositions of N into trees Γ agreeing with N, allowing us to replace $\mathrm{yw}(\Gamma)$ by $\mathrm{tw}(N)$, incurring an additional running time of $|N|\cdot 2^{O(\mathrm{tw}(N)^3)}$ [13].

Corollary 1 *Let (N,ϕ) be an instance of Hardwired Parsimony. Let $t\ge\mathrm{tw}(N)$ and let T be the time in which a width-t tree decomposition of N can be computed. Then, the hardwired parsimony score of (N,ϕ) can be computed in $O(T+|C|^{t+1}\cdot|A(N)|)$ time.*

Softwired parsimony

In contrast to the hardwired parsimony score, where the computation of the cost of the incident edges of a node x only required knowledge of the characters assigned to neighbors of x, computing the *softwired* score additionally requires knowledge of which parent of x remains a parent in the sought switching. A table entry $T^{\mathcal{SW}}[x,\ldots]$ contains the smallest combined cost of all arcs in $A_x(S)$ for a switching S of N minimizing this cost. To be able to compute an entry for $x\in V(N)$, we not only need to "guess" ψ_x but, additionally, some representation of the switching S. In particular, in S, no child of x may have another parent than x. However, since children of x in N may be above x in Γ, we have to "guess" which children of x in N are still children of x in S. Such a guess manifests itself as an additional index R^x of the dynamic programming table (note that we clearly only have to store this information for children of x that are reticulations). Indeed, this information has to be stored for all nodes considered below x who still have children in YW^Γ_x. Thus, we index our DP-table also by a subset $R^x\subseteq\mathrm{YW}^\Gamma_x\cap R(N)$ containing a reticulation $r\in R(N)$ if and only if Γ_x contains a parent v of r and vr is an arc of an optimal switching S for $N[\Gamma_x\cup\mathrm{YW}^\Gamma_x]$.

Definition 6 Let Γ be a tree that agrees with N, let $x\in V(N)$, let $\psi_x:\mathrm{YW}^\Gamma_x\to C$ with $\psi_x\trianglelefteq\phi$, and let $R^x\subseteq\mathrm{Succ}^{R\uparrow}_N(\Gamma_x)$. Let $v_1,v_2,\ldots,v_t$ denote the children of x in Γ ($t=0$ if x is a leaf in Γ). Then, set

$$T^{\mathcal{SW}}[x,\psi_x,R^x] := \min_{c_x\in\phi(x)} \min_{R^*\subseteq R^x\cap \mathrm{Succ}_N^{R\uparrow}(x)} \sum_{r\in R^*\cup \mathrm{Succ}_N^{T\uparrow}(x)} \delta\left(c_x,\psi_x(r)\right) + \min\begin{cases} Q^{\psi_x}_{x,c_x}[t,R^x\setminus R^*] + \min\limits_{y\in \mathrm{Pred}_N^{\downarrow}(x)} \delta\left(c_x,\psi_x(y)\right) & \text{if } \mathrm{Pred}_N^{\downarrow}(x)\neq\varnothing \\ Q^{\psi_x}_{x,c_x}[t,(R^x\setminus R^*)\cup(\{x\}\cap R(N))] & \text{if } \mathrm{Pred}_N^{\uparrow}(x)\neq\varnothing \end{cases} \tag{4}$$

where

$$Q^{\psi_x}_{x,c_x}[i,R'] := \begin{cases} \min\limits_{R^*\subseteq R'\cap \mathrm{Succ}_N^{R\uparrow}(\Gamma_{v_i})} Q^{\psi_x}_{x,c_x}[i-1,R'\setminus R^*] + T^{\mathcal{SW}}[v_i,\psi_i,R^*] & \text{if } i\neq 0 \\ 0 & \textit{if } i=0 \text{ and } R'=\varnothing \\ \infty & \textit{otherwise} \end{cases} \tag{5}$$

where $\psi_i := \psi_x[x\to c_x]\,|_{\mathrm{YW}^{\Gamma}_{v_i}}$ for all $i\leq t$. (Note how $Q^{\psi_x}_{x,c_x}[i,R']$ is used to assign the nodes in R^x to the v_i (with $v_0=x$) such that every node in R^x has a parent in some Γ_{v_i}).

In the following, for any anti-chain X in Γ and all $Z\subseteq\bigcup_{x\in X}\mathrm{YW}^{\Gamma}_x$, let $\mathcal{S}^{X\to Z}(N)$ denote the set of all switchings S of N with $\mathrm{Succ}_S^{R\uparrow}(X)=Z$.

Lemma 8 *Let Γ be a tree that agrees with N, let $x\in V(N)$, let $\psi_x:\mathrm{YW}^{\Gamma}_x\to C$ with $\psi_x\trianglelefteq\phi$, and let $R^x\subseteq \mathrm{Succ}_N^{R\uparrow}(\Gamma_x)$. If $\mathcal{S}^{\Gamma_x\to R^x}(N)=\varnothing$, then $T^{\mathcal{SW}}[x,\psi_x,R^x=\infty]$. Otherwise, let $S\in\mathcal{S}^{\Gamma_x\to R^x}(N)$ and $\psi:V(N)\to C$ such that (a) $\psi_x\trianglelefteq\psi\trianglelefteq\phi$ and (b) $\sum_{uw\in A_x(S)}\delta_{\psi}(u,w)$ is minimum among all such S and ψ. Then,*

$$T^{\mathcal{SW}}[x,\psi_x,R^x] = \sum_{uw\in A_x(S)} \delta_{\psi}(u,w). \tag{6}$$

Proof Sketch. Let us abbreviate $Z_i := \bigcup_{j\leq i}V(\Gamma_{v_j})$. We first show that the table Q does what we expect it to do.

Claim 2 $Q^{\psi_x}_{x,c_x}[i,R'] = \sum_{j\leq i}\sum_{uw\in A_{v_j}(S_i)}\delta_{\psi_i}(u,w)$ *for optimal $S_i\in\mathcal{S}^{Z_i\to R'}$ and ψ_i coincides with $\psi_x[x\to c_x]$ on $\bigcup_{j\leq i}\mathrm{YW}^{\Gamma}_{v_j}$.*

Proof Sketch. For "$\geq$", let $R^*\subseteq R'\cap\mathrm{Succ}_N^{R\uparrow}(\Gamma_{v_i})$ such that equality holds in (5). We consider a switching $S'\in\mathcal{S}^{Z_i\to R'}$ constructed from switchings $S_{i-1}\in\mathcal{S}^{Z_{i-1}\to R'\setminus R^*}$ and $S^*\in\mathcal{S}^{\Gamma_{v_i}\to R^*}$ as well as a mapping ψ' coinciding with $\psi_x[x\to c_x]$ on $\bigcup_{j<i}\mathrm{YW}^{\Gamma}_{v_j}$ constructed from mappings ψ_{i-1} and ψ^* such that (a) ψ_{i-1} coincides with $\psi_x[x\to c_x]$ on $\bigcup_{j<i}\mathrm{YW}^{\Gamma}_{v_j}$, (b) ψ^* coincides with $\psi_x[x\to c_x]$ on $\mathrm{YW}^{\Gamma}_{v_i}$, (c) the cost of ψ_{i-1} is optimal on $A_{Z_{i-1}}(S_{i-1})$ and (d) the cost of ψ^* is optimal on $A_{v_i}(S^*)$. By induction hypotheses, these costs are $Q^{\psi_x}_{x,c_x}[i-1,R'\setminus R^*]$ and $T^{\mathcal{SW}}[v_i,\psi_x[x\to c_x],R^*]$, respectively. Then, "$\geq$" follows by optimality of S_i and ϕ_i.

For "$\leq$", we let $R^*:=\mathrm{Succ}_{S_i}^{R\uparrow}(\Gamma_{v_i})$ and use independence of sub-solutions and the induction hypotheses to show that the cost of ϕ_i on $A_{Z_{i-1}}(S_i)$ is $Q^{\psi_x}_{x,c_x}[i-1,R'\setminus R^*]$ and the cost of ϕ_i on $A_{v_i}(S_i)$ is $T^{\mathcal{SW}}[v_i,\phi_i,R^*]$. Then, "$\leq$" follows from the fact that R^* is only one of the possible choices for the minimum in (5). ☐

For "$\geq$", let $c_x\in\phi(x)$ and $R^*\subseteq R^x\cap\mathrm{Succ}_N^{R\uparrow}(x)$ be such that equality holds in (4). We consider a switching $S'\in\mathcal{S}^{\Gamma_x\to R^x}$ constructed from switchings S_t and S^* with $S_t\in\mathcal{S}^{Z_t\to R^x\setminus R^*}$ (if $\mathrm{Pred}_N^{\downarrow}(x)\neq\varnothing$) or $S_t\in\mathcal{S}^{Z_t\to(R^x\setminus R^*)\cup\{x\}}$ (if $x\in R(N)$ and $\mathrm{Pred}_N^{\uparrow}(x)\neq\varnothing$), and $S^*\in\mathcal{S}^{\{x\}\to R^*}$, as well as a mapping ψ' coinciding with ψ_x on YW^{Γ}_x constructed from mappings ψ_t and ψ^* such that (a) ψ_t coincides with $\psi_x[x\to c_x]$ on $\bigcup_{i\leq t}\mathrm{YW}^{\Gamma}_{v_i}$, (b) ψ^* coincides with ψ_x on YW^{Γ}_x, (c) $\psi^*(x)=c_x$, (d) the cost of ψ_t is optimal on $A_{Z_t}(S_t)$ and (e) the cost of ψ^* is optimal on $A_{\{x\}}(S^*)$. Then, the cost of ψ^* on $A^{\uparrow}_{\{x\}}(S^*)$ is $\sum_{r\in R^*\cup\mathrm{Succ}_N^{T\uparrow}(x)}\delta\left(c_x,\psi_x(r)\right)$, the cost of ψ^* on $A^{\downarrow}_{\{x\}}(S^*)$ is $\min_{y\in\mathrm{Pred}_N^{\downarrow}(x)}\delta\left(c_x,\psi_x(y)\right)$ if the parent of x in S_t is above x in Γ (that is, $x\notin\mathrm{Succ}_{S_t}^{R\uparrow}(Z_t)$) and, by the claim above, the cost of ψ_t on $A_{Z_t}(S_t)$ is $Q^{\psi_x}_{x,c_x}[t,\mathrm{Succ}_{S_t}^{R\uparrow}(Z_t)]$. Then, as $S'\in\mathcal{S}^{\Gamma_x\to R^x}$, "$\geq$" follows by optimality of S and ϕ.

For "$\leq$", let $c_x:=\phi(x)$ and let $R^*:=\mathrm{Succ}_S^{R\uparrow}(\Gamma_x)$. We use independence of sub-solutions and the induction hypothesis to show that the cost of ϕ on $A_{Z_t}(S)$ is $Q^{\psi_x}_{x,c_x}[t,R'\setminus R^*]$ (if $x\notin R(N)$ or the parent of x in S is above x in Γ) or $Q^{\psi_x}_{x,c_x}[t,(R'\setminus R^*)\cup\{x\}]$ (if $x\in R(N)$ and the parent of x in S is in Γ_x). Further, the cost of ψ on $A^{\uparrow}_{\{x\}}(S)$ is $\sum_{r\in R^*\cup\mathrm{Succ}_N^{T\uparrow}(x)}\delta\left(c_x,\psi_x(r)\right)$, the cost of ψ on $A^{\downarrow}_{\{x\}}(S)$ is $\min_{y\in\mathrm{Pred}_N^{\downarrow}(x)}\delta\left(c_x,\psi_x(y)\right)$ if the parent of x in S is above x in Γ. Then, "$\leq$" follows from the fact that our choices of

c_x and R^* are only one of the possible choices for the minimum in (4). □

In order to solve the softwired parsimony problem given N, ϕ and Γ, all we have to do is compute $T^{\mathcal{SW}}[x, \psi_x, R^x]$ for each x bottom-up in Γ, each of the (at most) $|C|^{|\mathrm{YW}_x^\Gamma|}$ many choices of $\psi_x : \mathrm{YW}_x^\Gamma \to C$ with $\psi_x \trianglelefteq \phi$, and each $R^x \subseteq \mathrm{Succ}_N^{R\uparrow}(x) \subseteq \mathrm{YW}_x^\Gamma \cap R(N)$. To this end, $Q_{x,c_x}^{\psi_x}[i, R^x \setminus R^*]$ and $Q_{x,c_x}^{\psi_x}[i, (R^x \setminus R^*) \cup \{x\}]$ have to be computed for each child v_i of x in Γ and each $R^* \subseteq R^x \cap \mathrm{Succ}_N^{R\uparrow}(x)$. Then, by Lemma 8, the softwired parsimony score of N with respect to ϕ can be read from $T^{\mathcal{SW}}[\rho_\Gamma, \varnothing, \varnothing]$. In the following, let ψ_x be fix. Then, for fix c_x, we can compute $Q_{x,c_x}^{\psi_x}[i, R']$ for all choices of x, i and R' in $O(2^{|R' \cap \mathrm{Succ}_N^{R\uparrow}(v_i)|} + \sum_{x \in \Gamma} |\mathrm{Succ}_\Gamma(x)|) \subseteq O(2^{|\mathrm{YW}_x^\Gamma|+1} + |\Gamma|)$ time total. Further, the values of $\min_{y \in \mathrm{Pred}_N^\downarrow(x)} \delta(c_x, \phi_x(y))$ can be pre-computed for all $x \in \Gamma$ in $O(|A(N)|)$ time total. Then, to compute $T^{\mathcal{SW}}[x, \psi_x, R^x]$ for all x and R^x, we have to check $|V(N)|$ choices for x, as well as $|\phi(x)| \leq |C|$ choices for c_x and $3^{|\mathrm{Succ}_N^{R\uparrow}(x)|}$ choices for R^x and $R^* \subseteq R^x$ combined. Altogether, the table $T^{\mathcal{SW}}$ can be computed in $O(|C|^{|\mathrm{YW}_x^\Gamma|} \cdot (3^{|\mathrm{YW}_x^\Gamma|} \cdot |C| \cdot |V(N)| + |A(N)|))$ time. The computation of $Q_{x,c_x}^{\psi_x}$ in $O(2^{|\mathrm{YW}_x^\Gamma|} + |A(N)|)$ time is absorbed by this. For practical purposes, note that estimating $|\mathrm{Succ}_N^{R\uparrow}(x)| \leq |\mathrm{YW}_x^\Gamma|$ is quite crude and equality will almost never be attained. Then, the following result holds:

Theorem 2 *Given a network N, $\phi : V(N) \to 2^C$ and a tree Γ agreeing with N, the softwired parsimony score of (N, ϕ) can be computed in $O(|C|^{\mathrm{yw}(\Gamma)} \cdot (3^{\mathrm{yw}(\Gamma)} \cdot |C| \cdot |V(N)| + |A(N)|))$ time.*

Again, we can replace $\mathrm{yw}(\Gamma)$ by $\mathrm{tw}(N)$ using Proposition 1.

Corollary 2 *Let (N, ϕ) be an instance of Softwired Parsimony. Let $t \geq \mathrm{tw}(N)$ and let T be the time in which a width-t tree decomposition of N can be computed. Then, the softwired parsimony score of (N, ϕ) can be computed in $O(T + |C|^t \cdot (3^t \cdot |C| \cdot |V(N)| + |A(N)|))$ time.*

Parental parsimony

For ease of presentation, we introduce some additional notation. First, for any a and b, we abbreviate $\max\{a-b, 0\} =: a \dot{-} b$. Let ψ and ψ' be functions. If ψ maps all items to $\varnothing$ or to 0, then we say that ψ is a *zero-function* and we write $\psi = \vec{0}$. We use $\psi - \psi'$ to denote the function defined on the domain of ψ for which $(\psi - \psi')(x) = \psi(x)$ if $\psi'(x) = \perp$ and $(\psi - \psi')(x) = \psi(x) - \psi'(x)$, otherwise. This definition extends to functions mapping to sets in a natural way.

Each finite-cost lineage function f corresponds to a phylogenetic tree "embedded" in N whose branches are called *lineages* (see Fig. 1(right)). For each $x \in V(N)$, $f(x)$ represents the set of such lineages passing through x. Each such lineage may "choose" a parent among the parents of x in N. This models the biological circumstance that a character trait may be inherited from any parent. We compute (the cost of) an optimal lineage function on N using a tree Γ that agrees with N. To compute $\mathrm{cost}_f(x)$, we require knowledge of $\sum_{y \in \mathrm{Pred}(x)} |f(y)|$ as well as $\bigcup_{y \in \mathrm{Pred}(x)} f(y)$ (see Definition 4). We partition the predecessors of x over which the formula iterates into those above x in Γ and those below (since Γ agrees with N, all predecessors of x in N are comparable to y in Γ). For all $y \in \mathrm{YW}_x^\Gamma$, we thus store

1. the set $\lambda(y) := f(y)$ of lineages in y,
2. the subset $\psi(y)$ of lineages of y that also occur in parents (in N) of y that are below x in Γ, that is, in $\mathrm{Pred}_N^{\uparrow x}(y)$ (such lineages are inherited by y at no cost),
3. the total number $\eta(y)$ of lineages of y that can be inherited from parents (in N) of y that are below x in Γ, that is, from $\mathrm{Pred}_N^{\uparrow x}(y)$ (cost 0 or 1).

Then, in order to compute an entry $T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x]$, we "guess" the set $U \subseteq \phi(x)$ of lineages passing through x in an optimal solution, as well as the set $D \subseteq U$ of lineages inherited from nodes in $\mathrm{Pred}_N^\uparrow(x)$. This allows us to infer $\eta(x) = |\lambda(x)| \dot{-} \sum_{r \in \mathrm{Pred}_N^\downarrow(x)} |\lambda(r)|$ and $\psi(x) := D$. Then, by Definition 4, the cost incurred by f on x can be computed from $\sum_{y \in \mathrm{Pred}_N(x)} |f(y)| = \eta(x) + \sum_{y \in \mathrm{Pred}_N^\downarrow(x)} |\lambda(y)|$ and $\bigcup_{y \in \mathrm{Pred}_N(x)} f(y) = \psi(x) \cup \bigcup_{y \in \mathrm{Pred}_N^\downarrow(x)} \lambda(y)$.

We will compute table entries for x using the already computed table entries for the children v_i of x in Γ. In these lookups, we have $x \in \mathrm{YW}_{v_i}^\Gamma$ so, to be consistent with the semantics, we have to make sure that $\lambda(x) = U$, $\psi(x) = D$, and that all lineages of x that are not inherited from $\mathrm{Pred}_N^\downarrow(x)$ can be inherited from $\mathrm{Pred}_N^\uparrow(x)$, that is, $\eta(x) = |\lambda(x)| \dot{-} \sum_{r \in \mathrm{Pred}_N^\downarrow(x)} |\lambda(r)|$. Further, each child y of x in N may inherit a lineage from x and, if y is above x in Γ, this has to be registered by removing the lineages of U from $\psi(y)$ and subtracting $|U|$ from $\eta(y)$. Finally, the lineages represented by ψ and η are distributed among the children of x in Γ using the table Q. In the following, in order to avoid treating the case that $x = \rho_N$ separately, we define $\rho(x) := 1 - \delta(x, \rho_N)$, that is, $\rho(x) = 1$ if and only if $x = \rho_N$.

Definition 7 Let Γ be a tree that agrees with N, let $x \in V(N)$, let $\lambda_x : \mathrm{YW}_x^\Gamma \to 2^C$ with $\lambda_x \trianglelefteq \phi$ and let $\psi_x \trianglelefteq \lambda_x$. Let $\{v_1, v_2, \ldots, v_t\} = \mathrm{Succ}_\Gamma(x)$ ($t = 0$ if x is a leaf in Γ). Then, set $T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x]$ to

$$\min_{\substack{D \subseteq U \subseteq \phi(x) \\ U \neq \varnothing}} Q_x^{\lambda_x x \to U}\left[t, \psi_x\begin{bmatrix} x \to D \\ \forall_{y \in \mathrm{Succ}_N^{\uparrow}(x)}\ y \to \psi_x(y) \setminus U \end{bmatrix}, \eta_x\begin{bmatrix} x \to |U| \dot{-} \sum_{u \in \mathrm{Pred}_N^{\downarrow}(x)} |\lambda_x(u)| \\ \forall_{y \in \mathrm{Succ}_N^{\uparrow}(x)} y \to \eta_x(y) \dot{-} |U| \end{bmatrix}\right] + \left| U \setminus \left(D \cup \bigcup_{u \in \mathrm{Pred}_N^{\downarrow}(x)} \lambda_x(u) \right) \right| \tag{7}$$

where $Q_x^\lambda[i, \psi, \eta]$ equals

$$\begin{cases} \min\limits_{\psi' \trianglelefteq \psi|_{\mathrm{YW}_{v_i}^\Gamma}} \min\limits_{\eta' \trianglelefteq \eta|_{\mathrm{YW}_{v_i}^\Gamma}} Q_x^\lambda[i-1, \psi - \psi', \eta - \eta'] + T^{\mathcal{PT}}[v_i, \lambda|_{\mathrm{YW}_{v_i}^\Gamma}, \psi', \eta'] & \text{if } i > 0 \\ -\rho(x) & \text{if } i = 0 \text{ and } \psi = \vec{0} \text{ and } \eta = \vec{0}[x \to \rho(x)] \\ \infty & \text{otherwise} \end{cases} \tag{8}$$

Note how the table Q_x^λ distributes the lineage branches of x whose parents are in Γ_x among the children of x in Γ. We show that both $T^{\mathcal{PT}}$ and Q_x^λ are monotone in ψ and η (wrt. $\trianglelefteq$).

Lemma 9 *Let $x \in V(N)$, let $i \in \mathbb{N}$, let $\lambda : \mathrm{YW}_x^\Gamma \to 2^C$, let $\eta, \eta' : \mathrm{YW}_x^\Gamma \to \mathbb{N}$, and let $\psi, \psi' : \mathrm{YW}_x^\Gamma \to 2^C$ such that $\psi' \trianglelefteq \psi \trianglelefteq \lambda$ and $\vec{0}[x \to \rho(x)] \trianglelefteq \eta' \trianglelefteq \eta$. Then,*

$$T^{\mathcal{PT}}[x, \lambda, \psi', \eta'] \leq T^{\mathcal{PT}}[x, \lambda, \psi, \eta] \quad \text{and} \quad Q_x^\lambda[i, \psi', \eta'] \leq Q_x^\lambda[i, \psi, \eta]$$

Proof Sketch. The lemma can be proved by induction on the height of x in Γ and the value of i. If x is a leaf, then $Q_x^\lambda[0, \psi, \eta]$ is finite only if $\psi = \vec{0}$ and $\eta = \vec{0}[x \to \rho(x)]$, implying the second inequality. For monotony of $T^{\mathcal{PT}}$, fix the sets $D \subseteq U \subseteq \phi(x)$ for which the minimum in the formula of $T^{\mathcal{PT}}[x, \lambda, \psi, \eta]$ is attained. Then, by monotony of Q_x^λ, replacing ψ by ψ' and η by η' in this formula does not increase its value and this value is at most $T^{\mathcal{PT}}[x, \lambda, \psi', \eta']$ since it is obtained for one of several possible choices for D and U. If x is not a leaf in Γ then monotonicity of $Q_x^\lambda[i, \ldots]$ is implied by monotonicity of $Q_x^\lambda[i-1, \ldots]$ and monotonicity of $T^{\mathcal{PT}}[v, \ldots]$ for the children v of x. Finally, monotonicity of $T^{\mathcal{PT}}$ follows from monotonicity of Q_x^λ as in the induction base. □

Lemma 10 *Let Γ be a tree agreeing with N, let $x \in V(N)$, let $\psi_x, \lambda_x : \mathrm{YW}_x^\Gamma \to 2^c$ and $\eta_x : \mathrm{YW}_x^\Gamma \to \mathbb{N}$. Let f minimize* $\mathrm{cost}_{(f)}$ *among all lineage functions in $\mathcal{LF}_{N,\phi}$ such that, for all $w \in \mathrm{YW}_x^\Gamma$, $\lambda_x(w) = f(w)$, $\psi_x(w) = f(w) \cap \bigcup_{u \in \mathrm{Pred}_N^{\uparrow x}(w)} f(u)$, and $\eta_x(w) \leq \sum_{u \in \mathrm{Pred}_{N\uparrow x(w)}} |f(u)|$. If there are no such f, then $T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x = \infty]$. Otherwise,*

$$T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x] = \sum_{z \leq_\Gamma x} \mathrm{cost}_f(z)$$

Proof Sketch. Let us abbreviate $Z_i := \bigcup_{j \leq i} V(\Gamma_{v_j})$. We first show that the table Q does what we expect it to do.

Claim 3 *Let $\lambda, \psi : \mathrm{YW}_x^\Gamma \cup \{x\} \to 2^C$ and $\eta : \mathrm{YW}_x^\Gamma \cup \{x\} \to \mathbb{N}$ such that $\psi \trianglelefteq \lambda \trianglelefteq \phi$. Let $f_i \in \mathcal{LF}_{N,\phi}$ have minimum cost on $\bigcup_{j \leq i} \Gamma_{v_j}$ among all lineage functions for N that, for all $w \in \bigcup_{j \leq i} \mathrm{YW}_{v_j}^\Gamma$, satisfy (a) $\lambda(w) = f_i(w)$, (b) $\psi(w) = f_i(w) \cap \bigcup_{j \leq i} \bigcup_{u \in \mathrm{Pred}_N^{\uparrow v_j}(w)} f_i(u)$, and (c) $\eta(w) \leq \sum_{j \leq i} \sum_{u \in \mathrm{Pred}_N^{\uparrow v_j}(w)} |f_i(u)|$ Then, $Q_x^\lambda[i, \psi, \eta] = \sum_{j \leq i} \sum_{u \in \Gamma_{v_j}} \mathrm{cost}_{f_i}(u)$.*

Proof Sketch. For "≥", let $\psi' \trianglelefteq \psi|_{\mathrm{YW}_{v_i}^\Gamma}$ and $\eta' \trianglelefteq \eta|_{\mathrm{YW}_{v_i}^\Gamma}$ such that equality holds in (8). Let $f_{i-1} \in \mathcal{LF}_{N,\phi}$ minimize $\sum_{j<i} \sum_{u \in \Gamma_{v_j}} \mathrm{cost}_{f_{i-1}}(u)$ among all lineage functions satisfying (a)–(c) for $i-1$. Let $f^* \in \mathcal{LF}_{N,\phi}$ minimize $\sum_{u \in \Gamma_{v_i}} \mathrm{cost}_{f^*}(u)$ among all lineage functions that, for all $w \in \mathrm{YW}_{v_i}^\Gamma$, satisfy $\lambda(w) = f^*(w)$, $\psi'(w) = f^*(w) \cap \bigcup_{u \in \mathrm{Pred}_N^{\uparrow v_i}(w)} f^*(u)$ and $\eta'(w) = \sum_{u \in \Gamma_{v_i}} |f^*(u)|$. By induction hypotheses, the cost of f_{i-1} on Z_i is $Q_x^\lambda[i-1, \psi - \psi', \eta - \eta']$

and the cost of f^* on Γ_{v_i} is $T^{\mathcal{PT}}[v_i, \lambda\mid_{\mathrm{YW}^\Gamma_{v_i}}, \psi', \eta']$. From f_{i-1} and f^*, we construct a lineage function $f' \in \mathcal{LF}_{N,\phi}$ whose cost on Z_i is $\sum_{j<i}\sum_{u\in\Gamma_{v_j}} \mathrm{cost}_{f_{i-1}}(u) + \sum_{u\in\Gamma_{v_i}} \mathrm{cost}_{f^*}(u)$. Then, "$\geq$" follows by optimality of f_i on Z_i.

For "$\leq$", let ψ' and η' be such that, for all $w \in \mathrm{YW}^\Gamma_{v_i}$, we have $\psi'(w) = f_i(w) \cap \bigcup_{u\in\mathrm{Pred}^{\uparrow v_i}_N(w)} f_i(u) \subseteq \psi(w)$ and $\eta'(w) = \sum_{u\in\mathrm{Pred}^{\uparrow v_i}_N(w)} |f_i(u)|$. By independence of sub-solutions, f_i is optimal on Z_{i-1} and on Γ_{v_i} so, by induction hypotheses, the cost of f_i on Z_{i-1} is $Q^\lambda_x[i-1, \psi-\psi', \eta-\eta']$ and the cost of f_i on Γ_{v_i} is $T^{\mathcal{PT}}[v_i, \lambda\mid_{\mathrm{YW}^\Gamma_{v_i}}, \phi', \eta']$. Since ψ' and η' are only one of the possible choices for the minimum in (8), "$\leq$" follows. □

For "$\geq$", let $D \subseteq U \subseteq \phi(x)$ such that equality holds in (7). We construct a lineage function f' that assigns $f'(x) = U$ and such that the lineages of D are inherited from parents of x (in N) that are below x in Γ. To this end, we ask the dynamic programming table for the cost of a lineage function that is optimal on Z_t and such that 1. $\psi'(x) = D$ *(lineages in D are inherited from parents of x in* Γ_x*)* 2. $\psi'(w) = \psi'(w) \setminus U$ for all $w \in \mathrm{Succ}^\uparrow_N(x)$ *(children of x in* YW^Γ_x *no longer need to inherit the lineages in U from* Γ_x*)* 3. $\eta'(x) = |U| \dot{-} \sum_{u\in\mathrm{Pred}^\downarrow_N(x)} |\lambda_x(u)|$ *(x needs to inherit |U| lineages in total:* $|\lambda_x(u)|$ *come from every parent u of x in* YW^Γ_x *while the rest has to be inherited from* Γ_x*)* and 4. $\eta'(w) = \eta_x(w) \dot{-} |U|$ for all $w \in \mathrm{Succ}^\uparrow_N(x)$ *(children of x in* YW^Γ_x *can inherit a maximum of |U| lineages from x)*. Since the functions $\lambda' := \lambda_x[x \to U]$, $\psi' := \psi_x\left[x \to D, \forall_{u\in\mathrm{Succ}^\uparrow_N(x)} w \to \psi_x(w)\setminus U\right]$ and $\eta' := \eta_x\left[x \to |U| \dot{-} \sum_{u\in\mathrm{Pred}^\downarrow_N(x)} |\lambda_x(u)|, \forall_{u\in\mathrm{Succ}^\uparrow_N(x)} w \to \eta_x(w) \dot{-} |U|\right]$ satisfy the conditions of Claim 3, the optimal cost of such a lineage function f' on Z_t is $Q^\lambda_x[t, \psi', \eta']$. Further, the cost of f' on x is the number of lineages in U that is not inherited "for free" from parents of x, that is, $|U \setminus (D \cup \bigcup_{u\in\mathrm{Pred}^\downarrow_N(x)} \lambda_x(u))|$. Then, "$\geq$" follows by optimality of f on Γ_x.

For "$\leq$", let $U := f(x)$ and let $D := U \cap \bigcup_{u\in\mathrm{Pred}^\uparrow_N(x)} f(x)$ be the set of lineages of U that are inherited from parents of x in N that are below x in Γ. By independence of sub-solutions, f is optimal on Z_t so, by Claim 3, its cost on Z_t is $Q^\lambda_x[t, \psi', \eta']$ where $\psi' := \psi_x[\ldots]$ and $\eta' := \eta_x[\ldots]$ are defined as in (7) and its cost on x is $|f(x) \setminus (\bigcup_{u\in\mathrm{Pred}^\uparrow_N(x)} f(x) \cup \bigcup_{u\in\mathrm{Pred}^\downarrow_N(x)} f(x))| = |U \setminus (D \cup \bigcup_{\mathrm{Pred}^\downarrow_N(x)} f(x))|$. Then, "$\leq$" follows from the fact that U and D are only one of the possible choices for the minimum in (7). □

To solve the parental parsimony problem given N, ϕ and Γ, we compute $T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x]$ for each x bottom-up in Γ, each $\psi_x, \lambda_x : \mathrm{YW}^\Gamma_x \to 2^C$ with $\psi_x \trianglelefteq \lambda_x \trianglelefteq \phi$ and each $\eta_x : \mathrm{YW}^\Gamma_x \to \{0, \ldots, |C|\}$ (by Definition 7, no value larger than $|C|$ ever enters η_x and all modifications to η_x decrease the mapped-to values). To this end, $Q^\lambda_x[i, \psi, \eta]$ is computed for each x, i, λ, ψ, and η by making at most $2^{|C|\cdot|\mathrm{YW}^\Gamma_x|} \cdot |C|^{|\mathrm{YW}^\Gamma_x|}$ queries to $Q^{\psi_x}_{x,c_x}$ and $T^{\mathcal{PT}}$. As there are $O(|A(N)|)$ valid combinations of x and i, the table Q can be computed in $O(|A(N)| \cdot 3^{|C|\cdot \mathrm{yw}(N)} \cdot |C|^{\mathrm{yw}(N)} \cdot 2^{|C|\cdot \mathrm{ywN}} \cdot |C|^{\mathrm{yw}(N)}) = O(|A(N)| \cdot 6^{|C|\cdot \mathrm{yw}(N)|} \cdot 4^{\mathrm{yw}(N)\cdot\log|C|})$ time. Further, computing each $T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x]$ requires testing $3^{|\phi(x)|} \leq 3^{|C|}$ choices for $D \subseteq U \subseteq \phi(x)$ and computing $|U \setminus (D \cup \bigcup_{u\in\mathrm{Pred}^\downarrow_N(x)} \lambda_x(u))|$ in $O(|C|)$ time (we precompute $\bigcup_{u\in\mathrm{Pred}^\downarrow_N(x)} \lambda_x(u)$ for each fix x and λ_x). Thus, the table $T^{\mathcal{PT}}$ can be computed in $O(3^{|C|\cdot \mathrm{yw}(N)} \cdot (|C|^{\mathrm{yw}(N)+1} \cdot 3^{|C|} + |A(N)|))$ time, which is dominated by the construction of Q.

Theorem 3 *Given a network N,* $\phi : V(N) \to 2^C$ *and a tree* Γ *agreeing with N, the parental parsimony score of* (N, ϕ) *can be computed in* $O(6^{\mathrm{yw}(\Gamma)\cdot|C|} \cdot 4^{\mathrm{yw}(\Gamma)\cdot\log|C|} \cdot |A(N)|)$ *time.*

Again, we can replace $\mathrm{yw}(\Gamma)$ by $\mathrm{tw}(N)$ using Proposition 1.

Corollary 3 *Let* (N, ϕ) *be an instance of* Parental Parsimony*. Let* $t \geq \mathrm{tw}(N)$ *and let T be the time in which a width-t tree decomposition of N can be computed. Then, the parental parsimony score of* (N, ϕ) *can be computed in* $O(T + 6^{t\cdot|C|} \cdot 4^{t\cdot\log|C|} \cdot |A(N)|)$ *time.*

Note that the parental parsimony setting supports assigning multiple states of a character to a single species, thereby modeling species carrying multiple alleles of a single gene. By forcing $D \subseteq U = \phi(x)$ instead of $D \subseteq U \subseteq \phi(x)$ if x is a leaf, we can trivially modify our dynamic programming to explain multiple character states in extant species.

Corollaries 1, 2 and 3 give the running times of our algorithms as depending on the treewidth of N. The state-of-the-art solutions for Hardwired Parsimony, Softwired Parsimony and Parental Parsimony have the following respective running times: $O(|C|^{r+2}|V(N)|)$ [9], $O(2^\ell|C|^2|V(N)||A(N)|)$ [8] and $O(|2^C|^{\ell+3}|V(N)|)$ [12]. Since the scanwidth of N is potentially much smaller than its level ℓ [28], and the treewidth of N is smaller than its scanwidth [20], we have $\mathrm{tw}(N) - 1 \leq \ell \leq r$. Thus, we expect that there will be several cases where our algorithms will be faster than the current best-known ones.

Discussion

In this paper, we focused on the small version of the parsimony problem for networks given a specific position in the genome. When markers can be assumed to be independent, as it is the case when a certain distance is preserved between genomic locations included in the matrix, each position can be analyzed separately, and the parsimony score of a network w.r.t. the matrix is simply the sum of the parsimony scores of the network for each genomic location. Thus, the algorithms presented here can be easily expanded to several independent genomic locations. Moreover, our formulations are defined for networks that are not necessarily binary, can account for polymorphism and can impose restrictions on ancestral states. As discussed above, our algorithms can be orders of magnitude faster than the state-of-the-art solutions. A comparison of the reticulation number, the level, the scanwidth and the treewidth for practically relevant classes of networks would thus be an interesting project for future work.

Our results are slightly overshadowed by the fact that optimal tree decompositions are very hard to compute. However, practical exact and approximative algorithms are available today and we expect them do perform well, as phylogenetic networks can be expected to only be moderately tangled.

Furthermore, closer inspection of our dynamic programming formulations (most prominently Definition 6) unveils that their computation is faster when the maximum number of reticulations in each bag is small. Thus, it would be interesting to be able to compute tree decompositions in which this quantity is low, to the point where one could improve running time of the algorithm by sacrificing optimality of the decomposition in favor of reducing this "reticulation density". Research in this direction is, to the best of our knowledge, limited to a paper by Bachoore and Bodlaender [29], considering tree decompositions minimizing a *weight function* over the bags.

The ability to fast-score phylogenetic networks under the parsimony framework could be a big help in designing likelihood-based heuristics or bayesian methods to infer networks from independent markers [28, 30] by providing fast heuristics to compute the initial networks with which to start the likelihood or bayesian search, or to design fast local-search techniques.

In the future, we would like to tackle the SMALL PARSIMONY problem for several *dependent* genomic locations (e.g. a gene). Little is known for this problem, except that it stays NP-hard even for binary characters on level-1 networks [31] and that it is fixed-parameter tractable in the number of reticulations of the network [6]. Another important direction would be to study the BIG PARSIMONY problem, which is currently wide open, even lacking a consensus of the definition of optimality [6, 32–34].

Appendix

Lemma 7 *Let $x \in V(N)$ and let $\psi_x : \mathrm{YW}_x^\Gamma \to C$ with $\psi_x \trianglelefteq \phi$. Let $\psi : V(N) \to C$ with $\psi_x \trianglelefteq \psi \trianglelefteq \phi$ such that ψ minimizes $\sum_{uw \in A_x(N)} \delta_{\psi}(u, w)$. Then,*

$$T^{\mathcal{HW}}[x, \psi_x] = \sum_{uw \in A_x(N)} \delta_{\psi}(u, w)$$

Proof The proof is by induction on the height of x in Γ. For the induction base, suppose that x is a leaf in Γ and note that $A_x(N) = A_{\{x\}}(N)$ in this case. Then, (3) simplifies to

$$\begin{aligned} T^{\mathcal{HW}}[x, \psi_x] &= \min_{c_x \in \phi(x)} \sum_{z \in \mathrm{Pred}_N^{\downarrow}(x) \cup \mathrm{Succ}_N^{\uparrow}(x)} \delta(c_x, \psi_x(z)) \\ &= \min_{c_x \in \phi(x)} \left(\sum_{zx \in A_x^{\downarrow}(N)} \delta(c_x, \psi_x(z)) + \sum_{xz \in A_x^{\uparrow}(N)} \delta(c_x, \psi_x(z)) \right) \\ &= \min_{c_x \in \phi(x)} \sum_{uw \in A_x(N)} \delta_{\psi_x[x \to c_x]}(u, w) \end{aligned} \tag{9}$$

Since $\psi(x) \in \phi(x)$, we know that $\psi(x)$ participates in the minimum in (9), implying the "$\leq$"-direction. For the "$\geq$"-direction, assume that $T^{\mathcal{HW}}[x, \psi_x] < \sum_{uw \in A_x(N)} \delta_{\psi}(u, w)$. By (9), there is some $c_x \neq \psi(x)$ with $c_x \in \phi(x)$ and $\sum_{uw \in A_x(N)} \delta_{\psi_x[x \to c_x]}(u, w) < \sum_{uw \in A_x(N)} \delta_{\psi}(u, w)$. Since $c_x \in \phi(x)$, we still have $\psi_x \trianglelefteq \psi_x[x \to c_x] \trianglelefteq \phi$, contradicting optimality of ψ on $A_x(N)$. For the induction step, suppose that $t > 0$ and consider both directions separately.

"$\le$": Let $i \le t$, and let $\psi_i := \psi \mid_{\mathrm{YW}^{\Gamma}_{v_i}} = \psi_x[x \to \psi(x)] \mid_{\mathrm{YW}^{\Gamma}_{v_i}}$. Then, by Lemma 6 (with $Z = \{x\}$, $Y = \{v_i\}$, $\mathcal{G} = \{N\}$ and $(S^y)_{y \in Y} = (N)_{y \in Y}$), optimality of ψ on $A_x(N)$ implies optimality of ψ_i on $A_{v_i}(N)$. Thus, we can use the induction hypothesis on $T^{\mathcal{HW}}[v_i, \psi_i]$. Since $\psi(x)$ participates in the minimum of (3),

$$\begin{aligned} T^{\mathcal{HW}}[x, \psi_x] &\overset{(3)}{\le} \sum_{1 \le i \le t} T^{\mathcal{HW}}[v_i, \psi_i] + \sum_{z \in \mathrm{Pred}^{\downarrow}_N(x) \cup \mathrm{Succ}^{\uparrow}_N(x)} \delta(\psi(x), \psi_x(z)) \\ &\overset{\mathrm{IH}}{=} \sum_{1 \le i \le t} \sum_{uw \in A_{v_i}(N)} \delta_{\psi}(u, w) + \sum_{z \in \mathrm{Pred}^{\downarrow}_N(x) \cup \mathrm{Succ}^{\uparrow}_N(x)} \delta_{\psi}(x, z) \\ &= \sum_{uw \in A_x(N)} \delta_{\psi}(u, w) \end{aligned}$$

"$\ge$": Assume towards a contradiction that the lemma is false, that is, "<" holds. By (3), there is some $c_x \in \phi(x)$ such that

$$T^{\mathcal{HW}}[x, \psi_x] = \sum_{1 \le i \le t} T^{\mathcal{HW}}[v_i, \psi_x[x \to c_x] \mid_{\mathrm{YW}^{\Gamma}_{v_i}}] + \sum_{z \in \mathrm{Pred}^{\downarrow}_N(x) \cup \mathrm{Succ}^{\uparrow}_N(x)} \delta(c_x, \psi_x(z)) \tag{10}$$

Since $c_x \in \phi(x)$, we can extend $\psi_x[x \to c_x]$ to $V(N)$ without violating ϕ, that is, there are functions $\psi' : V(N) \to C$ with $\psi_x[x \to c_x] \trianglelefteq \psi' \trianglelefteq \phi$. Among them, let ψ' minimize $\sum_{i \le t} \sum_{uw \in A_{v_i}(N)} \delta_{\psi'}(u, w)$. By Lemma 6 (with $Z = \mathrm{Succ}_{\Gamma}(x)$, $Y = \{v_i\}$, $\mathcal{G} = \{N\}$, and $(S^y)_{y \in Y} = (N)_{y \in Y}$), ψ' also minimizes $\sum_{uw \in A_{v_i}(N)} \delta_{\psi'}(u, w)$ for all $1 \le i \le t$. Thus, the induction hypothesis applies to $T^{\mathcal{HW}}[v_i, \psi_x[x \to c_x] \mid_{\mathrm{YW}^{\Gamma}_{v_i}}]$ for all i. Then,

$$\begin{aligned} T^{\mathcal{HW}}[x, \psi_x] &\overset{(10)}{=} \sum_{1 \le i \le t} T^{\mathcal{HW}}[v_i, \psi_x[x \to c_x] \mid_{\mathrm{YW}^{\Gamma}_{v_i}}] + \sum_{z \in \mathrm{Pred}^{\downarrow}_N(x) \cup \mathrm{Succ}^{\uparrow}_N(x)} \delta(c_x, \psi_x(z)) \\ &\overset{\mathrm{IH}}{=} \sum_{1 \le i \le t} \sum_{uw \in A_{v_i}(N)} \delta_{\psi'}(u, w) + \sum_{z \in \mathrm{Pred}^{\downarrow}_N(x) \cup \mathrm{Succ}^{\uparrow}_N(x)} \delta(c_x, \psi_x(z)) \\ &\overset{\psi_x = \psi' \mid_{\mathrm{YW}^{\Gamma}_x}}{=} \sum_{1 \le i \le t} \sum_{uw \in A_{v_i}(N)} \delta_{\psi'}(u, w) + \sum_{z \in \mathrm{Pred}^{\downarrow}_N(x) \cup \mathrm{Succ}^{\uparrow}_N(x)} \delta_{\psi'}(x, z) \\ &= \sum_{uw \in A_x(N)} \delta_{\psi'}(u, w) \end{aligned}$$

Since, by assumption, $T^{\mathcal{HW}}[x, \psi_x]$ is strictly less than the cost of ψ on $A_x(N)$, we conclude that the cost of ψ' on $A_x(N)$ is strictly less than that of ψ, contradicting optimality of ψ. $\square$

Lemma 8 *Let Γ be a tree that agrees with N, let $x \in V(N)$, let $\psi_x : \mathrm{YW}^{\Gamma}_x \to C$ with $\psi_x \trianglelefteq \phi$, and let $R^x \subseteq \mathrm{Succ}^{R\uparrow}_N(\Gamma_x)$. If $\mathcal{S}^{\Gamma_x \to R^x}(N) = \varnothing$, then $T^{\mathcal{SW}}[x, \psi_x, R^x = \infty]$. Otherwise, let $S \in \mathcal{S}^{\Gamma_x \to R^x}(N)$ and $\psi : V(N) \to C$ such that (a) $\psi_x \trianglelefteq \psi \trianglelefteq \phi$ and (b) $\sum_{uw \in A_x(S)} \delta_{\psi}(u, w)$ is minimum among all such S and ψ. Then,*

$$T^{\mathcal{SW}}[x, \psi_x, R^x] = \sum_{uw \in A_x(S)} \delta_{\psi}(u, w). \tag{6}$$

Proof Note that arcs that are incoming to tree nodes cannot be switched off and, thus, $\mathrm{Succ}^{T\uparrow}_N(z) = \mathrm{Succ}^{T\uparrow}_{S'}(z)$ for all $z \in V(N)$ and all switchings $S' \in \mathcal{S}(N)$. The proof is by induction on the height of x in Γ.

Case 1: x is a leaf in Γ, that is, $t = 0$. First, note that $R^x \subseteq \mathrm{Succ}^{R\uparrow}_N(x)$ and no $r \in R^x \subseteq R(N)$ can have all their parents in $\Gamma_x = \{x\}$, thus implying $\mathcal{S}^{x \to R^x}(N) \ne \varnothing$. Next, let y be the predecessor of x in S and note that $y \in \mathrm{Pred}^{\downarrow}_N(x) = \mathrm{Pred}_N(x)$. Further, y minimizes $\delta_{\psi}(y, x)$ among all $y \in \mathrm{Pred}_N(x)$ as, otherwise, we can construct a new switching $S' \in \mathcal{S}^{\Gamma_x \to R^x}(N)$ by replacing yx by some $y'x$ with $y' \in \mathrm{Pred}_N(x)$, thereby contradicting (b). Clearly, $\mathrm{Pred}_N^{\uparrow}(x) = \varnothing$ and $Q_{x,c_x}^{\psi_x}[0, R^x \setminus R^*] \neq \infty$ only if $R^* = R^x$. Thus,

$$T^{\mathcal{SW}}[x,\psi_x,R^x] \overset{(4)}{=} \min_{c_x \in \phi(x)} \left(\sum_{r \in R^x \cup \mathrm{Succ}_N^{T\uparrow}(x)} \delta(c_x, \psi_x(r)) + \min_{y \in \mathrm{Pred}_N^{\downarrow}(x)} \delta(c_x, \psi_x(y)) \right)$$
$$\overset{\psi(x)\in\phi(x)}{\leq} \sum_{r \in R^x \cup \mathrm{Succ}_N^{T\uparrow}(x)} \delta(\psi(x), \psi_x(r)) + \min_{yx \in A_x^{\downarrow}(N)} \delta(\psi(x), \psi_x(y))$$
$$= \sum_{xr \in A_x^{\uparrow}(S)} \delta_{\psi}(x,r) + \sum_{yx \in A_x^{\downarrow}(S)} \delta_{\psi}(y,x) = \sum_{uw \in A_x(S)} \delta_{\psi}(u,w)$$

and there is some $c_x \in \phi(x)$ such that equality holds if $\psi(x) = c_x$. Let $\psi^* := \psi[x \to c_x]$ be the result of changing the assignment of x to c_x in ψ and note that $\psi_x \trianglelefteq \psi^*$. Clearly, we still have $S \in \mathcal{S}^{\Gamma_x \to R^x}(N)$. Thus,

$$T^{\mathcal{SW}}[x,\psi_x,R^x] \overset{(4)}{=} \sum_{r \in R^x \cup \mathrm{Succ}_N^{T\uparrow}(x)} \delta(c_x, \psi_x(r)) + \min_{yx \in A_x^{\downarrow}(N)} \delta(c_x, \psi_x(y))$$
$$\overset{\psi_x \trianglelefteq \psi^*}{=} \sum_{xr \in A_x^{\uparrow}(S)} \delta_{\psi^*}(x,r) + \sum_{yx \in A_x^{\downarrow}(S)} \delta_{\psi^*}(y,x)$$
$$= \sum_{uw \in A_x(S)} \delta_{\psi^*}(u,w) \overset{\text{Lemma 8(b)}}{\geq} \sum_{uw \in A_x(S)} \delta_{\psi}(u,w)$$

Case 2: x has children $v_1, v_2, ..., v_t$ in Γ. Recall that we suppose that $x \in \bigcup_{i \leq t} \mathrm{YW}_{v_i}^{\Gamma}$ by Lemma 3. For all $S^* \in \mathcal{S}(N)$ and all anti-chains Y in Γ, abbreviate $\mathcal{S}^{Y \to \bigcup_{y \in Y} \mathrm{Succ}_{S^*}^{R\uparrow}(\Gamma_y)}(N) =: \mathcal{S}^{Y,S^*}(N)$, that is, roughly, the set of switchings of N with the same "behavior" as S^* on Y. The proof of Case 2 relies on the independence of partial solutions established by Lemma 6 with $\mathcal{G} = \mathcal{S}^{Y,S^*}(N)$. To apply Lemma 6, we show that any set of switchings S^y such that $\{\mathrm{Succ}_{S^y}^{R\uparrow}(\Gamma_y) \mid y \in Y\}$ is a partition of $\bigcup_{y \in Y} \mathrm{Succ}_{S^*}^{R\uparrow}(\Gamma_y)$ is a Y-substitution system for $\mathcal{S}^{Y,S^*}(N)$.

Claim 4 *Let $S^* \in \mathcal{S}(N)$ and let Y be an anti-chain in Γ. For each $y \in Y$, let $S^y \in \mathcal{S}(N)$ such that $\{\mathrm{Succ}_{S^y}^{R\uparrow}(\Gamma_y) \mid y \in Y\}$ is a partition of $\bigcup_{y \in Y} \mathrm{Succ}_{S^*}^{R\uparrow}(\Gamma_y)$. Let*

$$S' := \left(V(N), \left(A(S^*) \setminus \bigcup_{y \in Y} A_y(S^*) \right) \cup \bigcup_{y \in Y} A_y(S^y) \right)$$

Then, $S' \in \mathcal{S}^{Y,S^*}(N)$.

Proof Since $\{\mathrm{Succ}_{S_i}^{R\uparrow}(\Gamma_y) \mid y \in Y\}$ is a partition of $\bigcup_{y \in Y} \mathrm{Succ}_{S^*}^{R\uparrow}(\Gamma_y)$, it is sufficient to show that $S' \in \mathcal{S}(N)$. Towards a contradiction, assume there is a node $w \in V(N) - \rho_N$ that does not have exactly one parent in S' and let u^* be the parent of w in S^*. Clearly, for each $y \in Y$, we have $w \notin \Gamma_y$ as, otherwise, $\mathrm{Pred}_{S'}(w) = \mathrm{Pred}_{S^y}(w)$. Further, $w \in \bigcup_{y \in Y} \mathrm{YW}_y^{\Gamma}$ as, otherwise, $\mathrm{Pred}_{S'}(w) = \mathrm{Pred}_{S^*}(w)$.

First, suppose w has no parent in S'. Then, $u^*w \in \bigcup_{y \in Y} A_y(S^*)$ that is, $u^* \in \Gamma_y$ for some $y \in Y$, but $w \notin A_y(S^y)$. But since $S^y \in \mathcal{S}(N)$, we know that w has a parent in S^y (which is not u^* since $w \notin A_y(S^y)$), implying that w is a reticulation in N. Thus, $w \in \mathrm{Succ}_{S^*}^{R\uparrow}(\Gamma_y) \subseteq \bigcup_{y' \in Y} \mathrm{Succ}_{S_{y'}}^{R\uparrow}(\Gamma_{y'})$ so there is some $y' \in Y$ with $w \in \mathrm{Succ}_{S_{y'}}^{R\uparrow}(\Gamma_{y'})$ (note that $y \neq y'$ is possible). But then, $S_{y'}$ contains an arc $uw \in A_{y'}(S_{y'})$ which is in S' by construction, thus contradicting w having no parents in S'.

Second, suppose that w has at least two distinct parents u and u^* in S' and note that, again, w is a reticulation in N. Since S^* is a switching, at least one of them, say u, is such that $uw \in \bigcup_{y \in Y} A_y(S^y)$. However, since the $\mathrm{Succ}_{S^y}^{R\uparrow}(\Gamma_y)$ are disjoint and each S^y is a switching, we cannot have $u^*w \in \bigcup_{y \in Y} A_y(S^y)$. Thus, $u^*w \in A(S^*) \setminus \bigcup_{y \in Y} A_y(S^*)$. However, since $\bigcup_{y \in Y} \mathrm{Succ}_{S^*}^{R\uparrow}(\Gamma_y) = \bigcup_{y \in Y} \mathrm{Succ}_{S^y}^{R\uparrow}(\Gamma_y)$, we know that $uw \in A_{S^*}(\Gamma_y)$ for some $y \in Y$. But then, w has two parents in S^* contradicting $S^* \in \mathcal{S}(N)$. □

In the following, we prove the semantics of the table $Q^{\psi_x}_{x,c_x}$. For all $i \leq t$, abbreviate $\bigcup_{1\leq j\leq i} \Gamma_{v_j} =: Z_i$.

Claim 5 *Let $1 \leq i \leq t$, let $c_x \in \phi(x)$, and let $R' \subseteq R(N)$. If $\mathcal{S}^{Z_i \to R'}(N) = \emptyset$, then $Q^{\psi_x}_{x,c_x}[i, R'] = \infty$. Otherwise, let $S_i \in \mathcal{S}^{Z_i \to R'}(N)$ and $\psi_i : V(N) \to C$ such that (a) $\psi_i \trianglelefteq \phi$, (b) ψ_i coincides with $\psi_x[x \to c_x]$ on $\bigcup_{j\leq i} \mathrm{YW}^\Gamma_{v_j}$ and (c) $\sum_{j\leq i} \sum_{uw\in A_{v_j}(S_i)} \delta_{\psi_i}(u,w)$ is minimum among all such S_i and ψ_i and*

$$Q^{\psi_x}_{x,c_x}[i, R'] = \sum_{j\leq i} \sum_{uw\in A_{v_j}(S_i)} \delta_{\psi_i}(u,w)$$

Proof The proof is by induction on i, noting that $\psi_x[x \to c_x]\,|_{\mathrm{YW}^\Gamma_{v_i}} = \psi_i\,|_{\mathrm{YW}^\Gamma_{v_1}}$ by Claim 5(b).

Case $i = 1$: By (5), $Q^{\psi_x}_{x,c_x}[0, R' \setminus R^*] \neq \infty$ only if $R^* = R' \subseteq \mathrm{Succ}^{R\uparrow}_N(Z_1)$ and $T^{\mathcal{SW}}[v_1, \psi_1\,|_{\mathrm{YW}^\Gamma_{v_1}}, R^*] \neq \infty$. However, if $\mathcal{S}^{Z_i \to R'}(N) = \emptyset$ then, by induction hypothesis (of the lemma), $T^{\mathcal{SW}}[v_1, \psi_1\,|_{\mathrm{YW}^\Gamma_{v_1}}, R'] = \infty$ and so $Q^{\psi_x}_{x,c_x}[0, R' \setminus R^*] = \infty$. Furthermore, S_1, ψ_1, and R' satisfy the conditions of the lemma for v_1, so we can employ the induction hypothesis of the lemma. Thus,

$$Q^{\psi_x}_{x,c_x}[1, R'] = 0 + T^{\mathcal{SW}}[v_1, \psi_1\,|_{\mathrm{YW}^\Gamma_{v_1}}, R'] \overset{\text{IH lemma}}{=} \sum_{uw\in A_{v_1}(S_1)} \delta_{\psi_1}(u,w)$$

Case $i > 1$: First, by (5), $Q^{\psi_x}_{x,c_x}[i, R'] \neq \infty$ only if $Q^{\psi_x}_{x,c_x}[i-1, R' \setminus R^*] \neq \infty$ and $T^{\mathcal{SW}}[v_i, \psi_i\,|_{\mathrm{YW}^\Gamma_{v_i}}, R^*] \neq \infty$. By induction hypotheses (of the claim and the lemma), there are switchings S_{i-1} and S' of N with $\mathrm{Succ}^{R\uparrow}_{S_{i-1}}(Z_{i-1}) = R' \setminus R^*$ and $\mathrm{Succ}^{R\uparrow}_{S'}(\Gamma_{v_i}) = R^*$. Now, we construct a digraph $S_i := (V(N), (A(S_{i-1} \setminus A_{v_i}(S_{i-1})) \cup A_{v_i}(S'))$ and show that $S_i \in \mathcal{S}^{Z_i \to R'}(N)$. Since $\mathrm{Succ}^{R\uparrow}_{S_i}(Z_i) = \mathrm{Succ}^{R\uparrow}_{S_{i-1}}(Z_{i-1}) \uplus \mathrm{Succ}^{R\uparrow}_{S'}(\Gamma_{v_i}) = (R' \setminus R^*) \uplus R^* = R'$, it is sufficient to show that S_i can be turned into a switching of N without changing $\mathrm{Succ}^{R\uparrow}_{S_i}(Z_i)$. To this end, suppose that there is a node $w \neq \rho_N$ of N that does not have exactly one parent in S_i. Since S_{i-1} and S' are switchings, w has parents u_{i-1} and u' in S_{i-1} and S', respectively. If w has no parent in S_i, then $u_{i-1}w \in A_{v_i}(S_{i-1})$ and $u'w \notin A_{v_i}(S')$ and, thus, $u_{i-1} \leq_\Gamma v_i <_\Gamma u'$, implying $u' \neq u_{i-1}$ as well as $w \in \mathrm{YW}^\Gamma_{v_i}$ and $w \notin R'$. Then, we can just add the arc $u'w$ to S_i without changing $\mathrm{Succ}^{R\uparrow}_{S_i}(Z_i)$. If w has at least two parents, then u_{i-1} and u' are both parents of w in S_i, that is, $u_{i-1}w \notin A_{v_i}(S_{i-1})$ and $u'w \in A_{v_i}(S')$ and, thus, $u' <_\Gamma v_i <_\Gamma u_{i-1}$, implying $u' \neq u_{i-1}$ as well as $w \in \mathrm{YW}^\Gamma_{v_i}$ and $w \in R^*$. But then, we can remove $u_{i-1}w$ from S_i without changing $\mathrm{Succ}^{R\uparrow}_{S_i}(v_i)$. Repeating this argument, we can turn S_i into a switching of N with $\mathrm{Succ}^{R\uparrow}_{S_i}(Z_i) = R'$, implying that $\mathcal{S}^{Z_i \to R'}(N) \neq \emptyset$. For the second part of the claim, we show both inequalities separately.

"$\leq$": Let $S_i \in \mathcal{S}^{Z_i \to R'}(N)$ and $\psi_i : V(N) \to C$ ψ_i coincides with $\psi_x[x \to c_x]$ on $\bigcup_{j\leq i} \mathrm{YW}^\Gamma_{v_j}$ and $\sum_{j\leq i} \sum_{uw\in A^{(S_i)}_{v_j}} \delta_{\psi_i}(u,w)$ is minimum among all such S_i and ψ_i. Further, let $R^* := \mathrm{Succ}^{R\uparrow}_{S_i}(\Gamma_{v_i})$. Note that $\mathrm{Succ}^{R\uparrow}_{S_i}(Z_{i-1})$ and $\mathrm{Succ}^{R\uparrow}_{S_i}(v_i) = R^*$ are disjoint since S_i is a switching, implying $\mathrm{Succ}^{R\uparrow}_{S_i}(Z_{i-1}) = R' \setminus R^*$ and, thus, $Q^{\psi_x}_{x,c_x}[i-1, R' \setminus R^*]$ and $T^{\mathcal{SW}}[v_i, \phi_i, R^*]$ are finite by induction hypotheses. Then, as $R^* \subseteq R' \cap \mathrm{Succ}^{R\uparrow}_N(\Gamma_{v_i})$, we know that R^* participates in the minimum of (5). Thus,

$$\begin{aligned} Q^{\psi_x}_{x,c_x}[i, R'] &\leq Q^{\psi_x}_{x,c_x}[i-1, R' \setminus R^*] + T^{\mathcal{SW}}[v_i, \psi_i\,|_{\mathrm{YW}^\Gamma_{v_i}}, R^*] \\ &\overset{\substack{\text{IH claim}\\ \text{IH lemma}}}{\leq} \sum_{j\leq i-1} \sum_{uw\in A_{v_j}(S_i)} \delta_{\psi_i}(u,w) + \sum_{uw\in A_{v_i}(S_i)} \delta_{\psi_i}(u,w) \\ &= \sum_{j\leq i} \sum_{uw\in A_{v_j}(S_i)} \delta_{\psi_i}(u,w) \end{aligned}$$

"$\geq$": Clearly, this direction is trivial if $Q^{\psi_x}_{x,c_x}[i, R']$ is infinite, so suppose it is finite. By (5), there is some $R^* \subseteq R' \cap \mathrm{Succ}^{R\uparrow}_N(\Gamma_{v_i})$ with $Q^{\psi_x}_{x,c_x}[i, R'] = Q^{\psi_x}_{x,c_x}[i-1, R' \setminus R^*] + T^{\mathcal{SW}}[v_i, \psi_i\mid_{\mathrm{YW}^{\Gamma}_{v_i}}, R^*]$. First, since $T^{\mathcal{SW}}[v_i, \psi_i\mid_{\mathrm{YW}^{\Gamma}_{v_i}}, R^*] \neq \infty$, the induction hypothesis (of the lemma) guarantees that there is some $S^* \in \mathcal{S}^{\Gamma_{v_i} \to R^*}(N)$ and $\psi^* : V(N) \to C$ such that (a) $\psi_i\mid_{\mathrm{YW}^{\Gamma}_{v_i}} \trianglelefteq \psi^* \trianglelefteq \phi$, (b) (S^*, ψ^*) is optimal on $A_{v_i}(S^*)$, and (c) $T^{\mathcal{SW}}[v_i, \psi_i\mid_{\mathrm{YW}^{\Gamma}_{v_i}}, R^*] = \sum_{uw\in A_{v_i}(S^*)} \delta_{\psi^*}(u,w)$. Second, since $Q^{\psi_x}_{x,c_x}[i-1, R' \setminus R^*] \neq \infty$, the induction hypothesis (of the claim) guarantees that there are $S_{i-1} \in \mathcal{S}^{Z_{i-1}\to R'\setminus R^*}(N)$ and $\psi_{i-1} : V(N) \to C$ such that (a) $\psi_{i-1} \trianglelefteq \phi$, (b) ψ_{i-1} coincides with $\psi_x[x \to c_x]$ on $\bigcup_{j<i} \mathrm{YW}^{\Gamma}_{v_j}$, (c) $\sum_{j<i}\sum_{uw\in A_{v_j}(S_{i-1})} \delta_{\psi_{i-1}}(u,w)$ is minimal among all such S_{i-1} and ψ_{i-1}, and (d) $Q^{\psi_x}_{x,c_x}[i-1, R' \setminus R^*] = \sum_{j<i}\sum_{uw\in A_{v_j}(S_{i-1})} \delta_{\psi_{i-1}}(u,w)$. Finally, we construct a new solution S' by replacing S_i by S^* on Γ_{v_i} and by S_{i-1} on Z_{i-1} and we use Claim 5(c) to show that the cost of S_i is at most that of S'. More formally, let

$$S' := \left(V(N), \left(A(S_i) \setminus \bigcup_{j\leq i} A_{v_j}(S_i)\right) \cup \bigcup_{j<i} A_{v_j}(S_{i-1}) \cup A_{v_i}(S^*)\right) \tag{11}$$

Since $\{v_1, v_2, \ldots, v_i\}$ is an anti-chain in Γ and $\{\mathrm{Succ}^{R\uparrow}_{S_{i-1}}(Z_{i-1}), \mathrm{Succ}^{R\uparrow}_{S^*}(\Gamma_{v_i})\} = \{R^*, R' \setminus R^*\}$ is a partition of $\mathrm{Succ}^{R\uparrow}_{S_i}(Z_i) = R'$, Claim 4 implies $S' \in \mathcal{S}^{Z_i \to R'}(N)$. Further, let $\psi' : V(N) \to C$ such that, for all $a \in A(S')$, $\psi'(a) := \psi_{i-1}(a)$ if $a \in A_{Z_i}(S_{i-1})$, $\psi'(a) := \psi^*(a)$ if $a \in A_{v_i}(S^*)$, and $\psi'(a) := \psi_i(a)$, otherwise. Note that $\psi' \trianglelefteq \phi$. Further, ψ_i and ψ_{i-1} coincide on $\mathrm{YW}^{\Gamma}_{Z_{i-1}}$ and, thus, ψ' and ψ_{i-1} coincide on all nodes touched by $A_{Z_{i-1}}(S') = A_{Z_{i-1}}(S_{i-1})$. Further, ψ_i and ψ^* coincide on $\mathrm{YW}^{\Gamma}_{v_i}$ and, thus, ψ' and ψ^* coincide on all nodes touched by $A_{v_i}(S') = A_{v_i}(S^*)$. Thus,

$$\begin{aligned}
Q^{\psi_x}_{x,c_x}[i, R'] &= Q^{\psi_x}_{x,c_x}[i-1, R' \setminus R^*] + T^{\mathcal{SW}}[v_i, \psi^*\mid_{\mathrm{YW}^{\Gamma}_{v_i}}, R^*] \\
&\overset{(c),(g)}{=} \sum_{j<i}\sum_{uw\in A_{v_j}(S_{i-1})} \delta_{\psi_{i-1}}(u,w) + \sum_{uw\in A_{v_i}(S^*)} \delta_{\psi^*}(u,w) \\
&\overset{\text{df. }\psi',(11)}{=} \sum_{j<i}\sum_{uw\in A_{v_j}(S')} \delta_{\psi'}(u,w) + \sum_{uw\in A_{v_i}(S')} \delta_{\psi'}(u,w) \\
&= \sum_{uw\in A_{Z_i}(S')} \delta_{\psi'}(u,w) \overset{\text{Claim 5(c)}}{\geq} \sum_{j\leq i}\sum_{uw\in A_{v_j}(S_i)} \delta_{\psi_i}(u,w)
\end{aligned}$$

$\square$

Having established the semantics of $Q^{\psi_x}_{x,c_x}$, we can finish proving Case 2 of Lemma 8s. First, consider the case that $\mathcal{S}^{\Gamma_x \to R^x}(N) = \varnothing$ and assume that $T^{\mathcal{SW}}[x, \psi_x, R^x] \neq \infty$. By Eq. (4) and Claim 5, there is some c_x and $R^* \subseteq R^x \cap \mathrm{Succ}^{R\uparrow}_N(x)$ such that $\mathcal{S}^{Z_t \to R^x\setminus R^*}(N) \neq \varnothing$ or $\mathcal{S}^{Z_t \to (R^x\setminus R^*)\cup(\{x\}\cup R(N))}(N) \neq \varnothing$. Let S' be a switching in one of these sets and note that $\mathrm{Succ}^{R\uparrow}_{S'}(\Gamma_x) = R^x \setminus R^*$. If there is some $y \in R^x \setminus \mathrm{Succ}^{R\uparrow}_{S'}(\Gamma_x)$, then $y \in R^*$ and S' contains an arc zy for some $z \notin \Gamma_x$, implying that we can swap zy for xy in S' without affecting $\mathrm{Succ}^{R\uparrow}_{S'}(Z_t)$ or S' being a switching. Thus, we can assume without loss of generality that $\mathrm{Succ}^{R\uparrow}_{S'}(\Gamma_x) = R^x$. But then, $S' \in \mathcal{S}^{\Gamma_x \to R^x}$ contradicting $\mathcal{S}^{\Gamma_x \to R^x} = \varnothing$. In the following, we thus assume that $\mathcal{S}^{\Gamma_x \to R^x} \neq \varnothing$ and we show both directions of the lemma separately.

"$\leq$": Let $c_x := \psi(x) \in \phi(x)$, let $R^* := \mathrm{Succ}^{R\uparrow}_S(x)$, and note that $R^* = \mathrm{Succ}^{R\uparrow}_S(x) \subseteq \mathrm{Succ}^{R\uparrow}_S(\Gamma_x) = R^x$. Further, let $y := \mathrm{Pred}_S(x)$ be the parent of x in S. Since Γ agrees with N (and, thus, with S) we know that either $x <_\Gamma y$ or $x >_\Gamma y$. If $x <_\Gamma y$, that is, $y \in \mathrm{Pred}^{\downarrow}_N(x)$, then $\mathrm{Succ}^{R\uparrow}_S(Z_t) = \mathrm{Succ}^{R\uparrow}_S(\Gamma_x) \setminus \mathrm{Succ}^{R\uparrow}_S(x) = R^x \setminus R^*$ and, by Claim 5,

$$\begin{aligned}
&\sum_{uw\in A_{Z_t}(S)} \delta_{\psi}(u,w) + \sum_{uw\in A^{\downarrow}_{\{x\}}(S)} \delta_{\psi}(u,w) \\
&\geq Q^{\psi_x}_{x,c_x}[t, R^x \setminus R^*] + \delta_{\psi}(x,y) \\
&\geq Q^{\psi_x}_{x,c_x}[t, R^x \setminus R^*] + \min_{yx\in A^{\downarrow}_{\{x\}}(N)} \delta(c_x, \psi(y))
\end{aligned} \tag{12}$$

If $x >_\Gamma y$, that is, $y \in \mathrm{Pred}^{\uparrow}_N(x)$, then $\mathrm{Succ}^{R\uparrow}_S(Z_t) = (R(N) \cap\{x\}) \cup (\mathrm{Succ}^{R\uparrow}_S(\Gamma_x) \setminus \mathrm{Succ}^{R\uparrow}_S(x)) = (R(N) \cap \{x\}) \cup (R^x \setminus R^*)$ and, by Claim 5,

$$\sum_{uw \in A_{Z_t}(S)} \delta_\psi(u,w) + \sum_{uw \in A^{\downarrow}_{\{x\}}(S)} \delta_\psi(u,w) \geq Q^{\psi_x}_{x,c_x}[t, (R(N) \cap \{x\}) \cup R^x \setminus R^*] \tag{13}$$

Then, since c_x and R^* are valid choices for the minima in (4), we have

$$T^{\mathcal{SW}}[x,\psi_x,R^x] \overset{(4),(12),(13)}{\leq} \sum_{r \in R^* \cup \mathrm{Succ}^{T\uparrow}_N(x)} \delta(c_x,\psi(r)) + \sum_{uw \in A_{Z_t}(S)} \delta_\psi(u,w) + \sum_{uw \in A^{\downarrow}_{\{x\}}(S)} \delta_\psi(u,w)$$
$$= \sum_{xr \in A^{\uparrow}_{\{x\}}(S)} \delta_\psi(x,r) + \sum_{uw \in A_{Z_t}(S)} \delta_\psi(u,w) + \sum_{uw \in A^{\downarrow}_{\{x\}}(S)} \delta_\psi(u,w) = \sum_{uw \in A_x(S)} \delta_\psi(u,w)$$

"$\geq$": Suppose that $T^{\mathcal{SW}}[x,\psi_x,R^x] \neq \infty$ as, otherwise, this direction is trivial. We consider each case of the minimum in (4) individually (although both cases are analogous).

Case 2.1: $\mathrm{Pred}^{\downarrow}_N(x) \neq \emptyset$ and there are $c_x \in \phi(x)$ and $R^* \subseteq R^x \cap \mathrm{Succ}^{R\uparrow}_N(x)$ such that

$$T^{\mathcal{SW}}[x,\psi_x,R^x] = \sum_{r \in R^* \cup \mathrm{Succ}^{T\uparrow}_N(x)} \delta(c_x,\psi_x(r)) + Q^{\psi_x}_{x,c_x}[t, R^x \setminus R^*] + \min_{y \in \mathrm{Pred}^{\downarrow}_N(x)} \delta(c_x,\psi_x(y)) \tag{14}$$

By Claim 5, there is some $S' \in \mathcal{S}^{Z_t \to R^x \setminus R^*}(N)$ and some $\psi' : V(N) \to C$ such that (a) $\psi' \trianglelefteq \phi$, (b) ψ' coincides with $\psi_x[x \to c_x]$ on $\bigcup_{i \leq t} \mathrm{YW}^{\Gamma}_{v_i}$ (recall that $x \in \bigcup_{i \leq t} \mathrm{YW}^{\Gamma}_{v_i}$) (c) $\sum_{uw \in A_{Z_t}(S')} \delta_{\psi'}(u,w)$ is minimum among all such S' and ψ' and

$$Q^{\psi_x}_{x,c_x}[t, R^x \setminus R^*] = \sum_{uw \in A_{Z_t}(S')} \delta_{\psi'}(u,w) \tag{15}$$

From S' we construct a switching $S^* \in \mathcal{S}^{\Gamma_x \to R^x}(N)$ by 1. swapping each arc $zr \in A(S')$ with $r \in R^*$ for xr (which exists in N since $R^* \subseteq \mathrm{Succ}^{R\uparrow}_N(x)$), 2. swapping each arc $xr \in A(S')$ with $r \notin R^x$ for an arc zr with $z \notin \Gamma_x$ (which exists in N since $\mathcal{S}^{\Gamma_x \to R^x}(N) \neq \emptyset$), and 3. swapping the arc $yx \in A^{\downarrow}_{\{x\}}(S')$ with an arc $zx \in \mathrm{Pred}^{\downarrow}_N(x) \times \{x\}$ minimizing $\delta_{\psi'}(x,z)$. Since this operation does not change the in-degree of any node, S^* is still a switching of N and we have $\mathrm{Succ}^{R\uparrow}_{S^*}(x) = R^*$ and $A_{Z_t}(S') = A_{Z_t}(S^*)$ by construction. Thus, $\mathrm{Succ}^{R\uparrow}_{S^*}(Z_t) = R^x \setminus R^*$ and $\mathrm{Succ}^{R\uparrow}_{S^*}(\Gamma_x) = R^x$. Altogether,

$$T^{\mathcal{SW}}[x,\psi_x,R^x] \overset{(14),(15)}{=} \sum_{r \in R^* \cup \mathrm{Succ}^{T\uparrow}_N(x)} \delta(c_x,\psi_x(r)) + \sum_{uw \in A_{Z_t}(S')} \delta_{\psi'}(u,w) + \min_{y \in \mathrm{Pred}^{\downarrow}_N(x)} \delta(c_x,\psi_x(y))$$
$$= \sum_{r \in R^* \cup \mathrm{Succ}^{T\uparrow}_N(x)} \delta(c_x,\psi_x(r)) + \sum_{uw \in A_{Z_t}(S^*)} \delta_{\psi'}(u,w) + \min_{y \in \mathrm{Pred}^{\downarrow}_N(x)} \delta(c_x,\psi_x(y))$$
$$\overset{\psi_x = \psi'|_{\mathrm{YW}^{\Gamma}_x}}{=} \sum_{xr \in A^{\uparrow}_{\{x\}}(S^*)} \delta_{\psi'}(x,r) + \sum_{uw \in A_{Z_t}(S^*)} \delta_{\psi'}(u,w) + \sum_{yx \in A^{\downarrow}_{\{x\}}(S^*)} \delta_{\psi'}(y,x)$$
$$= \sum_{uw \in A_x(S^*)} \delta_{\psi'}(u,w) \overset{\text{Lemma 8(b)}}{\geq} \sum_{uw \in A_x(S)} \delta_\psi(u,w)$$

Case 2.2: $\mathrm{Pred}^{\uparrow}_N(x) \neq \emptyset$ and there are $c_x \in \phi(x)$ and $R^* \subseteq R^x \cap \mathrm{Succ}^{R\uparrow}_N(x)$ such that

$$T^{\mathcal{SW}}[x,\psi_x,R^x] = \sum_{r \in R^* \cup \mathrm{Succ}^{T\uparrow}_N(x)} \delta(c_x,\psi_x(r)) + Q^{\psi_x}_{x,c_x}[t, (R(N) \cap \{x\}) \cup R^x \setminus R^*] \tag{16}$$

Abbreviate $R' := (R(N) \cap \{x\}) \cup R^x \setminus R^*$. By Claim 5, there is some $S' \in \mathcal{S}^{Z_t \to R'}(N)$ and some $\psi' : V(N) \to C$ such that (a) $\psi' \trianglelefteq \phi$, (b) ψ' coincides with $\psi_x[x \to c_x]$ on $\bigcup_{i \leq t} \mathrm{YW}^{\Gamma}_{v_i}$, (c) $\sum_{uw \in A_{Z_t}(S')} \delta_{\psi'}(u,w)$ is minimum among all such S' and ψ' and

$$Q^{\psi_x}_{x,c_x}[t, R'] = \sum_{uw \in A_{Z_t}(S')} \delta_{\psi'}(u,w) \tag{17}$$

We construct a switching $S^* \in \mathcal{S}^{\Gamma_x \to R^x}(N)$ by 1. swapping each arc $zr \in A(S')$ with $r \in R^*$ forxr (which exists in N since $R^* \subseteq \mathrm{Succ}_N^{R\uparrow}(x)$) and 2. swapping each arc $xr \in A(S')$ with $r \notin R^x$ for an arc zr with $z \notin \Gamma_x$ (which exists in N since $\mathcal{S}^{\Gamma_x \to R^x}(N) \neq \varnothing$). Since this operation does not change the in-degree of any node, S^* is still a switching of N and we have $\mathrm{Succ}_{S^*}^{R\uparrow}(x) = R^*$ and $A_{Z_t}(S') = A_{Z_t}(S^*)$ by construction. Thus, $\mathrm{Succ}_{S^*}^{R\uparrow}(Z_t) = R'$ and $\mathrm{Succ}_{S^*}^{R\uparrow}(\Gamma_x) = R^x$. Further, note that if x is a tree node, then $\mathrm{Pred}_N^{\uparrow}(x) \neq \varnothing$ implies $A_{\{x\}}^{\downarrow}(S^*) \subseteq A_{\{x\}}^{\downarrow}(N) = \varnothing$ and, otherwise, $x \in R'$ implying $A_{\{x\}}^{\downarrow}(S^*) = \varnothing$. Altogether,

$$\begin{aligned}
T^{\mathcal{SW}}[x, \psi_x, R^x] &\overset{(16),(17)}{=} \sum_{r \in R^* \cup \mathrm{Succ}_N^{\mathcal{T}\uparrow}(x)} \delta\,(c_x, \psi_x(r)) + \sum_{uw \in A_{Z_t}(S')} \delta_{\psi'}(u,w) \\
&= \sum_{r \in R^* \cup \mathrm{Succ}_N^{\mathcal{T}\uparrow}(x)} \delta\,(c_x, \psi_x(r)) + \sum_{uw \in A_{Z_t}(S^*)} \delta_{\psi'}(u,w) \\
&\overset{\psi_x = \psi'|_{\mathrm{YW}_x^\Gamma}}{=} \sum_{xr \in A_{\{x\}}^{\uparrow}(S^*)} \delta_{\psi'}(x,r) + \sum_{uw \in A_{Z_t}(S^*)} \delta_{\psi'}(u,w) \\
&\overset{A_{\{x\}}^{\downarrow}(S^*) = \varnothing}{=} \sum_{uw \in A_x(S^*)} \delta_{\psi'}(u,w) \overset{\text{Lemma 8(b)}}{\geq} \sum_{uw \in A_x(S)} \delta_{\psi}(u,w)
\end{aligned}$$

□

Lemma 9 *Let $x \in V(N)$, let $i \in \mathbb{N}$, let $\lambda : \mathrm{YW}_x^\Gamma \to 2^C$, let $\eta, \eta' : \mathrm{YW}_x^\Gamma \to \mathbb{N}$, and let $\psi, \psi' : \mathrm{YW}_x^\Gamma \to 2^C$ such that $\psi' \trianglelefteq \psi \trianglelefteq \lambda$ and $\vec{0}\,[x \to \rho(x)] \trianglelefteq \eta' \trianglelefteq \eta$. Then,*

$$T^{\mathcal{PT}}[x, \lambda, \psi', \eta'] \leq T^{\mathcal{PT}}[x, \lambda, \psi, \eta]$$
$$\text{and} \quad Q_x^\lambda[i, \psi', \eta'] \leq Q_x^\lambda[i, \psi, \eta]$$

Proof Note that the inequality on Q_x^λ trivially holds if $Q_x^\lambda[i, \psi, \eta] = \infty$ and, similarly for $T^{\mathcal{PT}}$. The proof is based on the observation that the transformations done to ψ and η in Equations (7) and (8) are monotone.

Claim 6 *Let $U, D \in \mathbb{N}$. The following functions (acting on functions) are montone*

$$f_{U,D}(\psi) := \psi \begin{bmatrix} x \to D \\ \forall_{y \in \mathrm{Succ}_N^{\uparrow}(x)} y \to \psi(y) \setminus U \end{bmatrix} \quad g_{U,D}(\eta) := \eta \begin{bmatrix} x \to |U| \dot{-} \sum_{u \in \mathrm{Pred}_N^{\downarrow}(x)} |\lambda_x(u)| \\ \forall_{y \in \mathrm{Succ}_N^{\uparrow}(x)} y \to \eta(y) \dot{-} |U| \end{bmatrix}$$

Let $\psi, \psi' : \mathrm{YW}_x^\Gamma \to 2^C$ with $\psi' \trianglelefteq \psi$. Then, for all $y \in \mathrm{YW}_x^\Gamma$, Further, for all $y \in \mathrm{Succ}_N^{\uparrow}(x)$, we have $f(\psi')(y) = \psi'(y) \setminus U \subseteq \psi(y) \setminus U = f(\psi)(y)$.

$$\begin{aligned}
f(\psi')(y) &= \begin{cases} D & \text{if } x = y \\ \psi'(y) \setminus U & \text{if } y \in \mathrm{Succ}_N^{\uparrow}(x) \\ \psi'(y) & \text{otherwise} \end{cases} \trianglelefteq \begin{cases} D & \text{if } x = y \\ \psi(y) \setminus U & \text{if } y \in \mathrm{Succ}_N^{\uparrow}(x) \\ \psi(y) & \text{otherwise} \end{cases} \\
&= f(\psi)(y)
\end{aligned}$$

The proof for $g_{U,D}$ is completely analogous. □

With Claim 6, we can show that monotonicity of Q_x^λ implies monotonicity of $T^{\mathcal{PT}}$.

Claim 7 *Let $v_1, v_2, \ldots, v_t$ be the children of x in Γ and suppose that Q_x^λ is monotone. Then, $T^{\mathcal{PT}}$ is monotone.*

Proof If $T^{\mathcal{PT}}[x, \lambda, \phi, \eta] \neq \infty$, there are $D \subseteq U \subseteq \phi(x)$ such that the minimum in Equation (7) in Definition 7 is attained, that is,

$$\begin{aligned}
T^{\mathcal{PT}}[x, \lambda, \phi, \eta] &= Q_x^{\lambda x \to U}[0, f_{U,D}(\phi), g_{U,D}(\eta)] + c_{U,D} \\
&= Q_x^{\lambda}[0, f_{U,D}(\phi), g_{U,D}(\eta)] + c_{U,D}^*
\end{aligned}$$

for some constants $c_{U,D}$ and $c_{U,D}^*$ that are independant of ϕ and η. Since, by assumption, Q_x^λ is monotone for all λ and both $f_{U,D}$ and $g_{U,D}$ are monotone by Claim 6, we conclude

$$T^{\mathcal{PT}}[x, \lambda, \psi, \eta] \geq Q_x^{\lambda}[0, f_{U,D}(\psi), g_{U,D}(\eta)] + c^*_{U,D}$$
$$\geq Q_x^{\lambda}[0, f_{U,D}(\psi'), g_{U,D}(\eta')] + c^*_{U,D} \geq T^{\mathcal{PT}}[x, \lambda, \psi', \eta']$$

Note the last "$\geq$" since we only know that this particular value participates in the minimum that forms $T^{\mathcal{PT}}[x, \lambda, \psi', \eta']$, while this minimum may be attained at an even smaller value. □

By Claim 7, in order to prove Lemma 9, it is sufficent to show that Q_x^{λ} is monotone. This proof is by induction on the height of x in Γ and the value of the first argument i of Q_x^{λ}.

For the induction base, suppose that x is a leaf of Γ and note that x has $t = 0$ children. If $Q_x^{\lambda}[0, \psi, \eta] \neq \infty$, we have $\psi = \overrightarrow{0}$ and $\eta = \overrightarrow{0}[x \to \rho(x)]$. But then, $\psi' = \psi$ and $\eta' = \eta$, implying $Q_x^{\lambda}[0, \psi', \eta'] = Q_x^{\lambda}[0, \psi, \eta]$.

For the induction step, let x have t children $v_1, v_2, \ldots, v_t$ and let $0 < i \leq t$. First, let $\psi^* \trianglelefteq \psi \mid_{\mathrm{YW}^{\Gamma}_{v_i}}$ and $\eta^* \trianglelefteq \eta \mid_{\mathrm{YW}^{\Gamma}_{v_i}}$ be such that the minimum in Equation (8) in Definition 7 is attained, that is, $Q_x^{\lambda}[i, \psi, \eta] = Q_x^{\lambda}[i-1, \psi - \psi^*, \eta - \eta^*] + T^{\mathcal{PT}}[v_i, \lambda \mid_{\mathrm{YW}^{\Gamma}_{v_i}}, \psi^*, \eta^*]$. Further, let ψ'^* and η'^* be defined as $\psi'^*(y) := \psi'(y) \cap \psi^*(y)$ and $\eta'^*(y) := \min\{\eta'(y), \eta^*(y)\}$. Clearly, $\psi'^* \trianglelefteq \psi'$ and $\psi'^* \trianglelefteq \psi^*$ and $\eta'^* \trianglelefteq \eta'$ and $\eta'^* \trianglelefteq \eta^*$. Further, for all y,

$$(\psi' - \psi'^*)(y) = \psi'(y) \setminus (\psi'(y) \cap \psi^*(y)) = \psi'(y) \setminus \psi^*(y)$$
$$\subseteq \psi(y) \setminus \psi^*(y) = (\psi - \psi^*)(y) \tag{18}$$

$$(\eta' - \eta'^*)(y) = \eta'(y) - \min\{\eta'(y), \eta^*(y)\} = \eta'(y) \dot{-} \eta^*(y)$$
$$\leq \eta(y) \dot{-} \eta^*(y) = (\eta - \eta^*)(y), \tag{19}$$

so $\psi' - \psi'^* \trianglelefteq \psi - \psi^*$ and $\eta' - \eta'^* \trianglelefteq \eta - \eta^*$. Since ψ'^* and η'^* participate in the minimum in the definition of $Q_x^{\lambda}[i, \psi', \eta']$,

$$Q_x^{\lambda}[i, \psi, \eta] = Q_x^{\lambda}[i-1, \psi - \psi^*, \eta - \eta^*] + T^{\mathcal{PT}}[v_i, \lambda \mid_{\mathrm{YW}^{\Gamma}_{v_i}}, \psi^*, \eta^*]$$
$$\overset{IH,(18),(19)}{\geq} Q_x^{\lambda}[i-1, \psi' - \psi'^*, \eta' - \eta'^*] + T^{\mathcal{PT}}[v_i, \lambda \mid_{\mathrm{YW}^{\Gamma}_{v_i}}, \psi'^*, \eta'^*]$$
$$\geq Q_x^{\lambda}[i, \psi', \eta']$$

□

Lemma 10 *Let Γ be a tree agreeing with N, let $x \in V(N)$, let $\psi_x, \lambda_x : \mathrm{YW}^{\Gamma}_x \to 2^c$ and $\eta_x : \mathrm{YW}^{\Gamma}_x \to \mathbb{N}$. Let f minimize $\mathrm{cost}_{(f)}$ among all lineage functions in $\mathcal{LF}_{N,\phi}$ such that, for all $w \in \mathrm{YW}^{\Gamma}_x$, $\lambda_x(w) = f(w)$, $\psi_x(w) = f(w) \cap \bigcup_{u \in \mathrm{Pred}_N^{\uparrow x}(w)} f(u)$, and $\eta_x(w) \leq \sum_{u \in \mathrm{Pred}_{N \uparrow x}(w)} |f(u)|$. If there are no such f, then $T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x = \infty]$. Otherwise,*

$$T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x] = \sum_{z \leq_{\Gamma} x} \mathrm{cost}_f(z)$$

Proof Note that, if the cost of f is finite, then $|f(v)| \leq \sum_{u \in \mathrm{Pred}(v)} |f(u)|$ for all $v \neq \rho_N$ and $|f(\rho_N)| = 1$ by Definition 4. Again, the proof is by induction on the height of x in Γ.

Case 1: x is a leaf in Γ, that is, $t = 0$ and $\mathrm{Pred}_N^{\uparrow x}(v) \subseteq \{x\}$ for all v. Then, by Definition 7, $T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x]$ is finite if and only if

$$\psi_x \begin{bmatrix} x \to D \\ \forall_{y \in \mathrm{Succ}_N^{\uparrow}(x)} y \to \psi_x(y) \setminus U \end{bmatrix} = \overrightarrow{0}$$

and

$$\eta_x \begin{bmatrix} x \to |U| \dot{-} \sum_{r \in \mathrm{Pred}_N^{\downarrow}(x)} |\lambda_x(r)| \\ \forall_{y \in \mathrm{Succ}_N^{\uparrow}(x)} y \to \eta_x(y) \dot{-} |U| \end{bmatrix} = \overrightarrow{0}[x \to \rho(x)],$$

if and only if (considering the assignments of the above modifications of ψ_x and η_x individually) (a) $D = \varnothing$, (b) $|U| \dot{-} \sum_{r \in \mathrm{Pred}_N^{\downarrow}(x)} |\lambda_x(r)| = \rho(x)$ (c) for each $y \in \mathrm{Succ}_N(x)$,

$$U \supseteq \psi_x(y) = f(y) \cap f(x). \text{ and } \quad |U| \geq \eta_x(y) = |f(x)| \tag{20}$$

In this case, the table entry is assigned the cost $|U \setminus \bigcup_{r \in \mathrm{Pred}_N^{\downarrow}(x)} \lambda_x(r)| - \rho(x) = |U \setminus \bigcup_{r \in \mathrm{Pred}(x)} f(r)| - \rho(x)$.

If $x = \rho_N$, this simplifies to $|U| - 1$ and, since $|f(\rho_N)| = 1$, the cost is minimized by $U = f(\rho_N)$ and the table entry equals $0 = \mathrm{cost}_f(\rho_N)$. Thus, in the following, let $x \neq \rho_N$.

"$\leq$": Since (20) is satisfied for $U = f(x)$, the minimum over all U is at most the cost when choosing $U = f(x)$, which is $|f(x) \setminus \bigcup_{r \in \mathrm{Pred}(x)} f(r)| = \mathrm{cost}_f(x)$

"$\geq$": Towards a contradiction, assume that there is a U satisfying (20) such that $T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x] = |U \setminus \bigcup_{r\in \mathrm{Pred}(x)} f(r)| < |f(x) \setminus \bigcup_{r\in \mathrm{Pred}(x)} f(r)| = \mathrm{cost}_f(x)$. We show that $f' := f[x \to U]$ has less overall cost than f, contradicting its optimality. Since changing $f(x)$ to U only influences the cost of x and its children in N, it suffices to consider them. To this end, let y be any child of x in N. First, by (b), $|f'(x)| = |U| \leq \sum_{r\in \mathrm{Pred}_N(x)} |f'(r)|$ so $\mathrm{cost}_{f'}(x) = |U \setminus \bigcup_{r\in \mathrm{Pred}_N(x)} f(r)| < \mathrm{cost}_f(x)$ by assumption. Further, $|f'(y)| = |f(y)| \leq \sum_{u\in \mathrm{Pred}(y)} |f(u)| \leq \sum_{u\in \mathrm{Pred}(y)} |f'(u)|$ since $|U| \geq |f(x)|$ by (20). Finally, for each $y \in \mathrm{Succ}_N(x)$,

$$\begin{aligned}\mathrm{cost}_f(y) &= |f(y) \setminus \bigcup_{v\in \mathrm{Pred}(y)} f(v)| = |f(y) \setminus (f(x) \cup \bigcup_{v\in \mathrm{Pred}(y)-x} f(v))| \\ &\overset{(20)}{\geq} |f'(y) \setminus (U \cup \bigcup_{v\in \mathrm{Pred}(y)-x} f'(v)))| = \mathrm{cost}_{f'}(y)\end{aligned}$$

Case 2: x has children $v_1, v_2, \ldots, v_t$ with $t \geq 1$ in Γ. In the following, we abbreviate $Y_i := \bigcup_{j\leq i} \mathrm{YW}^{\Gamma}_{v_j}$ and $Z_i := \bigcup_{j\leq i} \Gamma_{v_j}$. Further, we call a lineage function f' *eligible* with respect to an anti-chain Y in Γ and functions λ', ψ', and η' if, for all $w \in \bigcup_{y\in Y} \mathrm{YW}^{\Gamma}_y$, we have $\lambda(w) = f'(w)$, $\psi'(w) \subseteq f'(w) \cap \bigcup_{y\in Y} \bigcup_{u\in \mathrm{Pred}^{\uparrow y}_N(w)} f'(u)$ and $\eta'(w) \leq \sum_{y\in Y} \sum_{u\in \mathrm{Pred}^{\uparrow y}_N(w)} |f(u)| + \rho(w)$ and the cost of f' is finite on $\bigcup_{y\in Y} \Gamma_y$. We first show how the table Q^{λ}_x is used to distribute lineages among the v_i.

Claim 8 *Let* $1 \leq i \leq t$, *Let* $\eta_i : Y_i \to \mathbb{N}$ *and let* $\lambda, \psi_i : Y_i \to 2^C$ *with* $\psi_i \trianglelefteq \lambda$. *Let* f_i *minimize* $\sum_{z\in Z_i} \mathrm{cost}_{f_i}(z) - \rho(x)$ *among all lineage functions that are eligible with respect to* Y_i, λ, ψ_i, *and* η_i. *If no such* f *exists, then* $Q^{\lambda}_x[i, \psi_i, \eta_i] = \infty$. *Otherwise,*

$$Q^{\lambda}_x[i, \psi_i, \eta_i] = \sum_{z\in Z_i} \mathrm{cost}_{f_i}(z) - \rho(x).$$

Proof The proof of the claim is by induction on i.

Case $i = 1$: By Definition 7, $Q^{\lambda}_x[0, \psi_1 - \psi', \eta_1 - \eta']$ is finite if and only if $\psi' = \psi_1$, $\eta' = \eta_1$ and $T^{\mathcal{PT}}[v_1, \lambda|_{\mathrm{YW}^{\Gamma}_{v_1}}, \psi', \eta']$ is finite, that is, by induction hypothesis of the lemma, there is a lineage function f' that is eligible for Y_1, λ, $\psi_1 = \psi'$ and $\eta_1 = \eta'$. Thus, the first part of the claim follows. Since ψ_1 and η_1 are the only valid choices for the minima in (8) that result in finite values, we conclude

$$Q^{\lambda}_x[1, \psi_1, \eta_1] = -\rho(x) + T^{\mathcal{PT}}[v_1, \lambda|_{\mathrm{YW}^{\Gamma}_{v_1}}, \psi_1, \eta_1] \overset{\text{IH lemma}}{=} \sum_{z\in Z_1} \mathrm{cost}_{f_i}(z) - \rho(x)$$

since f_i is eligible with respect to Y_1, λ, ψ_1 and η_1 and minimizes $\sum_{z\in Z_1} \mathrm{cost}_{f_i}(z) - \rho(x)$.

Case $i > 1$: First, suppose that $Q^{\lambda}_x[i, \psi_i, \eta_i] \neq \infty$. By (8), there are $\psi' \trianglelefteq \psi_i$ and $\eta' \trianglelefteq \eta_i$ such that $Q^{\lambda}_x[i-1, \psi_{i-1}, \eta_{i-1}] \neq \infty$ and $T^{\mathcal{PT}}[v_i, \lambda|_{\mathrm{YW}^{\Gamma}_{v_i}}, \psi', \eta'] \neq \infty$, where $\psi_{i-1} := \psi_i - \psi'$ and $\eta_{i-1} := \eta_i - \eta'$. By induction hypotheses, there are functions f_{i-1} and f' such that f_{i-1} is eligible with respect to Y_{i-1}, λ, ψ_{i-1}, η_{i-1} and f' is eligible with respect to $\{v_i\}$, λ, ψ', η'. We construct a function f^* by setting

$$f^*(w) := \begin{cases} f'(w) & \text{if } w \in \mathrm{YW}^{\Gamma}_{v_i} \cup \Gamma_{v_i} \\ f_{i-1}(w) & \text{if } w \in \bigcup_{y\in Y_{i-1}} \left(\mathrm{YW}^{\Gamma}_y \cup \Gamma_{v_j}\right) \setminus \mathrm{YW}^{\Gamma}_{v_i} \\ C & \text{otherwise.} \end{cases}$$

(Note that the cost of f on N might be ∞ but we will see that its cost on Z_i is finite). First, we show that f^* is eligible with respect to Y_i, λ, ψ_i, and η_i. To this end, let $w \in \mathrm{YW}^{\Gamma}_y$ for any $y \in Y_i$. Then, by eligibility of f' and f_{i-1} and $\Gamma_{v_j} \cap \Gamma_{v_i} = \varnothing$ for all $j < i$, we have

$$
\begin{aligned}
f^*(w) &= \begin{cases} f'(w) = \lambda\mid_{\mathrm{YW}^{\Gamma}_{v_i}}(w) = \lambda(w) & \text{if } w \in \mathrm{YW}^{\Gamma}_{v_i} \\ f^*(w) = f_{i-1}(w) = \lambda\mid_{\bigcup_{j<i}\mathrm{YW}^{\Gamma}_{v_j}}(w) = \lambda(w) & \text{otherwise} \end{cases} \\
\psi_i(w) &= \psi_{i-1}(w) \cup \psi'(w) \\
&\subseteq \left(f_{i-1}(w) \cap \bigcup_{y \in Y_{i-1}} \bigcup_{u \in \mathrm{Pred}^{\uparrow y}_N(w)} f_{i-1}(u) \right) \cup \left(f'(w) \cap \bigcup_{u \in \mathrm{Pred}^{\uparrow v_i}_N(w)} f'(u) \right) \\
&= f^*(w) \cap \bigcup_{y \in Y_i} \bigcup_{u \in \mathrm{Pred}^{\uparrow y}_N(w)} f^*(w). \\
\eta_i(w) &= \eta_{i-1}(w) + \eta'(w) \leq \sum_{y \in Y_{i-1}} \sum_{u \in \mathrm{Pred}^{\uparrow y}_N(w)} |f_{i-1}(u)| + \sum_{u \in \mathrm{Pred}^{\uparrow v_i}_N(w)} |f'(u)| \\
&= \sum_{y \in Y_i} \sum_{u \in \mathrm{Pred}^{\uparrow y}_N(w)} |f^*(u)|.
\end{aligned}
$$

Finally, the cost of f^* on Z_i equals the cost of f_{i-1} on Z_{i-1} plus the cost of f' on Γ_{v_i} and is, therefore, finite. Thus, f^* is eligible for Y_i, λ, ψ_i and η_i, implying the contraposition of the first part of the lemma. For the cost equality, we consider both directions separately.

"$\leq$": Let $\psi' : \mathrm{YW}^{\Gamma}_{v_i} \to 2^C$ and $\eta' : \mathrm{YW}^{\Gamma}_{v_i} \to \mathbb{N}$ be defined on $\mathrm{YW}^{\Gamma}_{v_i}$ as

$$\psi'(w) := \psi_i(w) \cap f_i(w) \cap \bigcup_{u \in \mathrm{Pred}^{\uparrow v_i}_N(w)} f_i(u)$$

and

$$\eta'(w) := \min\{\eta_i(w), \sum_{u \in \mathrm{Pred}^{\uparrow v_i}_N(w)} |f_i(u)| + \rho(w)\}.$$

Clearly, $\psi' \trianglelefteq \psi_i\mid_{\mathrm{YW}^{\Gamma}_{v_i}}$ and $\eta' \trianglelefteq \eta_i\mid_{\mathrm{YW}^{\Gamma}_{v_i}}$. Furthermore, define ψ_{i-1} and η_{i-1} by, for all $w \in \bigcup_{y \in Y_{i-1}} \mathrm{YW}^{\Gamma}_y$, setting

$$
\begin{aligned}
\psi_{i-1}(w) := \psi_i(w) - \psi'(w) &\subseteq \left(f'(w) \cap \bigcup_{y \in Y_i} \bigcup_{u \in \mathrm{Pred}^{\uparrow y}_N(w)} f'(u) \right) \setminus \left(f'(w) \cap \bigcup_{u \in \mathrm{Pred}^{\uparrow v_i}_N(w)} f'(u) \right) \\
&\subseteq f'(w) \cap \bigcup_{y \in Y_{i-1}} \bigcup_{u \in \mathrm{Pred}^{\uparrow y}_N(w)} f'(u)
\end{aligned}
$$

and

$$
\begin{aligned}
\eta_{i-1}(w) := \eta_i(w) - \eta'(w) &\leq \sum_{y \in Y_i} \sum_{u \in \mathrm{Pred}^{\uparrow y}_N(w)} |f_i(u)| - \sum_{u \in \mathrm{Pred}^{\uparrow v_i}_N(w)} |f_i(u)| \\
&= \sum_{y \in Y_{i-1}} \sum_{u \in \mathrm{Pred}^{\uparrow y}_N(w)} |f_i(u)|.
\end{aligned}
$$

Thus, f_i is eligible with respect to Y_{i-1}, λ, ψ_{i-1} and η_{i-1}, implying

$$\begin{aligned}
Q_x^{\lambda}[i,\psi_i,\eta_i] &\overset{\text{Def. 7}}{\le} Q_x^{\lambda}[i-1,\psi_i-\psi',\eta_i-\eta'] + T^{\mathcal{PT}}[v_i,\lambda\mid_{\mathrm{YW}^{\Gamma}_{v_i}},\psi',\eta'] \\
&\overset{\text{Lem. 9}}{\le} Q_x^{\lambda}[i-1,\psi_{i-1},\eta_{i-1}] + T^{\mathcal{PT}}[v_i,\lambda\mid_{\mathrm{YW}^{\Gamma}_{v_i}},\psi',\eta'] \\
&\overset{\substack{\text{IH claim}\\ \text{IH lemma}}}{\le} \sum_{z\in Z_{i-1}} \mathrm{cost}_{f_i}(z) - \rho(x) + \sum_{z\le v_i} \mathrm{cost}_{f_i}(z) \\
&= \sum_{z\in Z_i} \mathrm{cost}_{f_i}(z) - \rho(x)
\end{aligned}$$

"$\geq$": Let $Q_x^{\lambda}[i,\psi_i,\eta_i]$ be finite as, otherwise, "$\geq$" trivially holds. By (8), there are $\psi' \trianglelefteq \psi_i\mid_{\mathrm{YW}^{\Gamma}_{v_i}}$ and $\eta' \trianglelefteq \eta_i\mid_{\mathrm{YW}^{\Gamma}_{v_i}}$ such that

$$\begin{aligned}
Q_x^{\lambda}[i,\psi_i,\eta_i] = Q_x^{\lambda}[i-1,\psi_i-\psi',\eta_i-\eta'] \\
+ T^{\mathcal{PT}}[v_i,\lambda\mid_{\mathrm{YW}^{\Gamma}_{v_i}},\psi',\eta']
\end{aligned}$$

Towards a contradiction, assume that this value is strictly smaller than $\sum_{z\in Z_i}\mathrm{cost}_{f_i}(z)-\rho(x)$. By the induction hypothesis of the lemma, there is a lineage function f' that is eligible with respect to $\{v_i\}$, λ, ψ', and η' with $T^{\mathcal{PT}}[v_i,\lambda\mid_{\mathrm{YW}^{\Gamma}_{v_i}},\psi',\eta'] = \sum_{z\le_\Gamma v_i}\mathrm{cost}_{f'}(z)$. Further, by the induction hypothesis of the claim, there is a lineage function f_{i-1} that is eligible with respect to Y_{i-1}, λ, $\psi_i-\psi'$, and $\eta_i-\eta'$ with $Q_x^{\lambda}[i-1,\psi_i-\psi',\eta_i-\eta'] = \sum_{z\in Z_{i-1}}\mathrm{cost}_{f_{i-1}}(z)-\rho(x)$. We construct a lineage function f^* by setting

$$f^*(w) := \begin{cases} f_{i-1}(w) & \text{if } w\in Z_{i-1} \\ f'(w) & \text{if } w\in \Gamma_{v_i} \\ f_i(w) & \text{otherwise} \end{cases}$$

By eligibility of f_{i-1}, f_i and f', we know that f_{i-1}, f_i and f^* coincide with λ on $\bigcup_{y\in Y_{i-1}}\mathrm{YW}^{\Gamma}_y$ and f', f_i and f^* coincide with λ on $\mathrm{YW}^{\Gamma}_{v_i}$. To contradict optimality of f, it thus suffices to show that f^* is eligible with respect to Y_i, λ, ψ_i, and η_i. To this end,note that, for all $w\in\bigcup_{y\in Y_i}\mathrm{YW}^{\Gamma}_y$, we have

$$\begin{aligned}
\psi_i(w) &= (\psi_i-\psi')(w)\cup\psi'(w) \\
&\subseteq \left(\lambda(w)\cap\bigcup_{y\in Y_{i-1}}\bigcup_{u\in \mathrm{Pred}_N^{\uparrow y}(w)} f_{i-1}(u)\right)\cup\left(\lambda(w)\cap\bigcup_{u\in \mathrm{Pred}_N^{\uparrow v_i}(w)} f'(u)\right) \\
&= f^*(w)\cap\bigcup_{y\in Y_i}\bigcup_{u\in \mathrm{Pred}_N^{\uparrow y}(w)} f^*(u)
\end{aligned}$$

as well as

$$\begin{aligned}
\eta_i(w) = (\eta_i-\eta')(w)+\eta'(w) &\le \sum_{y\in Y_{i-1}}\sum_{u\in \mathrm{Pred}_N^{\uparrow y}(w)} |f_{i-1}(u)| + \sum_{u\in \mathrm{Pred}_N^{\uparrow v_i}(w)} |f'(u)| + \rho(w) \\
&= \sum_{y\in Y_i}\sum_{u\in \mathrm{Pred}_N^{\uparrow y}(w)} |f^*(u)| + \rho(w)
\end{aligned}$$

$\square$

Having established the equality for Q_x^λ, we can now prove the lemma for $i > 1$. For the first part, suppose that $T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x] \neq \infty$. By (7), there are $D \subseteq U \subseteq \phi(x)$ such that $T^{\mathcal{PT}}[x, \lambda_x, \psi_x, \eta_x] = Q_x^{\lambda_x x \to U}[t, \psi_t, \eta_t] + |U \setminus (D \cup \bigcup_{u \in \mathrm{Pred}_N^{\downarrow}(x)} \lambda_x(u))|$, where

$$\psi_t := \psi_x\left[x \to D, \forall_{w \in \mathrm{Succ}_N^{\uparrow}(x)} w \to \psi_x(w) \setminus U\right] \text{ and}$$

$$\eta_t := \eta_x\left[x \to |U| \dot{-} \sum_{u \in \mathrm{Pred}_N^{\downarrow}(x)} |\lambda_x(u)|, \forall_{w \in \mathrm{Succ}_N^{\uparrow}(x)} w \to \eta_x(w) \dot{-} |U|\right].$$

By Claim 8, there is a lineage function f_t that is eligible for $\{v_t\}$, $\lambda_t := \lambda_x[x \to U]$, ψ_t, and η_t. Without loss of generality, suppose that $f_t(w) = \lambda_t(w)$ for all $w \in (\mathrm{YW}_x^\Gamma \cup \{x\}) \setminus \bigcup_{y \in Y_t} \mathrm{YW}_y^\Gamma$. In particular, $f_t(x) = \lambda_t(x) = U$ and f_t has finite cost on Z_t.

$$\begin{aligned} |f_t(x)| &> \sum_{u \in \mathrm{Pred}_N(x)} |f_t(u)| \\ &= \sum_{u \in \mathrm{Pred}_N^{\downarrow}(x)} |f_t(u)| + \sum_{y \in Y_t} \sum_{u \in \mathrm{Pred}_N^{\uparrow y}(x)} |f_t(u)| \\ &\geq \sum_{u \in \mathrm{Pred}_N^{\downarrow}(x)} |\lambda_x(u)| + n_t(x) \geq |U| \end{aligned}$$

contradicting $f_t(x) = U$. Further, for each $w \in \mathrm{YW}_x^\Gamma \setminus \mathrm{Succ}_N^{\uparrow}(x)$,

$$\psi_x(w) = \psi_t(w) \subseteq f_t(w) \cap \bigcup_{y \in Y_t} \bigcup_{u \in \mathrm{Pred}_N^{\uparrow y}(w)} f_t(u) = f_t(w) \cap \bigcup_{u \in \mathrm{Pred}_N^{\uparrow x}(w)} f_t(u)$$

$$\eta_x(w) = \eta_t(w) \leq \sum_{y \in Y_y} \sum_{u \in \mathrm{Pred}_N^{\uparrow y}(w)} |f_t(u)| = \sum_{u \in \mathrm{Pred}_N^{\uparrow x}(w)} |f_t(u)|$$

and, for each $w \in \mathrm{Succ}_N^{\uparrow}(x)$,

$$\psi_x(w) \subseteq \psi_t(w) \cup U \subseteq f_t(w) \cap \bigcup_{y \in Y_t} \bigcup_{u \in \mathrm{Pred}_N^{\uparrow y}(w)} f_t(u) \cup f_t(x) = f_t(w) \cap \bigcup_{u \in \mathrm{Pred}_N^{\uparrow x}(w)} f_t(u)$$

$$\eta_x(w) \leq \eta_t(w) + |U| \leq \sum_{y \in Y_y} \sum_{u \in \mathrm{Pred}_N^{\uparrow y}(w)} |f_t(u)| + |f_t(x)| = \sum_{u \in \mathrm{Pred}_N^{\uparrow x}(w)} |f_t(u)|$$

We show that f_t is eligible with respect to $\{x\}$, λ_x, ψ_x and η_x. First, assume that $\mathrm{cost}_{f_t}(x) = \infty$, that is, either $x = \rho_N$ and $|f_t(x)| = |U| > 1$ or $x \neq \rho_N$ and $|f_t(x)| > \sum_{u \in \mathrm{Pred}_N(x)} |f_t(u)|$. In the first case, $n_t(x) = |U| > 1$, contradicting $\eta_t(x) \leq \rho(x)$. In the second case, $n_t(x) = |U| \dot{-} \sum_{u \in \mathrm{Pred}_N^{\downarrow}(x)} |\lambda_x(u)|$, implying

Thus, f_t is eligible with respect to $\{x\}$, λ_x, ψ_x and η_x, implying the first part of the lemma. For the second part, we consider the directions seperately.

"≥": We pick up the definition of f_t and show that $T^{\mathcal{PT}}[x,\lambda_x,\psi_x,\eta_x] \geq \sum_{z\in\Gamma_x}\text{cost}_{f_t}(z)$. Then, "≥" follows from optimality of f on Γ_x. Indeed,

$$T^{\mathcal{PT}}[x,\lambda_x,\psi_x,\eta_x] = Q_x^{\lambda_x x\to U}[t,\psi_t,\eta_t] + |U \setminus (D \cup \bigcup_{u\in\text{Pred}_N^{\downarrow}(x)} \lambda_x(u))|$$

and, since $\psi_t(x) = D \subseteq f_t(x) \cap \bigcup_{u\in\text{Pred}_N^{\uparrow x}(x)} f_t(u)$,

$$\geq \sum_{z\in Z_t}\text{cost}_{f_t}(z) + |f_t(x) \setminus (\bigcup_{u\in\text{Pred}_N^{\uparrow}(x)} f_t(u) \cup \bigcup_{u\in\text{Pred}_N^{\downarrow}(x)} f_t(u))|$$
$$= \sum_{z\in Z_t}\text{cost}_{f_t}(z) + \text{cost}_{f_t}(x) = \sum_{z\in\Gamma_x}\text{cost}_{f_t}(z)$$

"≤": Let $U := f(x)$ and let $D := f(x) \cap \bigcup_{u\in\text{Pred}_N^{\uparrow x}(x)} f(u) \subseteq U$. Then, $|U \setminus (D \cup \bigcup_{u\in\text{Pred}_N^{\downarrow}(x)} f(u)\lambda_x(u))| = \text{cost}_f(x) + \rho(x)$. Further, let

$$\psi_t := \psi_x\left[x \to D, \forall_{w\in\text{Succ}_N^{\uparrow}(x)} w \to \psi_x(w) \setminus U\right] \text{ and}$$
$$\eta_t := \eta_x\left[x \to |U| \dot{-} \sum_{u\in\text{Pred}_N^{\downarrow}(x)} |\lambda_x(u)|, \forall_{w\in\text{Succ}_N^{\uparrow}(x)} w \to \eta_x(w) \dot{-} |U|\right].$$

We show that f is eligible with respect to Y_t, λ_x, ψ_t and η_t. Then, $Q_x^{\lambda_x x\to D}[t,\psi_t,\eta_t] = \sum_{z\in Z_t}\text{cost}_f(z) - \rho(x)$ by Claim 8, implying $T^{\mathcal{PT}}[x,\lambda_x,\psi_x,\eta_x] \leq \sum_{z\in Z_t}\text{cost}_f(z) - \rho(x) + \text{cost}_f(x) + \rho(x) = \sum_{z\in\Gamma_x}\text{cost}_f(z)$ since U and D are valid choices for the minimum in (7).

To see that f is eligible, note that $f(w) = \lambda_x[x \to U]$ for all $w \in \bigcup_{y\in Y_t} \text{YW}_y^\Gamma$ since $\bigcup_{y\in Y_t}\text{YW}_y^\Gamma \subseteq \text{YW}_x^\Gamma \cup \{x\}$. Further, for the conditions on ψ_t and η_t, consider three cases for nodes in $\bigcup_{y\in Y_t}\text{YW}_y^\Gamma$. First, if $w = x$, then

$$\psi_t(x) = D = f(x) \cap \bigcup_{u\in\text{Pred}_N^{\uparrow x}(x)} f(u) = f(x) \cap \bigcup_{y\in Y_t} \bigcup_{u\in\text{Pred}_N^{\uparrow y}(x)} f(u)$$
$$\eta_t(x) = |U| \dot{-} \sum_{u\in\text{Pred}_N^{\downarrow}(x)} |\lambda_x(u)| = |f(x)| \dot{-} \sum_{u\in\text{Pred}_N^{\downarrow}(x)} |f(u)| \overset{\text{Def. 4}}{\leq} \sum_{u\in\text{Pred}_N^{\uparrow}(x)} |f(u)|$$
$$= \sum_{y\in Y_t} \sum_{u\in\text{Pred}_N^{\uparrow}(x)} |f(u)|$$

Second, if $w \in \bigcup_{y\in Y_t}\text{YW}_y^\Gamma \cap \text{Succ}_N^{\uparrow}(x)$, then

$$\psi_t(w) = \psi_x(w) \setminus U \subseteq f(w) \cap \bigcup_{u\in\text{Pred}_N^{\uparrow x}(w)} f(u) \setminus f(x) = f(w) \cap \bigcup_{u\in\text{Pred}_N^{\uparrow x}(w)\setminus\{x\}} f(u)$$
$$= f(w) \cap \bigcup_{y\in Y_t} \bigcup_{u\in\text{Pred}_N^{\uparrow y}(w)} f(u)$$

as well as

$$\eta_t(w) = \eta_x(w) \dot{-} |U| \leq \sum_{u\in\text{Pred}_N^{\uparrow x}(w)} |f(u)| + \rho(x) \dot{-} |f(x)|$$
$$= \sum_{u\in\text{Pred}_N^{\uparrow x}(w)\setminus\{x\}} |f(u)| + |f(x)| + \rho(x) \dot{-} |f(x)| = \sum_{y\in Y_t} \sum_{u\in\text{Pred}_N^{\uparrow y}(w)} |f(u)| + \rho(x)$$

Otherwise, $w \in \bigcup_{y \in Y_t} \mathrm{YW}_y^{\Gamma} \setminus (\mathrm{Succ}_N^{\uparrow}(x) \cup \{x\})$ and we have

$$\psi_t(w) = \psi_x(w) \subseteq f(w) \cap \bigcup_{u \in \mathrm{Pred}_N^{\uparrow x}(w)} f(u) = f(w) \cap \bigcup_{y \in Y_t} \bigcup_{u \in \mathrm{Pred}_N^{\uparrow x}(y)} f(u)$$

$$\eta_t(w) = \eta_x(w) \leq \sum_{u \in \mathrm{Pred}_N^{\uparrow x}(w)} |f(u)| + \rho(x) = \sum_{y \in Y_t} \sum_{u \in \mathrm{Pred}_N^{\uparrow y}(w)} |f(u)| + \rho(x)$$

□

Acknowledgements
We thank Christophe Paul for sharing his expertise on treewidth formulations, and an anonymous reviewer for suggesting an interesting variant of decomposition minimizing the maximum number of reticulation nodes per bag for future work.

Author contributions
CS and MW contributed equally to the paper. Both authors read and approved the final manuscript.

Funding
This work was supported by French Agence Nationale de la Recherche through the CoCoAlSeq Project (ANR-19-CE45-0012).

Declarations

Competing interests
The authors declare that they have no competing interests

Author details
[1]ISEM, Université de Montpellier, CNRS, IRD, EPHE, Montpellier, France. [2]LIGM, Université Gustave Eiffel, CNRS, Paris, France.

Received: 24 January 2022 Accepted: 17 March 2022

References

1. Felsenstein J. Inferring phylogenies, vol. 2. Sunderland: Sinauer Associates; 2004.
2. Fitch WM. Toward defining the course of evolution: minimum change for a specific tree topology. Syst Biol. 1971;20(4):406–16.
3. Huson DH, Rupp R, Scornavacca C. Phylogenetic networks: concepts. Algorithms and applications. Cambridge: Cambridge University Press; 2010.
4. Kannan L, Wheeler WC. Maximum parsimony on phylogenetic networks. Algo Mol Biol. 2012;7(1):9.
5. Hein J. Reconstructing evolution of sequences subject to recombination using parsimony. Math Biosci. 1990;98(2):185–200.
6. Nakhleh L, Jin G, Zhao F, Mellor-Crummey J. Reconstructing phylogenetic networks using maximum parsimony. In: 2005 IEEE Computational Systems Bioinformatics Conference (CSB'05), pp. 93–102 (2005). IEEE
7. Zhu J, Yu Y, Nakhleh L. In the light of deep coalescence: revisiting trees within networks. BMC Bioinformat. 2016;17(14):271–82.
8. Fischer M, Iersel LV, Kelk S, Scornavacca C. On computing the maximum parsimony score of a phylogenetic network. SIAM J Discret Math. 2015;29(1):559–85.
9. Kannan L, Wheeler WC. Exactly computing the parsimony scores on phylogenetic networks using dynamic programming. J Comput Biol. 2014;21(4):303–19.
10. Jin G, Nakhleh L, Snir S, Tuller T. Parsimony score of phylogenetic networks: hardness results and a linear-time heuristic. IEEE/ACM Trans Comput Biol Bioinf. 2009;6(3):495–505.
11. Jin G, Nakhleh L, Snir S, Tuller T. Maximum likelihood of phylogenetic networks. Bioinformatics. 2006;22(21):2604–11.
12. Van Iersel L, Jones M, Scornavacca C. Improved maximum parsimony models for phylogenetic networks. Syst Biol. 2018;67(3):518–42.
13. Bodlaender HL. A linear-time algorithm for finding tree-decompositions of small treewidth. SIAM J Comput. 1996;25(6):1305–17.
14. Authors V. The graph parameter hierarchy. Available at https://gitlab.com/gruenwald/parameter-hierarchy. 2021.
15. Bodlaender HL. Discovering treewidth. In: Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM'05), pp. 1–16. Springer, Berlin, Heidelberg. 2005.
16. Bodlaender HL. Treewidth: structure and algorithms. In: International Colloquium on Structural Information and Communication Complexity, pp. 11–25. Springer. 2007.
17. Bryant D, Lagergren J. Compatibility of unrooted phylogenetic trees is FPT. Theoret Comput Sci. 2006;351(3):296–302.
18. Courcelle B. The monadic second-order logic of graphs. i. recognizable sets of finite graphs. Inf Comput. 1990;85(1):12–75.
19. Bulteau L, Weller M. Parameterized algorithms in bioinformatics: an overview. Algorithms. 2019;12(12):256.
20. Berry V, Scornavacca C, Weller M. Scanning phylogenetic networks is NP-hard. In: Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM'20), pp. 519–530, Springer, 2020.
21. Korhonen T. Single-exponential time 2-approximation algorithm for treewidth. CoRR abs/2104.07463. 2021.
22. Dell H, Komusiewicz C, Talmon N, Weller M. The PACE 2017 Parameterized Algorithms and Computational Experiments Challenge: The Second Iteration. In: 12th International Symposium on Parameterized and Exact Computation (IPEC 2017), vol. 89. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik; 2018. p. 30–13012.
23. Tamaki H. Positive-instance driven dynamic programming for treewidth. J Comb Optim. 2019;37(4):1283–311.
24. Dendris ND, Kirousis LM, Thilikos DM. Fugitive-search games on graphs and related parameters. Theoret Comput Sci. 1997;172(1):233–54.
25. Arnborg S. Efficient algorithms for combinatorial problems on graphs with bounded decomposability—a survey. BIT Numer Math. 1985;25(1):1–23.
26. Mescoff G, Paul C, Thilikos D. A polynomial time algorithm to compute the connected tree-width of a series-parallel graph. 2021. 2004.00547v5.
27. Kloks T. Treewidth: computations and approximations, vol. 842. Berlin: Springer; 1994.
28. Rabier C-E, Berry V, Stoltz M, Santos JaD, Wang W, Jean-Christophe G. Pardi F, Scornavacca C. On the inference of complicated phylogenetic networks by Markov Chain Monte-Carlo. Submitted.
29. Bachoore E, Bodlaender HL. Weighted treewidth algorithmic techniques and results. In: International Symposium on Algorithms and Computation, pp. 893–903. Springer; 2007.
30. Zhu J, Wen D, Yu Y, Meudt HM, Nakhleh L. Bayesian inference of phylogenetic networks from bi-allelic genetic markers. PLoS Comput Biol. 2018;14(1):1005932.
31. Kelk S, Pardi F, Scornavacca C, van Iersel L. Finding a most parsimonious or likely tree in a network with respect to an alignment. J Math Biol. 2019;78(1–2):527–47.

32. Jin G, Nakhleh L, Snir S, Tuller T. Inferring phylogenetic networks by the maximum parsimony criterion: a case study. Mol Biol Evol. 2006;24(1):324–37.
33. Wheeler WC. Phylogenetic network analysis as a parsimony optimization problem. BMC Bioinformatics. 2015;16(1):1–9.
34. Bryant C, Fischer M, Linz S, Semple C. On the quirks of maximum parsimony and likelihood on phylogenetic networks. J Theor Biol. 2017;417:100–8.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC