1-2017

Discriminative Measures for Comparison of Phylogenetic Trees

Omur Arslan
University of Pennsylvania, omur@seas.upenn.edu

Dan P. Guralnik
University of Pennsylvania, guralnik@seas.upenn.edu

Daniel E. Koditschek
University of Pennsylvania, kod@seas.upenn.edu

Recommended Citation

Omur Arslan, Dan P. Guralnik, and Daniel E. Koditschek, "Discriminative Measures for Comparison of Phylogenetic Trees", *Discrete Applied Mathematics* Volume 217, Part 3, 405-426. January 2017. http://dx.doi.org/10.1016/j.dam.2016.10.003

Discriminative Measures for Comparison of Phylogenetic Trees

Abstract

In this paper we introduce and study three new measures for efficient discriminative comparison of phylogenetic trees. The NNI navigation dissimilarity d_{nav} counts the steps along a “combing” of the Nearest Neighbor Interchange (NNI) graph of binary hierarchies, providing an efficient approximation to the (NP-hard) NNI distance in terms of “edit length”. At the same time, a closed form formula for d_{nav} presents it as a weighted count of pairwise incompatibilities between clusters, lending it the character of an edge dissimilarity measure as well. A relaxation of this formula to a simple count yields another measure on all trees — the crossing dissimilarity d_{CM}. Both dissimilarities are symmetric and positive definite (vanish only between identical trees) on binary hierarchies but they fail to satisfy the triangle inequality. Nevertheless, both are bounded below by the widely used Robinson–Foulds metric and bounded above by a closely related true metric, the cluster-cardinality metric d_{CC}. We show that each of the three proposed new dissimilarities is computable in time O(n^2) in the number of leaves n, and conclude the paper with a brief numerical exploration of the distribution over tree space of these dissimilarities in comparison with the Robinson–Foulds metric and the more recently introduced matching-split distance.

Keywords

Phylogenetic trees; Evolutionary trees; Nearest Neighbor Interchange; Comparison of classifications; Tree metric

Disciplines

Applied Mathematics | Bioinformatics | Discrete Mathematics and Combinatorics | Electrical and Computer Engineering | Engineering | Robotics | Systems Engineering

Available online at www.sciencedirect.com

ScienceDirect

Discrete Applied Mathematics 00 (2017) 1–25

DISCRETE APPLIED MATHEMATICS

Discriminative Measures for Comparison of Phylogenetic Trees

Omur Arslan*, Dan P. Guralnik, Daniel E. Koditschek

Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract

In this paper we introduce and study three new measures for efficient discriminative comparison of phylogenetic trees. The *NNI navigation dissimilarity* d_{nav} counts the steps along a "combing" of the Nearest Neighbor Interchange (NNI) graph of binary hierarchies, providing an efficient approximation to the (NP-hard) NNI distance in terms of "edit length". At the same time, a closed form formula for d_{nav} presents it as a weighted count of pairwise incompatibilities between clusters, lending it the character of an edge dissimilarity measure as well. A relaxation of this formula to a simple count yields another measure on *all* trees — the *crossing dissimilarity* d_{CM}. Both dissimilarities are symmetric and positive definite (vanish only between identical trees) on binary hierarchies but they fail to satisfy the triangle inequality. Nevertheless, both are bounded below by the widely used Robinson-Foulds metric and bounded above by a closely related true metric, the *cluster-cardinality metric* d_{CC}. We show that each of the three proposed new dissimilarities is computable in time $O(n^2)$ in the number of leaves n, and conclude the paper with a brief numerical exploration of the distribution over tree space of these dissimilarities in comparison with the Robinson-Foulds metric and the more recently introduced matching-split distance.

Keywords:
Phylogenetic Trees, Evolutionary Trees, Nearest Neighbor Interchange, Comparison of Classifications, Tree Metric.

1. Introduction

1.1. Motivation

A fundamental classification problem common to both computational biology and engineering is the efficient and informative comparison of hierarchical structures. In bioinformatics settings, these typically take the form of phylogenetic trees representing evolutionary relationships within a set S of taxa. In pattern recognition and data mining settings, hierarchical trees are often used to encode nested sequences of groupings of a set of observations. Dissimilarity between combinatorial trees has been measured in the past literature largely by recourse to one of two separate approaches: comparing edges and counting edit distances. Representing the former approach, a widely used tree metric is the Robinson-Foulds (RF) distance, d_{RF}, [1] whose count of the disparate edges between trees requires linear time, $O(n)$, in the number of leaves, n, to compute [2]. Empirically, d_{RF} offers only a very coarse measure of disparity, and among its many proposed refinements, the recent matching split distance d_{MS}, [3, 4] offers a more discriminative metric albeit with considerably higher computational cost, $O(n^{2.5} \log n)$. Alternatively, various edit distances have been proposed [5–8] but the most natural variant, the Nearest Neighbor Interchange (NNI) distance d_{NNI}, entails an NP-complete computation for both labelled and unlabelled trees [9].

*Corresponding author.
Email addresses: omur@seas.upenn.edu (Omur Arslan), guralnik@seas.upenn.edu (Dan P. Guralnik), kod@seas.upenn.edu (Daniel E. Koditschek)

1.2. Results

Our main contribution is the introduction of a dissimilarity measure on the space $\mathcal{BT}_S$ of labelled binary trees which bridges the above approaches by what is, effectively, a solution to the NNI navigation problem in $\mathcal{BT}_S$:

Problem 1 (NNI Navigation Problem)**.** *Given a target $\tau \in \mathcal{BT}_S$, provide an efficient algorithm $\mathcal{A}_\tau$ which, for any $\sigma \in \mathcal{BT}_S$, computes a Nearest Neighbor Interchange to be performed on σ while guaranteeing that successive application of $\mathcal{A}_\tau$ terminates in τ.*

This problem is motivated by applications in coordinated robot navigation [10–13], where a group of robots is required to reconfigure reactively in real time their (structural) adjacencies while navigating towards a desired goal configuration. Thus, our particular formulation of the problem is inspired by the notion of reactive planning [14], but may likely hold value for researchers interested in tree consensus and averaging as well.

Of course, since computation of d_{NNI} is NP-hard, one cannot hope for repeated applications of $\mathcal{A}_\tau$ to produce NNI geodesics without incurring prohibitive complexity in each iteration. However, as we will show, constructing an efficient navigation scheme is possible if we allow the algorithm to produce less restricted paths: for $|S| = n$, our navigation algorithms require $O(n)$ time for each iteration and produce paths of length $O(n^2)$ (as compared to the $O(n \log n)$ diameter of d_{NNI} — see (19)).

Additional insight into the geometry of the space $(\mathcal{BT}_S, d_{NNI})$ is gained by recognizing a significant degree of freedom with which our navigation algorithm may select the required tree restructuring operation at each stage. As it turns out, for any given target τ, the repeated application of $\mathcal{A}_\tau$ to a tree σ until reaching τ will yield paths of equal lengths regardless of any choices made along the way. This length, by definition, is the navigation dissimilarity $d_{nav}(\sigma, \tau)$ (and is obtained, in the manner described, in $O(n^3)$ time, though more efficient implementations will guarantee $O(n^2)$). At the same time, a closed form formula we derive for d_{nav} allows us to avoid computing a navigation path when only the value of d_{nav} is needed, and computes it in $O(n^2)$ time. Surprisingly, despite the asymmetric character of its construction, d_{nav} is a symmetric (and positive definite) dissimilarity on $\mathcal{BT}_S$, though it fails to be a metric.

Although d_{nav} does not satisfy the triangle inequality, it is related to the well accepted Robinson-Foulds distance by the following tight bounds:

$$d_{RF} \leq d_{nav} \leq \frac{1}{2} d_{RF}^2 + \frac{1}{2} d_{RF}\ , \tag{1}$$

We find it useful to introduce a "relaxation" of d_{nav}, the *crossing dissimilarity* d_{CM}. This dissimilarity simply counts all the pairwise cluster incompatibilities between two trees, hence it is symmetric, positive-definite, and computable in $O(n^2)$ time. In fact, the two dissimilarities are commensurable, leading to similar bounds in terms of d_{RF}:

$$d_{RF} \leq d_{nav} \leq \frac{3}{2} d_{CM}\ , \quad d_{RF} \leq d_{CM} \leq d_{RF}^2\ . \tag{2}$$

Finally, we introduce a true metric whose spatial resolution and computational complexity is comparable to those our new dissimilarities. Exploiting a well known relation between trees and ultrametrics [15], we also introduce *the cluster-cardinality distance* d_{CC} — constructed as the pullback of a matrix norm along an embedding of hierarchies into the space of matrices and computable in $O(n^2)$ time — which is a true metric bounding d_{CM} from above (and hence also d_{nav}, up to a constant factor). Thus, cumulatively we obtain:

$$\frac{2}{3} d_{RF} \leq \frac{2}{3} d_{nav} \leq d_{CM} \leq d_{CC}\ . \tag{3}$$

We have surveyed some of the new features of our tree proximity measures that might hold interest for pattern classification and phylogeny analysis relative to the diverse alternatives that have appeared in the literature. Closest among these many alternatives [16–18], d_{nav} has some resemblance to an early NNI graph navigation algorithm, d_{ra} [18] which used a divide-and-conquer approach with a balancing strategy to achieve an $O(n \log n)$ computation of tree dissimilarity. Notwithstanding its lower computational cost, in contrast to d_{nav}, the recursive definition of d_{ra}, as with many NNI distance approximations [16–18], does not admit a closed form expression.

It is often of interest to compare more than pairs of hierarchies at a time, and the notion of a "consensus" tree has accordingly claimed a good deal of attention in the literature [19]. For instance, the majority rule tree [20] of a set of trees is a median tree respecting the RF distance and provides statistics on the central tendency of trees [21]. When d_{nav} and d_{CM} are extended to degenerate trees they fail to be positive definite, and thus their behavior over (typically degenerate) consensus trees departs still further from the properties of a true metric. However, it turns out that both notions of a consensus tree (strict [22], and loose/semi-strict [23]) behave as median trees with respect to both our dissimilarities. In fact, the loose consensus tree is the maximal (finest) median tree with respect to inclusion for both d_{nav} and d_{CM}.

The paper is organized as follows. Section 2 briefly summarizes the necessary background while introducing the notation used throughout the sequel. Section 3 introduces and studies the cluster-cardinality distance d_{CC} and the crossing dissimilarity d_{CM}. In Section 4 we present a solution of the NNI navigation problem and study properties of the resulting NNI navigation dissimilarity d_{nav} and its relations with other tree dissimilarity measures. Section 5 discusses the relation between commonly used consensus models and our tree dissimilarities d_{CM} and d_{nav}, and compares our proposed tree measures with d_{RF} and d_{MS} based on some frequently used empirical distributions of tree measures. A brief discussion of future directions follows in Section 6.

2. Preliminaries

2.1. Hierarchies

By a *hierarchy* τ over a fixed non-empty finite index set S we shall mean a rooted tree with labeled leaves (see Figure 1). Formally, τ is a finite connected acyclic graph with leaves (vertices of degree one) bijectively labelled by S, and edges oriented in such a way that (i) all interior vertices have out-degree at least two, and (ii) there is a vertex, referred to as the *root of* τ, such that every edge is oriented away from the root. Under these assumptions all the vertices of τ are reachable from the root through a directed path in τ [24].

The *cluster* $\mathcal{C}(v)$ of a vertex $v \in V_\tau$ of a hierarchy τ is defined to be the set of leaves reachable from v by a directed path in τ. Singleton clusters and the root cluster S are common to all trees, and we refer to them as the trivial clusters. We denote by $\mathcal{C}(\tau)$ (respectively $\mathcal{C}_{int}(\tau)$) the set of all clusters (resp. non-trivial clusters) of τ:

$$\mathcal{C}(\tau) := \left\{\mathcal{C}(v) \,\middle|\, v \in V_\tau\right\} \subseteq \mathcal{P}(S)\ , \qquad \mathcal{C}_{int}(\tau) := \left\{I \in \mathcal{C}(\tau) \setminus \{S\} \,\middle|\, |I| \geq 2\right\}\ , \tag{4}$$

where $\mathcal{P}(S)$ denotes the power set of S.

2.1.1. Compatibility

Definition 1 ([8, 25])**.** *Subsets $A, B \subset S$ are said to be* compatible, *$A \bowtie B$, if*

$$A \cap B = \varnothing \ \vee\ A \subseteq B \ \vee\ B \subseteq A\ . \tag{5}$$

If $A \not\bowtie B$, then we say that A and *B* cross. *We further extend the compatibility relation ($\bowtie$) as follows:*

- *For $\mathcal{A}, \mathcal{B} \subseteq \mathcal{P}(S)$, write $\mathcal{A} \bowtie \mathcal{B}$ if $A \bowtie B$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$;*
- *For a cluster $I \subseteq S$ and a tree τ over the leaf set S, write $I \bowtie \tau$ if $\{I\} \bowtie \mathcal{C}(\tau)$;*
- *For two trees σ and τ over the leaf set S, write $\sigma \bowtie \tau$ if $\mathcal{C}(\sigma) \bowtie \mathcal{C}(\tau)$.*

By construction, any two elements of $\mathcal{C}(\tau)$ are compatible for any tree τ. This motivates the following definition:

Definition 2 ([25])**.** *A subset $\mathcal{A}$ of $\mathcal{P}(S)$ is said to be* nested *— also referred to in the literature as a "laminar family" — if any two elements of $\mathcal{A}$ are compatible. $\mathcal{C}(\tau)$ is known as the laminar family associated with τ.*

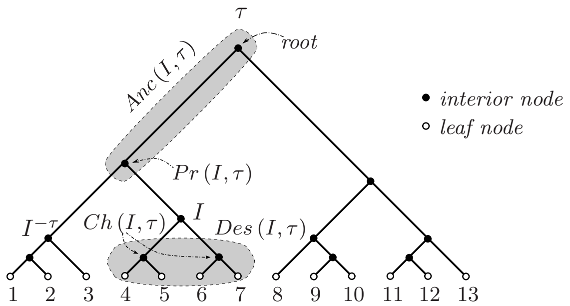

Figure 1. Hierarchical Relations: ancestors - $\mathrm{Anc}(I,\tau)$, parent - $\mathrm{Pr}(I,\tau)$, children - $\mathrm{Ch}(I,\tau)$, descendants - $\mathrm{Des}(I,\tau)$, and local complement (sibling) - $I^{-\tau}$ of cluster I of a rooted binary phylogenetic tree, $\tau \in \mathcal{BT}_{[13]}$. Filled and unfilled circles represent interior and leaf nodes, respectively. An interior node is referred to by its cluster, the list of leaves below it; for example, $I = \{4,5,6,7\}$. Accordingly, the cluster set of τ is $\mathcal{C}(\tau) = \{\{1\},\{2\},\ldots,\{13\},\{1,2\},\{1,2,3\},\{4,5\},\{6,7\},\{4,5,6,7\},\{1,2,\ldots,7\},\{9,10\},\{8,9,10\},\{11,12\},\{11,12,13\},\{8,9,\ldots,13\},\{1,2,\ldots,13\}\}$.

2.1.2. Hierarchical Relations

The cluster set $\mathcal{C}(\tau)$ of a hierarchy τ completely determines its representation as a rooted tree with labeled leaves: $\mathcal{C}(\tau)$ stands in bijective correspondence with the vertex set of τ, and (v, v') is an edge in τ if and only if $\mathcal{C}(v) \supset \mathcal{C}(v')$ and there is no $\tilde{v} \in V_\tau$ such that $\mathcal{C}(v) \supset \mathcal{C}(\tilde{v}) \supset \mathcal{C}(v')$. Consequently, the standard notions of ancestor, descendant, parent and child of a vertex in common use for rooted trees carry over to the cluster representation as follows:

$$\mathrm{Anc}(I,\tau) = \left\{V \in \mathcal{C}(\tau) \,\middle|\, I \subsetneq V\right\}, \qquad \mathrm{Des}(I,\tau) = \left\{V \in \mathcal{C}(\tau) \,\middle|\, V \subsetneq I\right\}, \tag{6a}$$

$$\mathrm{Pr}(I,\tau) = \min\left(\mathrm{Anc}(I,\tau)\right), \qquad \mathrm{Ch}(I,\tau) = \left\{V \in \mathcal{C}(\tau) \,\middle|\, \mathrm{Pr}(V,\tau) = I\right\}, \tag{6b}$$

where $\min(\mathrm{Anc}(I,\tau))$ is computed with respect to the inclusion order. Note that for the trivial clusters we have $\mathrm{Pr}(S,\tau) = \varnothing$ and $\mathrm{Ch}(\{s\},\tau) = \varnothing$ for $s \in S$.

Since the set of children partitions each parent, we find it useful to define the *local complement* $I^{-\tau}$ of $I \in \mathcal{C}(\tau)$ as

$$I^{-\tau} := \mathrm{Pr}(I,\tau) \setminus I\,, \tag{7}$$

not to be confused with the standard (global) complement, $I^C = S \setminus I$. Further, a grandchild in τ is a cluster $G \in \mathcal{C}(\tau)$ having a grandparent $\mathrm{Pr}^2(G,\tau) := \mathrm{Pr}(\mathrm{Pr}(G,\tau),\tau)$ in τ. We denote the set of all grandchildren in τ by $\mathcal{G}(\tau)$,

$$\mathcal{G}(\tau) := \left\{G \in \mathcal{C}(\tau) \,\middle|\, \mathrm{Pr}^2(G,\tau) \neq \varnothing\right\}\,. \tag{8}$$

If A, B are either elements of S or clusters of τ, it is convenient to have $(A \wedge B)_\tau$ denote the smallest (in terms of cardinality) common ancestor of A and B in τ. Finally, the depth $\ell_\tau(I)$ of a cluster in a hierarchy τ is defined to equal the number of distinct ancestors of I in τ.

2.1.3. Nondegeneracy

A rooted tree where every interior vertex has exactly two children is said to be *binary* or *non-degenerate*. All other trees are said to be *degenerate*. We will denote the set of hierarchies over a finite leaf set S, by $\mathcal{T}_S$. The subset of non-degenerate hierarchies will be denoted by $\mathcal{BT}_S$.

Note that the laminar family $\mathcal{C}(\tau)$ of a degenerate tree τ may always be augmented with additional clusters while remaining nested (Definition 2). This leads to the well known result:

Remark 1 ([25, 26]). *Let $\tau \in \mathcal{T}_S$. Then τ has at most $2|S|-1$ vertices, with equality if and only if τ is nondegenerate, if and only if $\mathcal{C}(\tau)$ is a maximal laminar family in $\mathcal{P}(S)$ with respect to inclusion.*[1]

[1]In this paper we adopt the convention that a laminar family does not contain the empty set (as an element).

2.1.4. Consensus

Definition 3 ([22, 23]). *For any set of trees T in $\mathcal{T}_S$, the strict and loose consensus trees of T, denoted T_* and T^* respectively, are defined by specifying their cluster sets as follows:*

$$\mathcal{C}(T_*) = \bigcap_{\tau \in T} \mathcal{C}(\tau), \quad \mathcal{C}(T^*) = \left\{ I \in \bigcup_{\tau \in T} \mathcal{C}(\tau) \,\middle|\, \forall \sigma \in T \quad I \bowtie \sigma \right\}. \tag{9}$$

Note that the loose consensus tree T^* of T refines the strict consensus tree T_*, that is $\mathcal{C}(T^*) \supseteq \mathcal{C}(T_*)$.

2.2. Some Operations on Trees

2.2.1. The NNI Graph

The standard definition of NNI walks on unrooted binary trees [5, 6] conveniently restricts to the space $\mathcal{BT}_S$ of rooted binary trees as follows:

Definition 4. *Let $\sigma \in \mathcal{BT}_S$. We say that $\tau \in \mathcal{BT}_S$ is the result of performing a* Nearest Neighbor Interchange (NNI) move *on σ at a grandchild $G \in \mathcal{G}(\sigma)$ (8) if*

$$\mathcal{C}(\tau) = \Big(\mathcal{C}(\sigma) \setminus \{\mathrm{Pr}(G,\sigma)\}\Big) \cup \Big\{\mathrm{Pr}^2(G,\sigma) \setminus G\Big\}. \tag{10}$$

We often indicate this by writing $\tau = NNI(\sigma, G)$.

Note that the NNI move at cluster G on σ swaps cluster G with its parent's sibling $\mathrm{Pr}(G,\sigma)^{-\sigma}$ to yield τ, depicted in Figure 2(left); and after an NNI move at cluster G of σ, grandchild G of grandparent $P = \mathrm{Pr}^2(G,\sigma)$ with respect to σ becomes child G of parent $P = \mathrm{Pr}(G,\tau)$ with respect to τ.

It is standard to say that $\sigma, \tau \in \mathcal{BT}_S$ are NNI-adjacent if and only if one can be obtained from the other by a single move. Figure 2(left) illustrates the moves on $\mathcal{BT}_S$ and their inverses.

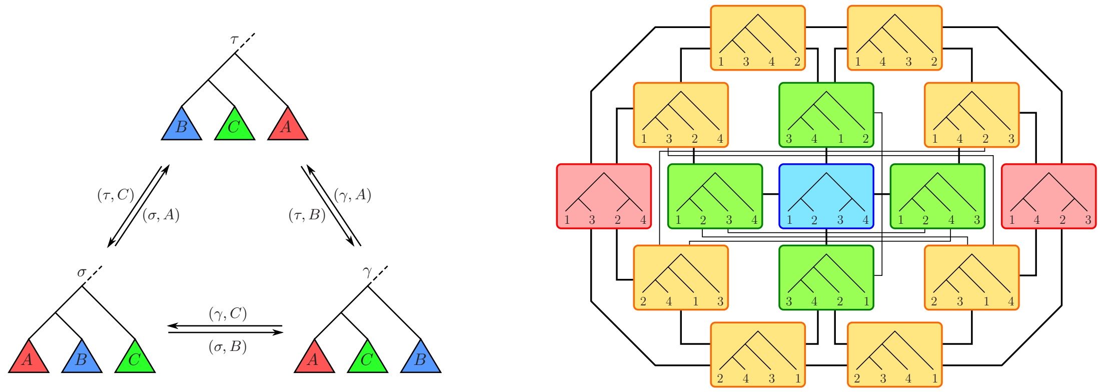

Figure 2. NNI moves (arrows, left) between binary trees, each move is labeled by its source tree and the grandchild defining the move, and the NNI Graph for $S = [4] = \{1, 2, 3, 4\}$ (right). These figures are adapted from [12].

The NNI-graph is formed over the vertex set $\mathcal{BT}_S$ by declaring two trees to be connected by an edge if and only if they are NNI-adjacent, see e.g. Figure 2(right). We will work with a directed version of this graph:

Definition 5. *The* directed NNI graph *$\mathcal{N}_S = (\mathcal{BT}_S, \mathcal{E}_S)$ is the directed graph on $\mathcal{BT}_S$ with $(\sigma, \tau) \in \mathcal{E}_S$ iff τ results from applying an NNI move to σ. We will henceforth identify the notation for an NNI move (σ, G), $G \in \mathcal{G}(\sigma)$ with the directed edge $(\sigma, NNI(\sigma, G)) \in \mathcal{E}_S$ wherever there is no danger of confusion.*

The (directed) NNI-graph on n leaves is a regular graph of out-degree $2(n-2)$ [5]. Our description clarifies this by parametrizing the set of neighbors of $\tau \in \mathcal{BT}_S$ with its grandchildren, $|\mathcal{G}(\tau)| = 2(|S|-2)$. The vertex set of the NNI graph is known to grow super exponentially with the number of leaves [24, 27, 28], [2]

$$\left|\mathcal{BT}_{[n]}\right| = (2n-3)!! \overset{\text{def.}}{=} (2n-3)(2n-5)\cdots 3 \cdot 1, \quad n \geq 2. \tag{11}$$

[2] In enumerative combinatorics, the double factorial is commonly encountered for counting combinatorial objects, such as binary trees [27, 28].

As a result, exploration of the NNI-graph (for example, searching for the shortest path between hierarchies or an optimal phylogenetic tree model) rapidly becomes impractical and costly as the number of leaves increases. A useful observation for NNI-adjacent trees is:

Lemma 1. *An ordered pair of hierarchies* (σ, τ) *is an edge in* $\mathcal{N}_S$ *if and only if there exists an ordered triple* (A, B, C) *of common clusters of* σ *and* τ *such that* $\{A \cup B\} = \mathcal{C}(\sigma) \setminus \mathcal{C}(\tau)$ *and* $\{B \cup C\} = \mathcal{C}(\tau) \setminus \mathcal{C}(\sigma)$*. The triple* (A, B, C) *is uniquely determined by* (σ, τ) *and will be referred to as the* NNI-triplet *associated with* (σ, τ)*.*

Proof. The proof amounts to a formal restatement of the observations made in Figure 2(left). See Appendix A.1. ■

Observe that the triplet in reverse order (C, B, A) is the NNI-triple associated with the edge (τ, σ). Also note that the NNI moves on σ at A and on τ at C yield τ and σ, respectively.

2.2.2. Tree Restriction

Definition 6. *Let* S *be a fixed finite set and* $K \subseteq S$*. The restriction map* $res_K : \mathcal{P}(S) \to \mathcal{P}(K)$ *is defined to be*

$$res_K(\mathcal{A}) := \left\{ A \cap K \,\middle|\, A \in \mathcal{A},\ A \cap K \neq \varnothing \right\} \tag{12}$$

for any $\mathcal{A} \subseteq \mathcal{P}(S)$*. It is convenient to have* $\mathcal{A}\big|_K$ *denote* $res_K(\mathcal{A})$*. For* $\sigma \in \mathcal{T}_K$ *and* $\tau \in \mathcal{T}_S$ *we will write:*

$$\sigma = res_K(\tau) = \tau\big|_K \iff \mathcal{C}(\sigma) = \mathcal{C}(\tau)\big|_K . \tag{13}$$

Remark 2. *Let* $\tau \in \mathcal{BT}_S$ *and denote* $\{L, R\} := \mathrm{Ch}(S, \tau)$*. Then one has* $\mathcal{C}(\tau) = \mathcal{C}\left(\tau\big|_L\right) \cup \{S\} \cup \mathcal{C}\left(\tau\big|_R\right)$*.*

Lemma 2. *For any finite set* S *and* $K \subseteq S$ *with* $|K| \geq 2$*,* $res_K(\mathcal{BT}_S) = \mathcal{BT}_K$*.*

Proof. See Appendix A.2. ■

2.3. Dissimilarities, Metrics and Ultrametrics

Recall that a *dissimilarity measure* on X, or simply a *dissimilarity*, is a real-valued nonnegative symmetric function d on $X \times X$ satisfying $d(x, x) = 0$ for all $x \in X$. Recall that a dissimilarity d on X is *positive definite* if $d(x, y) = 0$ implies $x = y$ for all $x, y \in X$. Many approximations of the (NP-hard) NNI metric are positive definite dissimilarities [16–18]. A dissimilarity d is a metric if it satisfies the triangle inequality, $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$. For example:

Definition 7 ([1] and[3, 4])**.** *The Robinson-Foulds distance* d_{RF} *on* $\mathcal{T}_S$ *is defined by:* [3]

$$d_{RF}(\sigma, \tau) = \frac{1}{2} \left| \mathcal{C}(\sigma) \ominus \mathcal{C}(\tau) \right| , \quad \sigma, \tau \in \mathcal{T}_S . \tag{14}$$

The matching split distance d_{MS} *between a pair of hierarchies* σ *and* τ *in* $\mathcal{BT}_S$ *is defined to be the value of a minimum-weighted perfect matching in the graph* $G_S(\sigma, \tau)$ *obtained from* $\sigma, \tau \in \mathcal{BT}_S$ *as the complete bipartite graph with sides* $\mathcal{C}_{int}(\sigma)$ *and* $\mathcal{C}_{int}(\tau)$ *with each edge* $(I, J) \in \mathcal{C}_{int}(\sigma) \times \mathcal{C}_{int}(\tau)$ *carrying the weight* [4] $A_S(I, J) = \min\left(\left|I \ominus J\right|, \left|I \ominus J^C\right|\right)$.

It is known that $d_{RF} \leq d_{MS} \leq \frac{|S|+1}{2} d_{RF}$ [3], which explains the improvement of d_{MS} over d_{RF} in discriminative power. At the same time, the cost of computing a minimum weighted perfect matching in any $G_S(\sigma, \tau)$ is $O(|S|^{2.5} \log |S|)$, which motivates the search for dissimilarities producing similar improvement in discriminative power (bounding d_{RF} from above) yet having a lower computational cost than that of d_{MS}.

Recall that an *ultrametric* d on X is a metric on X satisfying the strengthened triangle inequality, $d(x, y) \leq \max(d(x, z), d(z, y))$ for all $x, y, z \in X$. The following is a restatement of a well known fact (see, e.g. [15, 29, 30]) revealing the relation between hierarchies and ultrametrics:

[3] Here, $\ominus$ denotes the symmetric set difference, i.e. $A \ominus B = (A \setminus B) \cup (B \setminus A)$ for any sets A and B.

[4] This corresponds to the Hamming distance of clusters.

Lemma 3. *Let $\tau \in \mathcal{T}_S$ and $h_\tau : \mathcal{C}(\tau) \to \mathbb{R}_{\geq 0}$. For any $i, j \in S$ let $(i \wedge j)_\tau$ denote the smallest cluster in $\mathcal{C}(\tau)$ containing the pair $\{i, j\}$. Then the dissimilarity on S given by*

$$d_\tau(i, j) := h_\tau\left((i \wedge j)_\tau\right), \quad i, j \in S, \tag{15}$$

is an ultrametric if and only if the following are satisfied for any $I, J \in \mathcal{C}(\tau)$:

(a) if $I \subseteq J$, then $h_\tau(I) \leq h_\tau(J)$,

(b) $h_\tau(I) = 0$ if and only if $|I| = 1$.

Proof. See Appendix A.3. ■

Recall that a set X may always inherit a metric from a metric space (Y, d_Y) by pullback: any injective map f of X into Y yields a metric d_X on X defined by $d_X(x_1, x_2) := d_Y(f(x_1), f(x_2))$ and known as the pullback of d_Y along f. For example, the RF metric is a pullback: it is common knowledge that the set $F(X)$ of all finite subsets of a set X forms a metric space under the metric $d(A, B) = |A \ominus B|$, which is one of the ways of defining Hamming distance; thus, the RF distance is (one half times) the pullback of this metric on $F(\mathcal{P}(S))$ under the map $\tau \mapsto \mathcal{C}(\tau)$.

3. Quantifying Incompatibility

3.1. The Cluster-Cardinality Distance

We now introduce an embedding of hierarchies into the space of matrices based on the relation between hierarchies and ultrametrics, summarized in Lemma 3:

Definition 8. *The* ultrametric representation *is the map $\mathbf{U} : \mathcal{T}_S \to \mathbb{R}^{|S| \times |S|}$ defined by $\mathbf{U}(\tau)_{ij} := h\left((i \wedge j)_\tau\right)$, where $h : \mathcal{P}(S) \to \mathbb{N}$ is set to be $h(I) := |I| - 1$, $I \subseteq S$.*

Lemma 4. *The map $\mathbf{U}$ is injective.*

Proof. To see the injectivity of $\mathbf{U}$ (Definition 8), we shall show that $\mathbf{U}(\sigma) \neq \mathbf{U}(\tau)$ for any $\sigma \neq \tau \in \mathcal{T}_S$.

Two trees $\sigma, \tau \in \mathcal{T}_S$ are distinct if and only if they have at least one unshared cluster. Accordingly, for any $\sigma \neq \tau \in \mathcal{T}_S$ consider a common cluster $I \in \mathcal{C}(\sigma) \cap \mathcal{C}(\tau)$ with distinct parents $\mathrm{Pr}(I, \sigma) \neq \mathrm{Pr}(I, \tau)$. Depending on the cardinality of parent clusters:

- If $|\mathrm{Pr}(I, \sigma)| = |\mathrm{Pr}(I, \tau)|$, then observe that there exists some $j \in \mathrm{Pr}(I, \sigma)$ such that $j \notin \mathrm{Pr}(I, \tau)$ because $\mathrm{Pr}(I, \sigma) \neq \mathrm{Pr}(I, \tau)$. In fact, notice that $j \in I^{-\sigma}$ and $j \notin I^{-\tau}$ (recall (7)). Hence, for any $i \in I$ we have $(i \wedge j)_\sigma = \mathrm{Pr}(I, \sigma)$ and $\mathrm{Pr}(I, \tau) \subsetneq (i \wedge j)_\tau$. Thus, it follows from Definition 8 that for any $i \in I$

$$\mathbf{U}(\sigma)_{ij} = \left|\mathrm{Pr}(I, \sigma)\right| - 1 < \mathbf{U}(\tau)_{ij} = \left|(i \wedge j)_\tau\right| - 1 \,. \tag{16}$$

- Otherwise, without loss of generality, let $|\mathrm{Pr}(I, \sigma)| < |\mathrm{Pr}(I, \tau)|$. Then, observe that for any $i \in I$ and $j \in I^{-\sigma}$,

$$\mathbf{U}(\sigma)_{ij} = |\mathrm{Pr}(I, \sigma)| - 1 < \mathbf{U}(\tau)_{ij} = \left|(i \wedge j)_\tau\right| - 1 \,, \tag{17}$$

  since $(i \wedge j)_\tau \supseteq \mathrm{Pr}(I, \tau)$.

Therefore, for any $\sigma \neq \tau \in \mathcal{BT}_S$ one has $\mathbf{U}(\sigma) \neq \mathbf{U}(\tau)$, and the result follows. ■

Using the embedding $\mathbf{U}$ of $\mathcal{T}_S$ into $\mathbb{R}^{|S| \times |S|}$, we can construct tree metrics by pulling back metrics induced from matrix norms, such as the one below:

Definition 9. *The* cluster-cardinality metric, *$d_{CC} : \mathcal{T}_S \times \mathcal{T}_S \to \mathbb{R}_{\geq 0}$, on $\mathcal{T}_S$ is defined to be* [5]

$$d_{CC}(\sigma, \tau) := \frac{1}{2}\left\|\mathbf{U}(\sigma) - \mathbf{U}(\tau)\right\|_1, \quad \sigma, \tau \in \mathcal{T}_S \,. \tag{18}$$

[5] Here $\|.\|_1$ denotes the 1-norm of a matrix, i.e. $\|\mathbf{U}\|_1 := \sum_{i=1}^{n}\sum_{j=1}^{n}\left|\mathbf{U}_{ij}\right|$ for $\mathbf{U} \in \mathbb{R}^{n \times n}$. Our choice of the 1-norm was guided by the resulting relationships between d_{CC} and the dissimilarity measures d_{CM} and d_{nav} introduced below. Other choices of norm on $\mathbb{R}^{S \times S}$ may prove useful.

Proposition 1. *The cluster-cardinality distance d_{CC} on $\mathcal{T}_S$ is computable in $O(|S|^2)$ time.*

Proof. The 1-norm of the difference of a pair of $|S| \times |S|$ matrices obviously requires $O(|S|^2)$ time to compute, giving a lower bound on the computation cost of d_{CC}. It remains to show that the embedding $\mathbf{U}$ (Definition 8) may be obtained at this cost. We proceed by induction based on a post-order traversal of the trees involved, $\tau \in \mathcal{T}_S$. For the base case, consider the two-leaf tree $\tau \in \mathcal{BT}_{[2]}$, i.e. $|S| = 2$: then we simply assign $\mathbf{U}(\tau) = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. For the induction step, assume $|S| \geq 3$ and denote $\mathrm{Ch}(S, \tau) = \{S_k\}_{1 \leq k \leq K}$, where $K \geq 2$ is the number of children of the root S in τ. We observe:

- For every singleton child $\{i\}$ of S in τ (if any), then set $\mathbf{U}(\tau)_{ii} = 0$, which takes up $O(1)$ time.
- Note that all clusters of τ and their sizes can be obtained in $O(|S|^2)$ time by a single post-order traversal, as each individual cluster (as well as its cardinality) takes at most linear time to compute from those of its children.
- Suppose that for any $1 \leq k \leq K$ and $|S_k| \geq 2$ the elements of $\mathbf{U}(\tau)$ associated with the subtree rooted at S_k can be computed in $O(|S_k|^2)$ time. Then, the total number of updates associated with the root S is $\sum_{k=1}^{K} \sum_{l=1}^{K} |S_k| |S_l|$ and corresponds to setting $\mathbf{U}(\tau)_{ij} = \mathbf{U}(\tau)_{ji} = |S| - 1$ for all $i \in S_k$, $j \in S_l$ and $1 \leq k, l \leq K$.

In total, the cost of obtaining $\mathbf{U}(\tau)$ is $\sum_{k=1}^{K} O(|S_k|^2) + \sum_{k=1}^{K} \sum_{l=1}^{K} |S_k| |S_l| + O(|S|^2) = O(|S|^2)$, as required. ■

The diameter, $\mathrm{diam}(X, d) := \max\left\{ d(x, y) \,\middle|\, x, y \in X \right\}$, of a finite metric space (X, d) is always of interest in algorithmic applications. Some known diameters for hierarchies [3, 4, 31] are:

$$\mathrm{diam}(\mathcal{T}_S, d_{RF}) = |S| - 2\,, \quad \mathrm{diam}(\mathcal{BT}_S, d_{MS}) = O(|S|^2)\,, \quad \mathrm{diam}(\mathcal{BT}_S, d_{NNI}) = O(|S| \log |S|) \tag{19}$$

For the cluster-cardinality distance, we have:

Proposition 2. $\mathrm{diam}(\mathcal{T}_S, d_{CC}) = O(|S|^3)$.

Proof. From Definition 8, the minimum and maximum ultrametric distances between two distinct elements of S are, respectively, 1 and $|S| - 1$, implying the bound

$$\max_{i,j \in S} \left(\mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij} \right) \leq |S| - 2 \quad \forall \sigma, \tau \in \mathcal{T}_S \ . \tag{20}$$

Hence, it follows from (18) that the diameter of $\mathcal{T}_S$ with respect to d_{CC} is bounded above as follows:

$$\mathrm{diam}(\mathcal{T}_S, d_{CC}) \leq \frac{1}{2} |S| (|S| - 1)(|S| - 2)\,. \tag{21}$$

Now consider two NNI-adjacent binary trees $\sigma, \tau \in \mathcal{BT}_S$ such that the NNI triple (A, B, C) associated with (σ, τ) (see Lemma 1) satisfies $|A| = |B| = |C| = \left\lfloor \frac{|S|}{3} \right\rfloor$. It is straightforward to observe that for $|S| \geq 3$, there always exists such a pair of NNI-adjacent trees, because A, B, and C are disjoint and $|A| + |B| + |C| = 3 \left\lfloor \frac{|S|}{3} \right\rfloor \leq |S|$. Hence, we have from Proposition 3 that $d_{CC}(\sigma, \tau) = 2 |A| |B| |C| = 2 \left\lfloor \frac{|S|}{3} \right\rfloor^3$, which yields the following lower bound on the diameter of $\mathcal{T}_S$ with respect to d_{CC},

$$2 \left\lfloor \frac{|S|}{3} \right\rfloor^3 \leq \mathrm{diam}(\mathcal{T}_S, d_{CC})\,. \tag{22}$$

Note that these bounds on $\mathrm{diam}(\mathcal{T}_S, d_{CC})$ in (21) and (22) hold for all $|S| \geq 2$. Thus, the result follows. ■

A common question regarding any distance being proposed for the space of trees is how it behaves with respect to certain tree rearrangements. For instance, any pair of NNI-adjacent trees, $\sigma, \tau \in \mathcal{BT}_S$, are known to satisfy [3] [6]

$$d_{NNI}(\sigma, \tau) = 1 \iff d_{RF}(\sigma, \tau) = 1\,, \tag{23}$$

$$d_{NNI}(\sigma, \tau) = 1 \implies 2 \leq d_{MS}(\sigma, \tau) \leq \left\lfloor \frac{|S|}{2} \right\rfloor\,. \tag{24}$$

Similarly for d_{CC} we have:

[6] $\lfloor . \rfloor$ denotes the floor operator returning the largest integer not greater than its operand.

Proposition 3. *Let (σ, τ) be an edge of the NNI-graph $\mathcal{N}_S = (\mathcal{BT}_S, \mathcal{E})$ and (A, B, C) be the associated NNI triplet (Lemma 1). Then*

$$2 \leq d_{CC}(\sigma, \tau) = 2\,|A|\,|B|\,|C| \leq \left\lfloor \frac{2}{27} |S|^3 \right\rfloor , \tag{25}$$

and both bounds are tight.

Proof. Let $P = A \cup B \cup C$ and recall from Lemma 1 that $A \cup B \in \mathcal{C}(\sigma)$ and $B \cup C \in \mathcal{C}(\tau)$. Note that $P \in \mathcal{C}(\sigma) \cap \mathcal{C}(\tau)$ is a common (grand)parent cluster, and A, B and C are pairwise disjoint.

Since the NNI moves between σ and τ only change the relative relations of clusters A, B and C, the distance between σ and τ can be rewritten as

$$\begin{aligned} d_{CC}(\sigma, \tau) &= \frac{1}{2} \left\| \mathbf{U}(\sigma) - \mathbf{U}(\tau) \right\|_1 , && (26)\\ &= \sum_{\substack{i \in A\\ j \in B}} \left| \mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij} \right| + \sum_{\substack{i \in A\\ j \in C}} \left| \mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij} \right| + \sum_{\substack{i \in B\\ j \in C}} \left| \mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij} \right| , && (27)\\ &= \sum_{\substack{i \in A\\ j \in B}} \underbrace{\left| h(A \cup B) - h(P) \right|}_{=|C|} + \sum_{\substack{i \in A\\ j \in C}} \underbrace{\left| h(P) - h(P) \right|}_{=0} + \sum_{\substack{i \in B\\ j \in C}} \underbrace{\left| h(P) - h(B \cup C) \right|}_{=|A|} , && (28)\\ &= 2\,|A|\,|B|\,|C| . && (29) \end{aligned}$$

Clearly, the lower bound in (25) is realized when $|A| = |B| = |C| = 1$. Since the maximum product of three numbers with a prescribed sum occurs when all the numbers are equal — in our case, $|A| + |B| + |C| \leq |S|$ — we must have $|A|\,|B|\,|C| \leq \left\lfloor \frac{|S|^3}{27} \right\rfloor$, as $\lfloor . \rfloor$ is integer-valued. The result follows. ∎

Inequalities of the above form allow one to take advantage of the combinatorial nature of d_{NNI} through repeated application of the triangle inequality:

Corollary 1. *Over $\mathcal{BT}_S$ one has $d_{RF} \leq d_{NNI}$.*

Proof. Let $\sigma, \tau \in \mathcal{BT}_S$ and let $\Gamma = (\gamma_k)_{0 \leq k \leq K}$ be shortest path in the NNI graph $\mathcal{N}_S = (\mathcal{BT}_S, \mathcal{E})$ from $\sigma = \gamma_0$ to $\tau = \gamma_K$. This means that $(\gamma_{k-1}, \gamma_k) \in \mathcal{E}$ – or, equivalently, $d_{NNI}(\gamma_{k-1}, \gamma_k) = 1$ – for all $1 \leq k \leq K$, and that $K = d_{NNI}(\sigma, \tau)$. Repeatedly applying the triangle inequality for d_{RF} and then equation (23), we obtain:

$$d_{RF}(\sigma, \tau) \leq \sum_{k=1}^{K} d_{RF}(\gamma_{k-1}, \gamma_k) = \sum_{k=1}^{K} d_{NNI}(\gamma_{k-1}, \gamma_k) = K = d_{NNI}(\sigma, \tau) , \tag{30}$$

which completes the proof. ∎

Indeed, the length of a path in $\mathcal{N}_S$ produces a bound on the RF distance between its endpoints by repeatedly applying the triangle inequality to (23). A similar argument yields:

Corollary 2. *Let d be a dissimilarity on $\mathcal{BT}_S$ with the property that $d(\sigma, \tau) \leq 1$ for any pair of NNI-adjacent hierarchies $\sigma, \tau \in \mathcal{BT}_S$. If $d(\sigma, \tau) > d_{NNI}(\sigma, \tau)$ for some $\sigma, \tau \in \mathcal{BT}_S$, then d is not a metric.*

Proof. Assume, on the contrary, that d is a metric. Then the argument of the proof of Corollary 1 may be repeated, replacing d_{RF} with d and reaching the conclusion that $d(\sigma, \tau) \leq d_{NNI}(\sigma, \tau)$ for all $\sigma, \tau \in \mathcal{BT}_S$ – contradiction. ∎

3.2. The Crossing Dissimilarity

Definition 10. *Let $\sigma, \tau \in \mathcal{T}_S$. We define their* compatibility matrix $\mathbf{C}(\sigma, \tau)$ *and their* crossing matrix $\mathbf{X}(\sigma, \tau)$ *to be*[7]

$$\mathbf{C}(\sigma, \tau)_{I,J} := \mathbb{1}(I \bowtie J) \quad \textit{and} \quad \mathbf{X}(\sigma, \tau)_{I,J} := 1 - \mathbf{C}(\sigma, \tau)_{I,J} , \tag{31}$$

[7] $\mathbf{C}(\sigma, \tau)$ and $\mathbf{X}(\sigma, \tau)$ can be defined only in terms of nontrivial clusters of σ and τ since any trivial cluster of σ and τ is compatible with any cluster $K \subseteq S$. As a result, we are required to separately consider the special case in which one of the trees has only trivial clusters whenever $\mathbf{C}$ or $\mathbf{X}$ are used to reason about degenerate trees.

where $I \in \mathcal{C}(\sigma)$, $J \in \mathcal{C}(\tau)$ and $\mathbb{1}(.)$ denotes the indicator function returning unity if its argument holds true and zero otherwise. The crossing dissimilarity *d_{CM} is defined by $d_{CM}(\sigma, \tau) := \left\|\mathbf{X}(\sigma, \tau)\right\|_1$, counting*[8] *the pairs of incompatible clusters in $\mathcal{C}(\sigma) \cup \mathcal{C}(\tau)$.*

We list some useful properties of d_{CM}:

Remark 3. *The crossing dissimilarity d_{CM} on $\mathcal{BT}_S$ is positive definite and symmetric, but it is not a metric (apply Corollary 2 to the observations of Figure 3).*

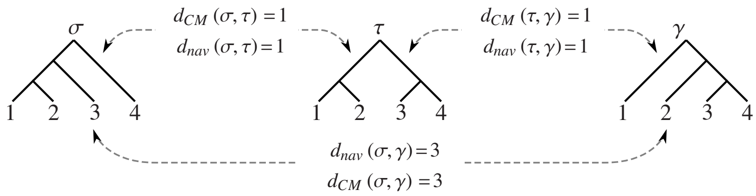

Figure 3. d_{CM} and d_{nav} are not metrics: an example of the triangle inequality failing for both dissimilarities.

Proposition 4. *The crossing dissimilarity d_{CM} over $\mathcal{T}_S$ can be computed in $O(|S|^2)$ time.*

Proof. The crossing matrix $\mathbf{X}(\sigma, \tau)$ (31) of a pair of hierarchies $\sigma, \tau \in \mathcal{T}_S$ has at most $2|S| - 1$ rows and columns. Hence, the 1-norm of $\mathbf{X}(\sigma, \tau)$ requires $O(|S|^2)$ time to compute, bounding the cost of d_{CM} from below. To obtain the upper bound, we show that $\mathbf{X}(\sigma, \tau)$ can be obtained in $O(|S|^2)$ time by post-order traversal.

Observe that for any cluster $J \in \mathcal{C}(\tau)$ (and symmetrically, for any cluster of $\mathcal{C}(\sigma)$) one can check whether J is disjoint with or a superset of each cluster I of σ by a post-order traversal of σ in $O(|S|)$ time using the following recursion:

- If either I or J is a singleton then the cluster inclusions $I \subseteq J$, $J \subseteq I$ and their disjointness can be determined in constant time using a hash map.
- Otherwise ($|I| \geq 2$ and $|J| \geq 2$), we have

$$I \subseteq J \iff \forall D \in \mathrm{Ch}(I, \sigma) \quad D \subseteq J, \tag{32}$$

$$I \cap J = \emptyset \iff \forall D \in \mathrm{Ch}(I, \sigma) \quad D \cap J = \varnothing. \tag{33}$$

Thus, it follows from Definition 1 that a complete list of compatibilities between σ and τ can be produced in $O(|S|^2)$ time, and so $\mathbf{X}(\sigma, \tau)$ can be obtained at the same cost, $O(|S|^2)$. ■

Proposition 5. $\mathrm{diam}(\mathcal{T}_S, d_{CM}) = (|S| - 2)^2$.

Proof. Two clusters of a pair of trees can only be incompatible if they are both nontrivial. Recall from Remark 1 that the number of nontrivial clusters of a tree in $\mathcal{T}_S$ is at most $|S| - 2$. Hence, by Definition 10, an upper bound on $\mathrm{diam}(\mathcal{T}_S, d_{CM})$ is $(|S| - 2)^2$. To observe that this upper bound is realized, see Figure 4. ■

Proposition 6. *Two nondegenerate trees $\sigma, \tau \in \mathcal{BT}_S$ are NNI-adjacent if and only if $d_{CM}(\sigma, \tau) = 1$.*

Proof. The result is evident from Remark 1 and Definition 4. ■

Despite the result of the last proposition, d_{CM} does not provide a linear lower bound on d_{NNI} since $\mathrm{diam}(\mathcal{BT}_S, d_{NNI}) = O(|S| \log |S|) < \mathrm{diam}(\mathcal{BT}_S, d_{CM}) = O(|S|^2)$ (Proposition 5). This inequality provides us with an additional, more conceptual, argument that d_{CM} is not a metric, by applying Corollary 2.

[8] We find that choosing to use the 1-norm of the crossing matrix easily reveals combinatorial relations between d_{CM} and d_{CC} (18); of course, one could use other matrix norms to construct alternative dissimilarities.

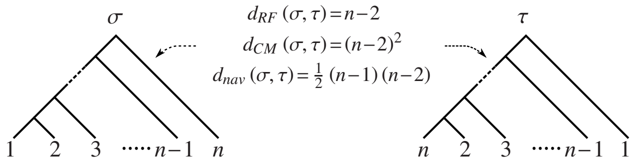

Figure 4. A pair of nondegenerate hierarchies realizing $\operatorname{diam}(\mathcal{T}_{[n]}, d_{CM}) = (n-2)^2$ and $\operatorname{diam}(\mathcal{BT}_{[n]}, d_{nav}) = \frac{1}{2}(n-1)(n-2)$.

Proposition 7. *Over $\mathcal{T}_S$ one has $d_{RF} \leq d_{CM} \leq d_{RF}^2$. These bounds are tight.*

Proof. The lower bound directly follows from Remark 1. Because a pair of distinct binary hierarchies always have uncommon clusters whose count is equal to d_{RF}, and an unshared cluster of one tree crosses at least one unshared cluster of the other tree. This bound is tight since for any $\sigma, \tau \in \mathcal{BT}_S$

$$d_{RF}(\sigma,\tau) = 1 \Leftrightarrow d_{NNI}(\sigma,\tau) = 1 \Leftrightarrow d_{CM}(\sigma,\tau) = 1. \tag{34}$$

For any $\sigma, \tau \in \mathcal{BT}_S$, the columns and rows of $\mathbf{X}(\sigma,\tau)$ (31) associated with common clusters of σ, τ are necessarily null. Hence, $\mathbf{X}(\sigma,\tau)_{I,J} \neq 0$ implies $I \notin \mathcal{C}(\tau)$ and $J \notin \mathcal{C}(\sigma)$. By the definition of d_{RF}, there are no more than $d_{RF}(\sigma,\tau)^2$ such pairs — hence the claimed upper bound. To observe that this bound is also tight, see Figure 4. ∎

Proposition 8. *Over $\mathcal{T}_S$ one has $d_{CM} \leq d_{CC}$.*

Proof. Given any $\sigma, \tau \in \mathcal{T}_S$ we claim that there is a function $q : \mathcal{C}(\sigma) \times \mathcal{C}(\tau) \to S \times S$ with the following properties:

(a) for any $I \in \mathcal{C}(\sigma)$ and $J \in \mathcal{C}(\tau)$, $I \bowtie J$ if and only if $(i, j) = q(I, J)$ with $i = j$,

(b) for any $i \neq j \in S$, $\left|q^{-1}(i,j)\right| \leq \left|\mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij}\right|$.

Observe that, if such a function does exist, then (a) implies:

$$\bigcup_{i \neq j \in S} q^{-1}(i,j) = \left\{(I,J) \in \mathcal{C}(\sigma) \times \mathcal{C}(\tau) \,\middle|\, I \not\bowtie J\right\}. \tag{35}$$

It is then evident from (35) and (b) that

$$d_{CM}(\sigma,\tau) \leq \sum_{i \neq j \in S} \left|q^{-1}(i,j)\right| \leq d_{CC}(\sigma,\tau), \tag{36}$$

proving our proposition.

We proceed to construct the function q. If $I \not\bowtie J$, then there exist $i \in I \cap J$ and $j \in I \setminus J$ with the property that $(i \wedge j)_\sigma = I$. Accordingly, define

$$Q(I,J) := \left\{(i,j) \in S \times S \,\middle|\, i \in I \cap J, j \in I \setminus J, (i \wedge j)_\sigma = I\right\}, \tag{37}$$

$$R(I,J) := \left\{(i,j) \in S \times S \,\middle|\, i \in I \cap J, j \in J \setminus I, (i \wedge j)_\tau = J\right\}. \tag{38}$$

Note that if $(i, j) \in Q(I, J) \cup R(I, J)$, then $i \neq j$.

Have S totally ordered (say, by enumerating its elements) and have $S \times S$ ordered lexicographically according to the order of S. Then, define $q : \mathcal{C}(\sigma) \times \mathcal{C}(\tau) \to S \times S$ to be

$$q(I,J) := \begin{cases} (\min(I \cup J), \min(I \cup J)) & , \text{if } I \bowtie J, \\ \min Q(I,J) & , \text{if } I \not\bowtie J, |I| \leq |J|, \\ \min R(I,J) & , \text{if } I \not\bowtie J, |I| > |J|. \end{cases} \tag{39}$$

Recall that $Q(I, J)$ and $R(I, J)$ both contain pairs of distinct elements of S. Hence, q satisfies the property (a) above.

By construction, for any $i \neq j$ we have:

$$q^{-1}(i, j) \subseteq A(i, j) \cup B(i, j), \tag{40}$$

where

$$A(i, j) := \left\{ (I, J) \in \mathcal{C}(\sigma) \times \mathcal{C}(\tau) \,\middle|\, I \not\bowtie J, |I| \leq |J|, (i, j) \in Q(I, J) \right\}, \tag{41}$$

$$B(i, j) := \left\{ (I, J) \in \mathcal{C}(\sigma) \times \mathcal{C}(\tau) \,\middle|\, I \not\bowtie J, |I| \geq |J|, (i, j) \in R(I, J) \right\}. \tag{42}$$

Remark from (37) that if $(I, J) \in A(i, j)$ then $(i \wedge j)_\sigma = I$ and $(i \wedge j)_\tau \supsetneq J$. Hence, if $\left|(i \wedge j)_\sigma\right| \geq \left|(i \wedge j)_\tau\right|$, then $A(i, j) = \varnothing$. Similarly, $(i \wedge j)_\sigma \supsetneq I$ and $(i \wedge j)_\tau = J$ whenever $(I, J) \in B(i, j)$; and $B(i, j) = \varnothing$ if $\left|(i \wedge j)_\sigma\right| \leq \left|(i \wedge j)_\tau\right|$. Thus, one can observe that for any $i, j \in S$,

$$A(i, j) \neq \varnothing \Longrightarrow B(i, j) = \varnothing. \tag{43}$$

Recall that for any $i, j \in S$ and $(I, J) \in A(i, j)$ we have:

$$I = (i \wedge j)_\sigma, \; J \subsetneq (i \wedge j)_\tau, \; |I| \leq |J| \text{ and } J \in \mathrm{Anc}(\{i\}, \tau). \tag{44}$$

Hence, one can conclude that

$$\left|A(i, j)\right| \leq \left| \left|(i \wedge j)_\tau\right| - \left|(i \wedge j)_\sigma\right| \right| = \left|\mathbf{U}(\tau)_{ij} - \mathbf{U}(\sigma)_{ij}\right|. \tag{45}$$

Similarly, for any $i, j \in S$

$$\left|B(i, j)\right| \leq \left| \left|(i \wedge j)_\sigma\right| - \left|(i \wedge j)_\tau\right| \right| = \left|\mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij}\right|. \tag{46}$$

Thus, overall, using (40) and (43), one can obtain the second property of q as follows: for any $i \neq j \in S$

$$\left|q^{-1}_{\sigma,\tau}(i, j)\right| \leq \left|A(i, j)\right| + \left|B(i, j)\right| \leq \left|\mathbf{U}(\tau)_{ij} - \mathbf{U}(\sigma)_{ij}\right|, \tag{47}$$

which completes the proof. ∎

4. The Navigation Dissimilarity

Problem 1 may be loosely restated in graph-theoretic terms as follows:

Problem 2. *For each tree $\tau \in \mathcal{BT}_S$, find a subgraph $\mathcal{N}_{S,\tau}$ of the NNI graph $\mathcal{N}_S$ containing no directed cycles and such that every $\sigma \in \mathcal{BT}_S$ satisfies:*

(†) If $\sigma \neq \tau$ then there exists an edge of $\mathcal{N}_{S,\tau}$ exiting σ; moreover, such an edge may be produced in low time complexity.

Clearly, the reactive navigation algorithm $\mathcal{A}_\tau$ of Problem 1 is, in this case, to compute an edge of $\mathcal{N}_{S,\tau}$ exiting the input tree σ and then follow that edge. The challenge for us is to produce a graph (Definition 18) where (i) the complexity of $\mathcal{A}_\tau$ is low (Corollary 5), and (ii) the length of any directed path is bounded by a reasonable function of $d_{NNI}(\sigma, \tau)$, or, at least of $n = |S|$ (Definition 19, Theorem 2 and Corollary 4). Observe the similarity between our requirements of $\mathcal{N}_{S,\tau}$ and a skeletal variant of the stricter notion of a *combing* from the early days of geometric group theory (see, e.g. [32]): a 'coherent' system of paths $\{p_x\}_x \in X$ in a topological space X, one for each point of the space, with $p_x(0) = x_0$ for all $x \in X$ and $p_y(t) = p_x(t)$ for all $t \leq s$ whenever $y = p_x(s)$. Specializing to the differentiable setting, one might hope to be able to (efficiently) compute a tangent vector t_x to p_x at x in some open dense (and necessarily contractible) sub-manifold of X so that the p_x become integral curves of $\dot{x} = t_x$; following these curves in reverse comprises reactive navigation towards x_0, as seen through the eyes of a roboticist [14].

We start out with a study of the coarse structure of the directed NNI graph $\mathcal{N}_S$. We consider special subspaces of the vertex space $\mathcal{BT}_S$:

Definition 11. *Let $K_1, \ldots, K_m$, $m \geq 1$, be a compatible family of subsets of S. Denote:*

$$\mathcal{BT}_S(K_1, \ldots, K_m) := \left\{ \sigma \in \mathcal{BT}_S \,\middle|\, \sigma \Join \{K_1, \ldots, K_m\} \right\} . \tag{48}$$

Recalling that $\mathcal{C}(\sigma)$ is a maximal nested family in $\mathcal{P}(S)$ if and only if $\sigma \in \mathcal{BT}_S$, one has, in fact:

$$\mathcal{BT}_S(K_1, \ldots, K_m) = \left\{ \sigma \in \mathcal{BT}_S \,\middle|\, K_1, \ldots, K_m \in \mathcal{C}(\sigma) \right\} . \tag{49}$$

Intuitively, it is clear that the problem of navigating $\mathcal{N}_S$ towards a specified tree τ may be parsed into a sequence of problems, each being that of navigating in $\mathcal{BT}_S(K)$ towards $\mathcal{BT}_S(K) \cap \mathcal{BT}_S(\mathrm{Ch}(K, \tau))$, where K ranges over $\mathcal{C}(\tau)$, starting with $K = S$ and continuing inductively, provided each step preserves the achievements of its predecessors.

4.1. Resolving incompatibilities with a prescribed split

Throughout this section, let $\mathbb{K} = \{K_1, K_2\}$ be a fixed pair of disjoint non-empty subsets of S, and set $K = K_1 \cup K_2$. We will refer to such pairs as *partial splits*. Let us make a simple observation:

Lemma 5. *The following equivalence holds for all $I \subsetneq K$:*

$$I \Join \mathbb{K} \iff (I \subseteq K_1) \vee (I \subseteq K_2) . \tag{50}$$

Proof. Suppose $I \Join \mathbb{K}$ but neither $I \subseteq K_1$ nor $I \subseteq K_2$ holds. By Definition 1 we must then have $I \supseteq K_1$ and $I \supseteq K_2$, implying $I \supseteq K$ — contradiction to $I \subsetneq K$. The converse is trivial. ■

Let $\sigma \in \mathcal{BT}_S(K)$ be a tree which splits K into a pair of children not coinciding with $\mathbb{K}$. According to the preceding lemma, this is equivalent to $\mathrm{Ch}(K, \sigma) \not\Join \mathbb{K}$. Observe now that any cluster $I \in \mathcal{C}(\sigma)$ which is not a σ-descendant of K is automatically compatible with $\mathbb{K}$. Thus, incompatibilities of σ with $\mathbb{K}$ could only occur among σ-descendants of K. This motivates the following definition:

Definition 12 (Recombinants)**.** *For $\sigma \in \mathcal{BT}_S(K)$ we distinguish two classes of σ-descendants of the cluster K:*

$$\begin{aligned} \mathcal{I}(\sigma; \mathbb{K}) &:= \left\{ I \in \mathrm{Des}(K, \sigma) \,\middle|\, I \not\Join \mathbb{K} \right\}, &(51)\\ \mathcal{R}(\sigma; \mathbb{K}) &:= \left\{ I \in \mathcal{I}(\sigma; \mathbb{K}) \,\middle|\, \mathrm{Ch}(I, \sigma) \Join \mathbb{K},\ \mathrm{Ch}(I^{-\sigma}, \sigma) \Join \mathbb{K} \right\} . &(52)\end{aligned}$$

For lack of a better term, we will refer to the elements of $\mathcal{R}(\sigma; \mathbb{K})$ as recombinants *of $\mathbb{K}$ in σ. See Figure 5.*

The set of recombinants suffices to characterize the compatibility of a tree with a given split:

Lemma 6. *Observe that $\sigma \in \mathcal{BT}_S(K)$ has recombinants of $\mathbb{K}$ if and only if $\sigma \notin \mathcal{BT}_S(\mathbb{K})$.*

Proof. Indeed, if $\sigma \in \mathcal{BT}_S(\mathbb{K})$, then all clusters of σ are compatible with $\mathbb{K}$, causing $\mathcal{I}(\sigma; \mathbb{K})$ — and hence also $\mathcal{R}(\sigma; \mathbb{K})$ — to be empty. Conversely, suppose there is a cluster of σ incompatible with $\mathbb{K}$. Then the σ-children of any deepest such cluster and its local complement's children are compatible with $\mathbb{K}$ in σ, and their children are compatible with $\mathbb{K}$ as well (even if vacuously). ■

Definition 13 (Incompatibility Types)**.** *Given $\sigma \in \mathcal{BT}_S(K)$, a cluster $I \in \mathcal{I}(\sigma; \mathbb{K})$ is said to be of* type 1 *with respect to $\mathbb{K}$ if $I^{-\sigma} \Join \mathbb{K}$. If $I \in \mathcal{I}(\sigma; \mathbb{K})$ is not of type 1, then it is said to be of* type 2 *(see Figure 5).*

Another, perhaps less intuitive, quantifier of incompatibility arises as follows:

Definition 14 (Essential Crossing Index)**.** *Let $\mathbb{K} = \{K_1, K_2\}$ and $\mathbb{L} = \{L_1, L_2\}$ be partial splits. Their* essential crossing index *is defined as:*

$$[\![\mathbb{L} \,|\, \mathbb{K}]\!] := \begin{cases} 0 & \text{if } \mathbb{L}\big|_{K_1 \cup K_2} \Join \mathbb{K}\big|_{L_1 \cup L_2} \\ 1 & \text{if } L_j\big|_{K_1 \cup K_2} \Join \mathbb{K}\big|_{L_1 \cup L_2} \text{ for only one } j \in \{1, 2\} \\ 3 & \text{otherwise} \end{cases} \tag{53}$$

For a tree $\sigma \in \mathcal{BT}_S$ we define:

$$\|\sigma\|_{\mathbb{K}} := \sum_{I \in \mathcal{C}(\sigma)} [\![\mathrm{Ch}(I, \sigma) \,|\, \mathbb{K}]\!] . \tag{54}$$

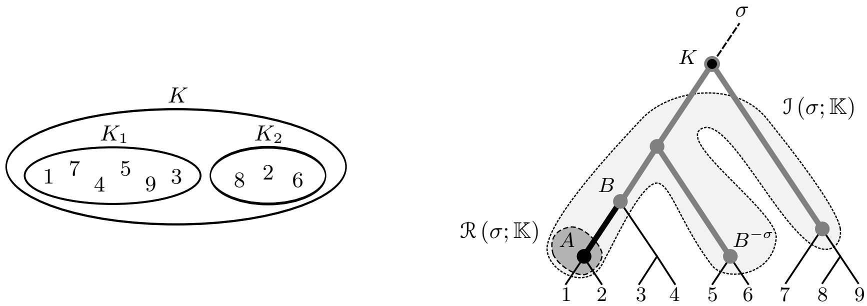

Figure 5. An illustration of $\mathcal{I}(\sigma;\mathbb{K})$ (51) and $\mathcal{R}(\sigma;\mathbb{K})$ (52) of $\sigma \in \mathcal{BT}_{[n]}(K)$, where $n \geq 9$ and $K = [9]$. The vertices and edges associated with clusters of σ incompatible with the split $\mathbb{K}$ are thickened. The only recombinant of $\mathbb{K}$ in σ is $A = \{1,2\}$ and it has Type 1. B and $B^{-\sigma}$ are examples of Type 2 clusters of σ incompatible with $\mathbb{K}$ which are not recombinants.

The following elementary observations will be useful:

Lemma 7. *Let* $\mathbb{K} = \{K_1, K_2\}$ *and* $\mathbb{L} = \{L_1, L_2\}$ *be partial splits. Then* $[\![\mathbb{K} \,|\, \mathbb{L}]\!] = [\![\mathbb{L} \,|\, \mathbb{K}]\!]$.

Proof. Write $K = K_1 \cup K_2$ and $L = L_1 \cup L_2$. Without loss of generality we may assume $K = L = S$, since:

$$\bigcup_{i=1}^{2}(K_i \cap L) = \bigcup_{j=1}^{2}(L_j \cap K) = K \cap L\,. \tag{55}$$

We study the possible cases:

- $[\![\mathbb{L} \,|\, \mathbb{K}]\!] = 0$: By definition, this means none of the K_i crosses any of the L_j; equivalently, no L_j crosses any of the K_i and we have $[\![\mathbb{K} \,|\, \mathbb{L}]\!] = 0$.
- $[\![\mathbb{L} \,|\, \mathbb{K}]\!] = 1$: WLOG, only L_1 crosses $\mathbb{K}$, hence L_2 is contained in one of the K_i, say K_2. Then L_1 contains K_1 and at least one element of K_2, by Lemma 5. Thus, $K_1 \bowtie \mathbb{L}$ while $K_2 \bowtie L_2$, $K_2 \not\bowtie L_1$. This means $[\![\mathbb{K} \,|\, \mathbb{L}]\!] = 1$.
- $[\![\mathbb{L} \,|\, \mathbb{K}]\!] = 3$: if both L_1 and L_2 cross $\mathbb{K}$, then $L_j \cap K_i \neq \varnothing$ for all $i, j \in 1, 2$, implying both K_1 and K_2 cross $\mathbb{L}$, as desired. ■

We are now ready to construct the graph $\Gamma_S(\mathbb{K})$:

Definition 15 (Projector Graph)**.** *Let* $\mathbb{K} = \{K_1, K_2\}$ *be a partial split, and set* $K = K_1 \cup K_2$*. Then* $\Gamma_S(\mathbb{K})$ *is defined to be the directed graph with vertex set* $\mathcal{BT}_S(K)$*, and all edges of the form* $(\sigma, G) \in \widetilde{\mathcal{E}}$ *such that* $I := \mathrm{Pr}(G,\sigma) \in \mathcal{R}(\sigma;\mathbb{K})$ *and one of the following holds:*

1. *I is of type 1, and* $G^{-\sigma}, I^{-\sigma} \subseteq K_i$ *for some* $i \in \{1,2\}$*;*
2. *I is of type 2.*

The following elementary property of edges in $\Gamma_S(\mathbb{K})$ is crucial:

Lemma 8. *Suppose* $\sigma \in \mathcal{BT}_S(K)$*,* (σ, G) *is an edge of* $\Gamma_S(\mathbb{K})$ *and* $\tau = NNI(\sigma, G)$*. Then* $\|\sigma\|_{\mathbb{K}} = \|\tau\|_{\mathbb{K}} + 1$.

Proof. Let $I = \mathrm{Pr}(G,\sigma)$ and let $J = I^{-\sigma} \cup G^{-\sigma}$ be the cluster replacing I in τ. Also, set $M = \mathrm{Pr}^2(G,\sigma) \in \mathcal{C}(\sigma) \cap \mathcal{C}(\tau)$. In the transition from σ to τ only the clusters I, J and M change (or lose, or acquire) their child splits. Therefore:

$$\|\tau\|_{\mathbb{K}} = \|\sigma\|_{\mathbb{K}} - [\![\mathrm{Ch}(I,\sigma) \,|\, \mathbb{K}]\!] + [\![\mathrm{Ch}(J,\tau) \,|\, \mathbb{K}]\!] - [\![\mathrm{Ch}(M,\sigma) \,|\, \mathbb{K}]\!] + [\![\mathrm{Ch}(M,\tau) \,|\, \mathbb{K}]\!]\ . \tag{56}$$

Figure 6 demonstrates without loss of generality that, in the case when I is of type 1 with respect to $\mathbb{K}$ the values of the above crossing indices are 0, 0, 1 and 0, respectively, resulting in a total decrease of one unit. The case when I is of type 2 produces the respective values of 0, 1, 3 and 1, also resulting in a total decrease of one unit. ■

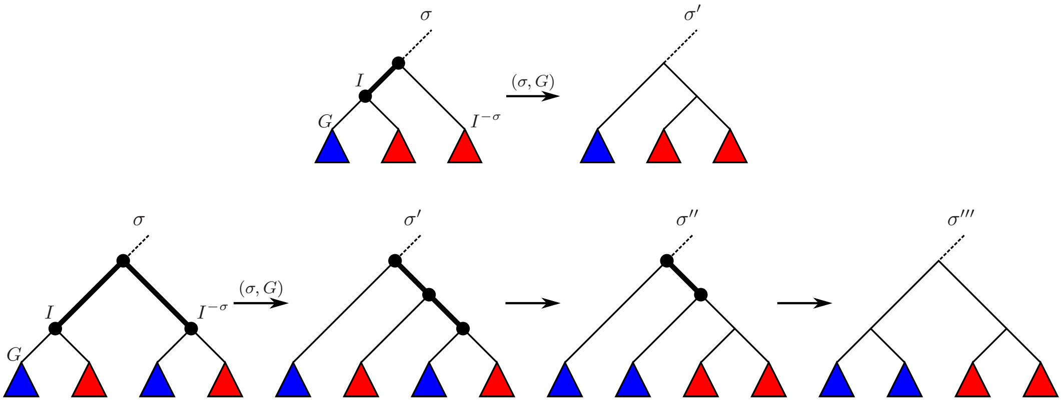

Figure 6. Different types of incompatibility — Type 1 (above) and Type 2 (below) — of a tree $\sigma \in \mathcal{BT}_S(K)$, $K = K_1 \cup K_2$, with the split $\mathbb{K} = \{K_1, K_2\}$, and the NNI moves suggested by $\Gamma_S(\mathbb{K})$ to resolve them. Clusters are colored blue or red according to their being contained in K_1 or K_2, respectively. Thickened vertices represent the recombinants affected by these moves.

Lemma 9. *The following are equivalent for a vertex $\sigma \in \mathcal{BT}_S(K)$ of $\Gamma_S(\mathbb{K})$:*

1. $\|\sigma\|_{\mathbb{K}} > 0$*;*

2. $\Gamma_S(\mathbb{K})$ *contains an edge exiting* σ*;*

3. $\sigma \notin \mathcal{BT}_S(\mathbb{K})$.

Proof. First observe that, since K is a cluster of σ, all clusters $I' \in \mathcal{C}(\sigma)$ not contained in K have $[\![I' \,|\, \mathbb{K}]\!] = 0$.

(1) ⇒ (2). By the preceding observation, if $\|\sigma\|_{\mathbb{K}} > 0$ then σ has a sub-cluster of K whose child split is incompatible with $\mathbb{K}$. By Lemma 6, σ then has a cluster $I \subsetneq K$ which is a recombinant of $\mathbb{K}$. Picking G to be an appropriate σ-child of I provides the required edge (σ, G).

(2) ⇒ (3). Suppose (σ, G) is an edge in $\Gamma_S(\mathbb{K})$. Then $I := \mathrm{Pr}(G, \sigma)$ is incompatible with $\mathbb{K}$, proving (3).

(3) ⇒ (1). Finally, if $\sigma \notin \mathcal{BT}_S(\mathbb{K})$ then σ contains a recombinant I whose parent $M = \mathrm{Pr}(I, \sigma)$ then must satisfy $[\![\mathrm{Ch}(M, \sigma) \,|\, \mathbb{K}]\!] > 0$, resulting in $\|\sigma\|_{\mathbb{K}} > 0$. ■

Definition 16 (Projection)**.** *Let $\mathbb{K} = \{K_1, K_2\}$ be a partial split, and set $K = K_1 \cup K_2$. For any $\sigma \in \mathcal{BT}_S(K)$ we define its* projection to $\mathcal{BT}_S(K) \cap \mathcal{BT}_S(\mathbb{K})$ *to be the tree $\gamma = \mathbf{P}_S(\sigma; \mathbb{K}) \in \mathcal{BT}_S(K) \cap \mathcal{BT}_S(\mathbb{K})$ whose clusters are of one of the following forms:*

(a) $I \in \mathcal{C}(\sigma)$ with $I \cap K = \varnothing$ or $K \subseteq I$;

(b) $I \cap K_i \in \mathcal{C}(\gamma)$, $i \in \{1, 2\}$ where $I \in \mathcal{C}(\sigma)$ (and $I \subseteq K$).

Remark 4. *The tree $\mathbf{P}_S(\sigma; \mathbb{K})$ is a well-defined binary tree in $\mathcal{BT}_S(K) \cap \mathcal{BT}_S(\mathbb{K})$ by Lemma 2 (applied to $\mathcal{BT}_K$).*

We are ready to state the main result of this section:

Theorem 1. *The directed graph $\Gamma_S(\mathbb{K})$ contains no directed cycles. Moreover, for every $\sigma \in \mathcal{BT}_S(K)$, every maximal directed path of $\Gamma_S(\mathbb{K})$ emanating from σ terminates at the tree $\mathbf{P}_S(\sigma; \mathbb{K}) \in \mathcal{BT}_S(K) \cap \mathcal{BT}_S(\mathbb{K})$ and has length $\|\sigma\|_{\mathbb{K}}$.*

Proof. Denote $\Gamma := \Gamma_S(\mathbb{K})$ for short. By Lemma 8, the function $\|\cdot\|_{\mathbb{K}}$ decreases by a unit along each edge of Γ, implying the absence of directed cycles in the graph. In particular, for each $\sigma \in \mathcal{BT}_S(K)$, the length of a directed path in Γ emanating from σ is bounded above by $\|\sigma\|_{\mathbb{K}}$. Since, by Lemma 9, $\sigma \in \mathcal{BT}_S(K)$ has an exiting edge in Γ if and only if $\|\sigma\|_{\mathbb{K}} > 0$, we conclude that all maximal directed paths in Γ emanating from σ have length exactly $\|\sigma\|_{\mathbb{K}}$ and terminate in $\mathcal{BT}_S(K) \cap \mathcal{BT}_S(\mathbb{K})$.

It will be useful to henceforth denote

$$\mathbf{Path}_{\mathbb{K}}(\sigma) := \left\{\mathbf{p} \,\middle|\, \mathbf{p} \text{ is a maximal directed path in } \Gamma_S(\mathbb{K}) \text{ emanating from } \sigma\right\} . \quad (57)$$

It remains to prove that every $\mathbf{p} \in \mathbf{Path}_{\mathbb{K}}(\sigma)$ terminates in $\mathbf{P}_S(\sigma;\mathbb{K})$.

We will prove the remaining assertion of the proposition by induction on $\|\sigma\|_{\mathbb{K}}$. More precisely, for any non-negative integer k let $S(k)$ denote the statement that for every $\tau \in \mathcal{BT}_S(K)$ satisfying $\|\tau\|_{\mathbb{K}} \leq k$ every path in $\mathbf{Path}_{\mathbb{K}}(\tau)$ terminates in $\mathbf{P}_S(\tau;\mathbb{K})$. Observing that $S(0)$ holds true by construction, we assume $S(k)$ holds for some $k \geq 0$ and deduce $S(k+1)$.

Suppose σ has $\|\sigma\|_{\mathbb{K}} = k+1$. Once again, consider any directed edge (σ, G) in $\Gamma_S(\mathbb{K})$, and write $\tau = \mathrm{NNI}(\sigma, G)$ with $\|\tau\|_{\mathbb{K}} = k$. Let γ and γ' denote the projections of σ and τ to $\mathcal{BT}_S(K) \cap \mathcal{BT}_S(\mathbb{K})$. Finally, letting $I = \mathrm{Pr}(G,\sigma)$ and $J = G^{-\sigma} \cup I^{-\sigma}$ we recall that $\mathcal{C}(\tau) = (\mathcal{C}(\sigma) \smallsetminus \{I\}) \cup \{J\}$. We observe the following:

- For any set $Q \subseteq S$ satisfying $Q \cap K = \varnothing \;\vee\; K \subseteq Q$ and for any tree σ' lying on a path in $\mathbf{Path}_{\mathbb{K}}(\sigma)$ — for the trees τ, γ and γ' in particular — one has $Q \in \mathcal{C}(\sigma)$ if and only if $Q \in \mathcal{C}(\sigma')$. Thus, $\mathcal{C}(\gamma) \setminus \mathcal{C}(\gamma')$ consists only of proper subsets of K.

- For a cluster $Q \subseteq K$ of σ with $Q \neq I$ we have $Q \cap K_i \in \mathcal{C}(\gamma) \implies Q \cap K_i \in \mathcal{C}(\gamma')$ for $i \in \{1,2\}$ because $\mathcal{C}(\sigma) \smallsetminus \{I\} \subset \mathcal{C}(\tau)$.

- Finally, we consider the clusters $I \cap K_i$: since $I \in \mathcal{R}(\sigma;\mathbb{K})$, the sets $I \cap K_i$ are precisely the children of I in σ, which makes them clusters of τ; since $I \cap K_i \subset K_i$, they are also clusters of γ'.

To summarize, we have found out that $\mathcal{C}(\gamma) \subseteq \mathcal{C}(\gamma')$. By the maximality of $\mathcal{C}(\gamma)$ as a nested family (Remark 1 and Remark 4) they must be equal and we conclude that $\gamma = \gamma'$. Applying the induction hypothesis, we deduce that every path in $\mathbf{Path}_{\mathbb{K}}(\sigma)$ starting with the edge (σ, G) must terminate in γ. Since the choice of edge (σ, G) was arbitrary, we are done. ■

4.2. The Navigation Distance

The following result has the flavor of a commutation relation between *different* projector graphs:

Lemma 10. *Fix a pair of distinct partial splits $\mathbb{K} = \{K_1, K_2\}$ and $\mathbb{L} = \{L_1, L_2\}$. Setting $K = K_1 \cup K_2$ and $L = L_1 \cup L_2$ assume in addition that $\{K, K_1, K_2\} \bowtie \{L, L_1, L_2\}$. Then, for any $\sigma \in \mathcal{BT}_S(K)$ and any edge $(\sigma, G) \in \Gamma_S(\mathbb{K})$ one has $\|NNI(\sigma, G)\|_{\mathbb{L}} = \|\sigma\|_{\mathbb{L}}$.*

Proof. As before, set $\tau = \mathrm{NNI}(\sigma, G)$ and consider the sets $I = \mathrm{Pr}(G,\sigma)$, $J = G^{-\sigma} \cup I^{-\sigma}$ and $M = \mathrm{Pr}^2(G,\sigma)$ — all contained in the cluster $K \in \mathcal{C}(\sigma) \cap \mathcal{C}(\tau)$ — and recall that $\mathcal{C}(\tau) = (\mathcal{C}(\sigma) \smallsetminus \{I\}) \cup \{J\}$. Without loss of generality, $G \subseteq K_1$ and $G^{-\sigma} \subseteq K_2$.

Once again we observe that the transition from σ to τ affects only the crossing indices of the clusters I, J, M (which are all contained in K) as follows:

$$\|\tau\|_{\mathbb{L}} = \|\sigma\|_{\mathbb{L}} - \underbrace{[\![\mathrm{Ch}(I,\sigma) \mid \mathbb{L}]\!]}_{\alpha} + \underbrace{[\![\mathrm{Ch}(J,\tau) \mid \mathbb{L}]\!]}_{\beta} - \underbrace{[\![\mathrm{Ch}(M,\sigma) \mid \mathbb{L}]\!]}_{\gamma} + \underbrace{[\![\mathrm{Ch}(M,\tau) \mid \mathbb{L}]\!]}_{\delta} . \quad (58)$$

Note that $K \neq L$, since otherwise the compatibility assumption and Lemma 5 would have forced $\mathbb{K} = \mathbb{L}$.

Suppose now that $K \cap L = \varnothing$. In this case the restrictions of $\mathbb{L}$ to I, J, M are all trivial and the corresponding crossing indices are all zero.

Suppose $K \subsetneq L$. Then, without loss of generality, we have $K \subseteq L_1$ by Lemma 5 and all children of I, J, M in σ and τ (as relevant) are compatible with $\mathbb{L}$, resulting again in zero crossing indices.

Since $K \bowtie L$, $K \neq L$, we need only consider two cases (we refer the reader again to Figure 6 for an illustration):

- **$L \subseteq K_1$.** We have $\mathrm{Ch}(I,\sigma)\big|_L = \{G \cap L, \varnothing\}$ and therefore $\alpha = 0$. Also, $\mathrm{Ch}(J,\tau)\big|_L = \{\varnothing, I^{-\sigma} \cap L\}$, so that $\beta = 0$. Finally, $\mathrm{Ch}(M,\sigma)\big|_L = \mathrm{Ch}(M,\tau)\big|_L = \{G \cap L, I^{-\sigma} \cap L\}$ produces $\gamma = \delta$.

- **$L \subseteq K_2$.** In this case we have $\mathrm{Ch}(I,\sigma)\big|_L = \{\varnothing, G^{-\sigma} \cap L\}$ and α is zero again. Similarly, observe that $\mathrm{Ch}(M,\tau)\big|_L = \{\varnothing, J \cap L\}$ gives $\delta = 0$. At the same time, $\mathrm{Ch}(J,\tau)\big|_L = \mathrm{Ch}(M,\sigma)\big|_L = \{G^{-\sigma} \cap L, I^{-\sigma} \cap L\}$, so that $\beta = \gamma$.

This finishes the proof. ■

Any pair of binary trees in $\mathcal{BT}_S$ has a common cluster (the cluster S, for example), and one might hope to quantify the discrepancy between a pair of trees by counting common clusters which split differently in the two trees (perhaps, somehow accounting for the depth of these clusters). This motivates:

Definition 17. *For any $\sigma, \gamma \in \mathcal{BT}_S$, let $\mathcal{K}(\sigma, \gamma)$ denote the set*

$$\mathcal{K}(\sigma, \gamma) := \left\{ K \in \mathcal{C}(\sigma) \cap \mathcal{C}(\gamma) \,\middle|\, \mathrm{Ch}(K, \sigma) \neq \mathrm{Ch}(K, \gamma) \right\}. \tag{59}$$

Remark 5. *In $\mathcal{BT}_S$, $\sigma = \tau$ if and only if $\mathcal{K}(\sigma, \tau) = \varnothing$.*

Corollary 3. *For all $\sigma, \tau \in \mathcal{BT}_S$ we have $\mathcal{K}(\sigma, \tau) := \left\{ K \in \mathcal{C}(\sigma) \cap \mathcal{C}(\tau) \,\middle|\, \mathrm{Ch}(K, \sigma) \not\bowtie \mathrm{Ch}(K, \tau) \right\}$.*

Proof. Follows directly from Lemma 5 and the definitions. ■

Given a prescribed target tree $\tau \in \mathcal{BT}_S$, the projector graphs introduced above give rise to a tool for achieving planned reductions in the number of clusters in $\mathcal{K}(\sigma, \tau)$ at a given depth, for any tree $\sigma \in \mathcal{BT}_S$. More formally, consider the following construction:

Definition 18 (Navigation Graph)**.** *Let $\tau \in \mathcal{BT}_S$. Then $\mathcal{N}_{S,\tau}$ denotes the directed subgraph of the NNI graph $\mathcal{N}_S$ with vertex set $\mathcal{BT}_S$ and all the edges (σ, G) for which there exists a cluster $K \in \mathcal{K}(\sigma, \tau)$ satisfying $(\sigma, G) \in \Gamma_S(\mathrm{Ch}(K, \tau))$.*

We proceed to prove statements about the navigation graph analogous to those we have shown to hold for the projector graphs. It is time to introduce:

Definition 19 (Navigation Distance)**.** *Let $\sigma, \tau \in \mathcal{BT}_S$. We define the navigation distance from σ to τ to be:*

$$d_{nav}(\sigma, \tau) := \sum_{K \in \mathcal{C}(\tau)} \|\sigma\|_{\mathrm{Ch}(K,\tau)}, \tag{60}$$

$$= \sum_{K \in \mathcal{C}(\tau)} \sum_{L \in \mathcal{C}(\sigma)} [\![\mathrm{Ch}(L, \sigma) \mid \mathrm{Ch}(K, \sigma)]\!]. \tag{61}$$

We also define the special crossing matrix $\mathbf{S}(\sigma, \tau)$ *by*

$$\mathbf{S}(\sigma, \tau)_{K,L} := [\![\mathrm{Ch}(L, \sigma) \mid \mathrm{Ch}(K, \tau)]\!], \quad \forall K \in \mathcal{C}(\sigma), L \in \mathcal{C}(\tau). \tag{62}$$

Thus, d_{nav} coincides with the standard 1-norm of the special crossing matrix.

Theorem 2. *For any $\tau \in \mathcal{BT}_S$ the graph $\mathcal{N}_{S,\tau}$ has no directed cycles. Moreover, for any $\sigma \in \mathcal{BT}_S$ every maximal directed path in $\mathcal{N}_{S,\tau}$ emanating from σ terminates in τ and has length $d_{nav}(\sigma, \tau)$. We will refer to such paths as* navigation paths from σ to τ.

Proof. First, observe from equation (60) that $d_{nav}(\sigma, \tau)$ is zero if and only if $\|\sigma\|_{\mathbb{K}} = 0$ for every pair $\mathbb{K}$ of siblings in τ. By Lemma 9, this is equivalent to saying that $\sigma \in \mathcal{BT}_S(\mathbb{K})$ for every pair of siblings in τ, or, in other words, that $\sigma = \tau$. Moreover, note that $d_{nav}(\sigma, \tau) > 0$ implies there is an edge of $\mathcal{N}_{S,\tau}$ exiting σ: indeed, if $\sigma \neq \tau$ then there exists a $K \in \mathcal{K}(\sigma, \tau)$ (Remark 5), so that $\sigma \notin \mathcal{BT}_S(\mathrm{Ch}(K, \tau))$; Lemma 9 guarantees an edge of $\Gamma_S(\mathrm{Ch}(K, \tau))$ exiting σ, which, by definition, is also an edge of $\mathcal{N}_{S,\tau}$.

Suppose now (σ, G) is an edge in $\mathcal{N}_{S,\tau}$. That is, there exists $K \in \mathcal{K}(\sigma, \tau)$ such that $(\sigma, G) \in \Gamma_S(\mathbb{K})$ where $\mathbb{K} = \mathrm{Ch}(K, \tau)$.

Suppose there were more than one such K, that is: suppose $K, L \in \mathcal{K}(\sigma, \tau)$, $K \neq L$, such that $I := \mathrm{Pr}(G, \sigma)$ is contained in both K and L, and such that $\mathrm{Ch}(I, \sigma)$ is incompatible both with $\mathrm{Ch}(K, \tau)$ and $\mathrm{Ch}(L, \tau)$. Since $\varnothing \neq I \subseteq K \cap L$ and $K \bowtie L$, we may assume $K \subsetneq L$. But then $K, L \in \mathcal{C}(\tau)$ and $K \neq L$ implies K is contained in a τ-child of L, denoted L_1. As $I \subseteq K$, we conclude that both σ-children of I are contained in L_1 — a contradiction to the assumption that $\mathrm{Ch}(I, \sigma)$ and $\mathrm{Ch}(L, \tau)$ are incompatible.

Let $\sigma' = \mathrm{NNI}(\sigma, G)$. Then, by Lemma 8, we have $\|\sigma'\|_{\mathrm{Ch}(K,\tau)} = \|\sigma\|_{\mathrm{Ch}(K,\tau)} - 1$. Moreover, Lemma 10 guarantees $\|\sigma'\|_{\mathrm{Ch}(L,\tau)} = \|\sigma\|_{\mathrm{Ch}(L,\tau)}$ for all $L \in \mathcal{C}(\tau)$, $L \neq K$. Applying equation (60) we obtain

$$d_{nav}(NNI(\sigma, G), \tau) = d_{nav}(\sigma, \tau) - 1\,. \tag{63}$$

Thus, $\mathcal{N}_{S,\tau}$ contains no directed cycles, and every maximal directed path in $\mathcal{N}_{S,\tau}$ emanating from a fixed $\sigma \in \mathcal{BT}_S$ terminates after precisely $d_{nav}(\sigma, \tau)$ steps. By the preceding paragraph, every such path may only terminate in τ. ∎

The solution to the navigation problem implied by this theorem yields the following (very crude) bounds on the performance of the corresponding reactive navigation algorithm:

Corollary 4 (Navigation Complexity)**.** *The length of a navigation path through $\mathcal{N}_S$ does not exceed* $\mathrm{O}(|S|^2)$.

Proof. Let $n = |S|$. For all $\sigma, \tau \in \mathcal{BT}_S$ we have $|\mathcal{C}(\tau)| = \mathrm{O}(n)$, implying $\mathbf{S}(\sigma, \tau)$ has $\mathrm{O}(n^2)$ entries. The value of $d_{nav}(\sigma, \tau)$ never exceeds three times the number of entries in $\mathbf{S}(\sigma, \tau)$. ∎

Corollary 5. *Given $\sigma, \tau \in \mathcal{BT}_S$, computing an edge of $\mathcal{N}_{S,\tau}$ exiting σ may be done in* $\mathrm{O}(|S|)$ *time.*

Proof. Using a look-up table for the clusters of τ [2], a cluster $K \in \mathcal{K}(\sigma, \tau)$ may be found in linear time by a traversal of σ. Next, an appropriate recombinant cluster may be found in linear time by post-order traversal of $\sigma\big|_K$ (compare with proof of Proposition 1). ∎

The last theorem emphasizes the crucial role of the fact that all navigation paths from σ to τ have the same length, equal to $d_{nav}(\sigma, \tau)$, irrespective of the order in which one chooses to resolve the incompatibilities between the two trees. We will now consider additional applications of the last theorem which will help us clarify the geometry of the navigation distance and its relationship to the other dissimilarities mentioned in this paper.

Lemma 11. *Let $\mathbb{K} = \{K_1, K_2\}$ be a partial split, let $\tau \in \mathcal{BT}_S(K_1, K_2, K_1 \cup K_2)$ and $\sigma \in \mathcal{BT}_S(K_1 \cup K_2)$. Then:*

(a) $\Gamma_S(\mathbb{K})$ is contained in $\mathcal{N}_{S,\tau}$;

(b) Let $\sigma' = \mathbf{P}_S(\sigma; \mathbb{K})$, then:
$$d_{nav}(\sigma, \tau) = d_{nav}(\sigma, \sigma') + d_{nav}(\sigma', \tau)\ . \tag{64}$$

(c) Finally, $d_{nav}(\sigma, \mathcal{BT}_S(K_1, K_2, K_1 \cup K_2)) = \|\sigma\|_{\mathbb{K}}$.

Proof. For Lemma 11(a), let (σ, G) be an edge of $\Gamma_S(\mathbb{K})$. In particular, $\sigma \notin \mathcal{BT}_S(\mathbb{K})$ so that $K \in \mathcal{K}(\sigma, \tau)$ which produces $(\sigma, G) \in \mathcal{N}_{S,\tau}$ by definition.

For Lemma 11(b), let $\mathbf{p}$ be a maximal path in $\Gamma_S(\mathbb{K})$ emanating from σ. Then the endpoint of $\mathbf{p}$ is $\sigma' := \mathbf{P}_S(\sigma; \mathbb{K})$ by Theorem 1. Now apply Lemma 11(a) and Theorem 2 to extend $\mathbf{p}$ to a navigation path $\widetilde{\mathbf{p}}$ in $\mathcal{N}_{S,\tau}$ from σ to τ. Then:

$$d_{nav}(\sigma, \tau) = \ell(\,\widetilde{\mathbf{p}}\,) = \ell(\,\mathbf{p}\,) + d_{nav}(\sigma', \tau) = \|\sigma\|_{\mathbb{K}} + d_{nav}(\sigma', \tau)\ , \tag{65}$$

as required.

Finally, for Lemma 11(c), pick τ above to be a tree of $\mathcal{BT}_S(K_1, K_2, K_1 \cup K_2)$ with $d_{nav}(\sigma, \tau)$ minimal. By the construction above, $\sigma' \in \mathcal{BT}_S(K_1, K_2, K_1 \cup K_2)$ satisfies $d_{nav}(\sigma, \sigma') \leq d_{nav}(\sigma, \tau)$ while $\mathbf{p}$ is a navigation path from σ to σ'. Thus σ' must coincide with τ, and (65) reduces to the desired equality. ∎

Corollary 6. *For any bipartition $\{L, R\}$ of S and $\sigma \in \mathcal{BT}_S$, the navigation distance $d_{nav}(\sigma, \mathcal{BT}_S(L, R))$ can be computed in linear time,* $\mathrm{O}(|S|)$.

Proof. Similarly to the proof of Proposition 4, the crossing indices of σ-clusters with $\{L, R\}$ can be determined in $\mathrm{O}(|S|)$ time using Lemma 5 and by post order traversal of σ. Therefore, by Lemma 11 and Theorem 1, the quantity $d_{nav}(\sigma, \mathcal{BT}_S(S_L, S_R))$ can be computed in $\mathrm{O}(|S|)$ by a complete traversal of σ. ∎

Lemma 12. *For any bipartition $\{L, R\}$ of S and $\sigma \in \mathcal{BT}_S$, an NNI navigation path in $\Gamma_S(L, R)$ joining σ to $\mathcal{BT}_S(L, R)$ can be computed in* $\mathrm{O}(|S|)$ *time.*

Proof. As illustrated in Figure 5, since $\mathrm{Anc}(I, \sigma) \subseteq \mathcal{I}(\sigma; \{L, R\}) \cup \{S\}$ for any $I \in \mathcal{I}(\sigma; \{L, R\})$, the vertices and branches of σ associated with clusters in $\mathcal{I}(\sigma; \{L, R\}) \cup \{S\}$ defines a tree structure, containing all the information required to compute the navigation distance $d_{nav}(\sigma, \mathcal{BT}_S)(L, R) = \|\sigma\|_{\{L,R\}}$ (Lemma 11.(c)). Hence, one can construct an NNI navigation path by a complete post-order traversal of this tree structure as follows:

1. Set $k \leftarrow 0$ and $\sigma_0 \leftarrow \sigma$, and compute $\mathcal{I}(\sigma_0; \{L, R\})$.
2. Find a cluster $I_0 \in \mathcal{R}(\sigma_0; \{L, R\})$ by a post-order traversal of incompatible clusters $\mathcal{I}(\sigma_0; \{L, R\})$ of σ_0.
3. While $(\mathcal{I}(\sigma_k; \{L, R\}) \neq \varnothing)$
   (a) If $I_k \in \mathcal{R}(\sigma_k; \{L, R\})$ is Type 1, then, as illustrated in Figure 6(top), choice a grandchild $G_k \in \mathrm{Ch}(I_k, \sigma_k)$ such that $G_k{}^{-\sigma_k}, I_k{}^{-\sigma_k} \subseteq L$ or $G_k{}^{-\sigma_k}, I_k{}^{-\sigma_k} \subseteq R$, and set
   $$\sigma_{k+1} \leftarrow \mathrm{NNI}(\sigma_k, G_k), \quad \mathcal{I}(\sigma_{k+1}; \{L, R\}) \leftarrow \mathcal{I}(\sigma_k; \{L, R\}) \setminus \{I_k\}, \quad I_{k+1} \leftarrow \mathrm{Pr}(I_k, \sigma_k), \quad k \leftarrow k + 1.$$
   (b) If $I_k \in \mathcal{R}(\sigma_k; \{L, R\})$ is Type 2, then, as illustrated in Figure 6(bottom), choice $G_k \in \mathrm{Ch}(I_k, \sigma_k)$ and $G_{k+1} \in \mathrm{Ch}(I_k{}^{-\sigma_k}, \sigma_k)$ such that $G_k, G_{k+1} \subseteq L$ or $G_k, G_{k+1} \subseteq R$, and $G_{k+2} = G_k{}^{-\sigma_k} \cup G_{k+1}{}^{-\sigma_k}$; and set
   $$\sigma_{k+1} \leftarrow \mathrm{NNI}(\sigma_k, G_k), \quad \sigma_{k+2} \leftarrow \mathrm{NNI}(\sigma_{k+1}, G_{k+1}), \quad \sigma_{k+3} \leftarrow \mathrm{NNI}(\sigma_{k+2}, G_{k+2}),$$
   $$\mathcal{I}(\sigma_{k+3}; \{L, R\}) \leftarrow \mathcal{I}(\sigma_k; \{L, R\}) \setminus \{I_k, I_k{}^{-\sigma_k}\}, \quad I_{k+3} \leftarrow \mathrm{Pr}(I_k, \sigma_k), \quad k \leftarrow k + 3.$$
   (c) Otherwise (I_k and $I_k{}^{-\sigma_k}$ are Type 2 with $\mathrm{Ch}(I_k, \sigma_k) \bowtie \{L, R\}$ and $\mathrm{Ch}(I_k{}^{-\sigma_k}, \sigma_k) \not\bowtie \{L, R\}$), find a cluster $J_k \in \mathcal{R}(\sigma_k; \{L, R\})$ by a post-order traversal of incompatible clusters of the subtree of σ_k rooted at $I_k{}^{-\sigma_k}$, and set $I_k \leftarrow J_k$.
4. Return $(\sigma_k)_{k \in [0, \|\sigma\|_{\{L,R\}}]}$ as an NNI navigation path starting at σ and ending in $\mathcal{BT}_S(L, R)$.

As discussed in the proof of Proposition 4, all clusters of σ incompatible with $\{L, R\}$, i.e. $\mathcal{I}(\sigma; \{L, R\})$ in Step 1, can be determined in $O(|S|)$ time. Given $\mathcal{I}(\sigma; \{L, R\})$, a cluster $I \in \mathcal{R}(\sigma; \{L, R\})$, in Step 2, can be found in $O(|\mathcal{I}(\sigma; \{L, R\})|) \leq O(|S|)$ time by a post-order traversal of incompatible clusters of σ. Observe that the while loop terminates after at most $2|\mathcal{I}(\sigma; \{L, R\})|$ iterations after a complete traversal of the tree structure defined by $\mathcal{I}(\sigma; \{L, R\}) \cup \{S\}$ since $|\mathcal{I}(\sigma_k; \{L, R\})|$ decreases at least by one unit after every two consecutive iterations and a post-order subtree traversal in Step 3(c) is required only if the associated subtree is not explored yet. Hence, an NNI navigation path joining σ to $\mathcal{BT}_S(L, R)$ can be found by a complete post-order traversal of σ in $O(|S|)$ time. ■

The observation made in Lemma 11 is a good example of how the dual representation of d_{nav} — both in terms of paths in the NNI graph, and in terms of a closed-form formula quantifying inter-cluster incompatibility — offers a practical compromise between the heretofore separate traditional approaches to constructing dissimilarities on $\mathcal{BT}_S$, those of edge comparison and of estimation of edit distances. A particular application of this dual nature is the decomposability of d_{nav} (as defined in [33]):

Lemma 13 (Root Split Reduction)**.** *Fix $\tau \in \mathcal{BT}_S$ and denote $\{L, R\} := \mathrm{Ch}(S, \tau)$. Then for any $\sigma \in \mathcal{BT}_S$ one has:*

$$d_{nav}(\sigma, \tau) = d_{nav}(\sigma, \mathcal{BT}_S(L, R)) + d_{nav}\left(\sigma\big|_L, \tau\big|_L\right) + d_{nav}\left(\sigma\big|_R, \tau\big|_R\right). \tag{66}$$

Proof. By Lemma 11(2) it suffices to prove

$$d_{nav}(\mathbf{P}_S(\sigma; L, R), \tau) = d_{nav}\left(\sigma\big|_L, \tau\big|_L\right) + d_{nav}\left(\sigma\big|_R, \tau\big|_R\right). \tag{67}$$

By definition, $\mathcal{C}(\mathbf{P}_S(\sigma; L, R)) = \{S\} \cup \mathcal{C}\left(\sigma\big|_L\right) \cup \mathcal{C}\left(\sigma\big|_R\right)$ so it suffices to prove:

$$\sigma \in \mathcal{BT}_S(L, R) \Rightarrow d_{nav}(\sigma, \tau) = d_{nav}\left(\sigma\big|_L, \tau\big|_L\right) + d_{nav}\left(\sigma\big|_R, \tau\big|_R\right). \tag{68}$$

At this stage, however, observe that $\mathcal{C}\left(\sigma\big|_L\right)$ and $\mathcal{C}\left(\sigma\big|_R\right)$ together exhaust the list of of clusters of σ not equal to S, with the same holding *ab initio* for τ. This allows us to finish the proof by applying Theorem 2 separately in $\mathcal{BT}_L$ and $\mathcal{BT}_R$. ■

The root split reduction of the NNI navigation dissimilarity may be used for its efficient computation:

Corollary 7. *The NNI navigation dissimilarity d_{nav} on $\mathcal{BT}_S$ is computable in* $O(|S|^2)$ *time.*

Proof. Let $\sigma, \tau \in \mathcal{BT}_S$ and $\{L, R\} = \mathrm{Ch}(S, \tau)$. By the root split reduction above and the last corollary, $d_{nav}(\sigma, \tau)$ requires the computation of $d_{nav}(\sigma, \mathcal{BT}_S(\mathrm{Ch}(S, \tau)))$ at a cost of $O(|S|)$ time, plus the computation of the restrictions $\sigma\big|_L$ and $\sigma\big|_R$, each of which can be computed using post-order traversal of σ in $O(|S|)$ time. Hence, computing $d_{nav}(\sigma, \tau)$ requires a complete (depth-first) traversal of τ with each stage incurring at most a linear time cost in $|S|$. ■

Corollary 8. *An NNI navigation path joining $\sigma \in \mathcal{BT}_S$ to $\tau \in \mathcal{BT}_S$ can be computed in* $O(|S|^2)$ *time.*

Proof. Similar to the recursive expression of d_{nav} in Lemma 13, an NNI navigation path joining σ to τ can be found using the decomposability property within a divide-and-conquer approach as follows: first obtain an NNI navigation path from σ to $\mathcal{BT}_S(\mathrm{Ch}(S, \tau))$ in $O(|S|)$ (Lemma 12) and then find NNI navigation paths between subtrees. Hence, this requires the pre-order traversal of τ each of whose step costs $O(|S|)$. Thus, an NNI navigation path joining σ to τ can be recursively computed in $O(|S|^2)$ time, which completes the proof. ■

4.3. Properties of the Navigation Dissimilarity

Proposition 9. *The NNI navigation dissimilarity d_{nav} is positive definite and symmetric, but it is not a metric.*

Proof. That d_{nav} is positive definite follows directly from its definition. Lemma 7 proves it is symmetric and Corollary 2 with Figure 3 shows where the triangle inequality fails. ■

Lemma 14. *Let $\{L, R\}$ be a bipartition of S and $\sigma \in \mathcal{BT}_S$. Then we have the tight bound:*

$$d_{nav}(\sigma, \mathcal{BT}_S(L, R)) \leq |S| + \min(|L|, |R|) - 3. \tag{69}$$

Proof. Denote $\mathbb{S} = \{L, R\}$. For any $\sigma \in \mathcal{BT}_S$ and $I \in \mathcal{C}(\sigma)$ observe that (i) $[\![\mathrm{Ch}(I, \sigma) \mid \mathbb{S}]\!] = 0$ if I is a singleton or $|I| = 2$, and (ii) otherwise for larger clusters $[\![\mathrm{Ch}(I, \sigma) \mid \mathbb{S}]\!]$ equals 3 or 1 only if, respectively, both clusters or only one cluster of $\mathrm{Ch}(I, \sigma)$ are incompatible with $\mathbb{S}$. Since there are at least $|S| + 1$ clusters of the first kind, there are at most $|S| - 2$ clusters of the second kind. Thus, applying Lemma 11 and Theorem 1 we have

$$d_{nav}(\sigma, \mathcal{BT}_S(L, R)) \leq (|S| - 2) + |\mathcal{X}|, \tag{70}$$

where $\mathcal{X}$ is the set of all $I \in \mathcal{C}(\sigma)$ both of whose children are incompatible with $\mathbb{S}$. For each $I \in \mathcal{X}$ both $I \cap L$ and $I \cap R$ are non-singleton clusters of $\sigma\big|_L$ and $\sigma\big|_R$, respectively (each child of I intersects each of L, R). Suppose now that $I, J \in \mathcal{X}$ are distinct. There are two cases, without loss of generality:

- If $I \cap J = \varnothing$, then $I \cap L \neq J \cap L$ (and similarly for R);
- If $I \subsetneq J$, then J has a child I' disjoint from I, and this child must intersect L. Hence, $I \cap L \subsetneq J \cap L$.

We conclude that the map $I \mapsto I \cap L$ (respectively $I \cap R$) of $\mathcal{X}$ to $\mathcal{C}\left(\sigma\big|_L\right)$ (resp. to $\mathcal{C}\left(\sigma\big|_R\right)$) is injective, and has no singleton clusters in its image. Thus, $|\mathcal{X}| \leq \min(|L| - 1, |R| - 1)$, proving the desired inequality.

The example $\sigma, \tau \in \mathcal{BT}_{[n]}$ in Figure 4 with $\{L, R\} = \mathrm{Ch}([n], \tau) = \{\{1\}, \{2, 3, \ldots, n\}\}$ shows that the upper bound in (69) is tight (where $d_{nav}(\sigma, \mathcal{BT}_S(L, R)) = n - 2$). ■

Proposition 10. $\mathrm{diam}(\mathcal{BT}_S, d_{nav}) = \frac{1}{2}(|S| - 1)(|S| - 2)$.

Proof. We proceed by induction over $|S|$, with the base case $|S| = 2$ satisfying $|\mathcal{BT}_S| = 1$. The formula then holds trivially, as $d_{nav} = 0$.

For the induction step assume $|S| \geq 3$ and that $\sigma, \tau \in \mathcal{BT}_S$ satisfy $d_{nav}\left(\sigma\big|_K, \tau\big|_K\right) \leq \frac{1}{2}(|K| - 1)(|K| - 2)$ for every $K \in \mathrm{Ch}(S, \tau) = \{L, R\}$.

Let $\mu = \min(|L|, |R|)$, and note that $|L|\,|R| = \mu(|S| - \mu)$. We now apply the root split reduction (Lemma 13):

$$d_{nav}(\sigma,\tau) = \underbrace{d_{nav}(\sigma, \mathcal{BT}_S(L,R))}_{\substack{\text{by Lemma 14} \\ \leq |S|+\mu-3}} + \underbrace{d_{nav}\left(\sigma|_L, \tau|_L\right)}_{\substack{\text{by induction} \\ \leq \frac{1}{2}(|L|-1)(|L|-2)}} + \underbrace{d_{nav}\left(\sigma|_R, \tau|_R\right)}_{\substack{\text{by induction} \\ \leq \frac{1}{2}(|R|-1)(|R|-2)}}, \tag{71}$$

$$\leq \tfrac{1}{2}(|S|-1)(|S|-2) + \underbrace{(1-\mu)(|S|-\mu-2)}_{\text{non-positive whenever } |S|\geq 3}, \tag{72}$$

$$\leq \tfrac{1}{2}(|S|-1)(|S|-2). \tag{73}$$

Finally, note that the trees in Figure 4 realize this bound on the diameter. ■

4.4. Relations with Other Tree Measures

Like d_{CM} (Proposition 7), d_{nav} is tightly bounded in terms of d_{RF} as follows:

Proposition 11. *Over $\mathcal{BT}_S$ one has $d_{RF} \leq d_{nav} \leq \frac{1}{2}d_{RF}^2 + \frac{1}{2}d_{RF}$ and both bounds are tight.*

Proof. Since d_{nav} is realized by paths in the NNI graph we have $d_{NNI} \leq d_{nav}$. The lower bound then follows from $d_{RF} \leq d_{NNI}$ (Corollary 1). The bound is tight because

$$d_{RF}(\sigma,\tau) = 1 \Leftrightarrow d_{NNI}(\sigma,\tau) = 1 \Leftrightarrow d_{nav}(\sigma,\tau) = 1. \tag{74}$$

For the upper bound we argue by induction over $|S|$, keeping in mind that for $|S| = 2$ the result holds trivially. Suppose $|S| \geq 3$. Now, if σ and τ have no common nontrivial clusters then $d_{RF}(\sigma,\tau) = |S| - 2$ and the result follows from Proposition 10. Otherwise, let $I \in \mathcal{C}(\sigma) \cap \mathcal{C}(\tau)$ be a nontrivial cluster and consider the tree σ' obtained from σ by replacing the branch $\sigma|_I$ with the branch $\tau|_I$.

By theorem Theorem 2 and by the definition of d_{RF}, respectively, we have:

$$d_{nav}(\sigma,\tau) = d_{nav}(\sigma,\sigma') + d_{nav}(\sigma',\tau) \tag{75}$$

$$d_{RF}(\sigma,\tau) = d_{RF}(\sigma,\sigma') + d_{RF}(\sigma',\tau) \tag{76}$$

Let $\alpha = d_{RF}(\sigma,\sigma')$ and $\beta = d_{RF}(\sigma',\tau)$. Since $d_{nav}(\sigma,\sigma') = d_{nav}\left(\sigma|_I, \sigma'|_I\right)$ we may apply the induction hypothesis in $\mathcal{BT}_I$ to conclude $d_{nav}(\sigma,\sigma') \leq \frac{1}{2}\alpha(\alpha+1)$. By pruning the trees σ' and τ at cluster I we may apply the induction hypothesis in $\mathcal{BT}_{\bar{S}}$, where $\bar{S}$ is the result of contracting I to a single vertex, to conclude that $d_{nav}(\sigma',\tau) \leq \frac{1}{2}\beta(\beta+1)$. It then follows that:

$$d_{nav}(\sigma,\tau) \leq \tfrac{1}{2}\alpha(\alpha+1) + \tfrac{1}{2}\beta(\beta+1) \leq \tfrac{1}{2}(\alpha+\beta)(\alpha+\beta+1) = \tfrac{1}{2}d_{RF}(\sigma,\tau)(d_{RF}(\sigma,\tau)+1). \tag{77}$$

Proposition 10 ensures this bound is tight. ■

Proposition 12. *Over $\mathcal{BT}_S$ one has $d_{nav}(\sigma,\tau) \leq \frac{3}{2}d_{CM}(\sigma,\tau)$.*

Proof. Consider the closed form expression of d_{nav} (61) in terms of crossing indices. Since the trivial clusters are compatible with any subset of S, it will suffices to verify that, for each $I \in \mathcal{C}(\sigma)$ and $J \in \mathcal{C}(\tau)$, one has:

$$[\![\mathrm{Ch}(I,\sigma) \,|\, \mathrm{Ch}(J,\tau)]\!] \leq \tfrac{3}{2} \sum_{A \in \mathrm{Ch}(I,\sigma)} \sum_{B \in \mathrm{Ch}(J,\tau)} \mathbb{1}(A \not\pitchfork B). \tag{78}$$

This verification is straightforward. ■

The overall ordering of tree dissimilarities in Corollary 1, Proposition 8 and Proposition 12 can be combined as:

Theorem 3. *For non-degenerate hierarchies,*

$$\frac{2}{3}d_{RF} \leq \frac{2}{3}d_{NNI} \leq \frac{2}{3}d_{nav} \leq d_{CM} \leq d_{CC}. \tag{79}$$

Finally, we remark that the NNI navigation dissimilarity d_{nav} (Definition 19) can be generalized to a pair of trees, σ and τ, in $\mathcal{T}_S$ as

$$d_{nav}(\sigma, \tau) = \frac{1}{2}\left(\left\|\mathbf{S}(\sigma, \tau)\right\|_1 + \left\|\mathbf{S}(\tau, \sigma)\right\|_1\right), \tag{80}$$

which is non-negative and symmetric. For non-degenerate trees $\sigma, \tau \in \mathcal{BT}_S$ one has $\mathbf{S}(\sigma, \tau) = \mathbf{S}(\tau, \sigma)^{\mathrm{T}}$ (which is evident from (62) and Lemma 7), so that d_{nav} in (80) simplifies back to (61).[9] Although the closed form expression of d_{nav} in Theorem 2 enables the generalization of d_{nav} to degenerate trees as above, the notion of NNI moves (Definition 4) is generally not valid in $\mathcal{T}_S$.

As for non-degenerate trees in Proposition 12, the generalized d_{nav} in $\mathcal{T}_S$ can be bounded above by d_{CM} as follows:

Proposition 13. *Over $\mathcal{T}_S$ one has $d_{nav} \leq \left(\frac{1}{8}|S|^2 + \frac{1}{4}|S|\right) d_{CM}$.*

Proof. Note that the number of nontrivial children of a cluster in a tree can be at most $\frac{1}{2}|S|$. Hence one can verify the result following similar steps as in the proof of Proposition 12. ■

5. Discussion and Statistical Analysis

5.1. Consensus Models and Median Trees

Let us recall a definition : a *median tree* of a set of sample trees is a tree whose sum of distances to the sample trees is minimum. Although the notion of a median tree is simple and well-defined, finding a median tree of a set of trees is generally a hard combinatorial problem. On the other hand, a consensus model of a set of sample trees is a computationally efficient tool to identify common structures of sample trees. In particular, a remark relating d_{CM} and d_{nav} to commonly used consensus models of a set of trees and their median tree(s) is:

Proposition 14. *Both the strict and loose consensus trees, T_* and T^*, of any set of trees T in $\mathcal{T}_S$ (Definition 3) are median trees with respect to both the crossing (d_{CM}) and navigation (d_{nav}) dissimilarities. In fact, for any $d \in \{d_{CM}, d_{nav}\}$ one has:*

$$\sum_{\tau \in T} d(\tau, T_*) = \sum_{\tau \in T} d(\tau, T^*) = 0. \tag{81}$$

Proof. By Definition 3, both strict and loose consensus trees only contain clusters that are compatible with the clusters of every tree in T, and the loose consensus tree is the finest median tree containing only clusters from the sample trees. Thus, the result follows for both d_{CM} and d_{nav} due their relation in Proposition 13. ■

5.2. Sample Distribution of Dissimilarities

To compare their discriminative power, we use a standard statistical analysis of empirical distributions of different tree measures. The shape of the distribution of a tree measure tells how informative it is; for example, a highly concentrated distribution means that the associated tree measure behaves like the discrete metric[10] as in the case of the Robinson-Foulds distance — see Figure 7. Finding a closed form expression for the distribution of a tree measure is a hard problem, and so extensive numerical simulations are generally applied to obtain its sample distribution. In particular, using the uniform and Yule model [34] for generating random trees, we compute the empirical distributions of d_{RF}, d_{MS}, d_{CC}, d_{CM}, and d_{nav} as illustrated in Figure 7.[11] Moreover, in Table 1 we present two commonly used statistical measures, skewness and kurtosis, for describing the shapes of the probability distributions of all these tree measures. Here, recall that the skewness of a probability distribution measures its tendency on one side of the mean, and the concept of kurtosis measures the peakedness of the distribution [35]. In addition to their computational advantage over d_{MS}, as illustrated in both Figure 7 and Table 1, like d_{MS}, our tree measures, d_{CC}, d_{CM} and d_{nav}, are significantly more discriminative, with wider ranges of values and symmetry, than d_{RF}.

[9] $\mathbf{A}^{\mathrm{T}}$ is the transpose of matrix $\mathbf{A}$.

[10] The discrete metric $d : X \times X \to \mathbb{R}_{\geq 0}$ on a set X is defined as for any $x \neq y \in X$ $d(x, x) = 0$ and $d(x, y) = 1$.

[11] In our numerical simulations for any chosen tree measure we observe the same pattern of sample distribution for different numbers of leaves, and so here we only include results for $\mathcal{BT}_{[25]}$.

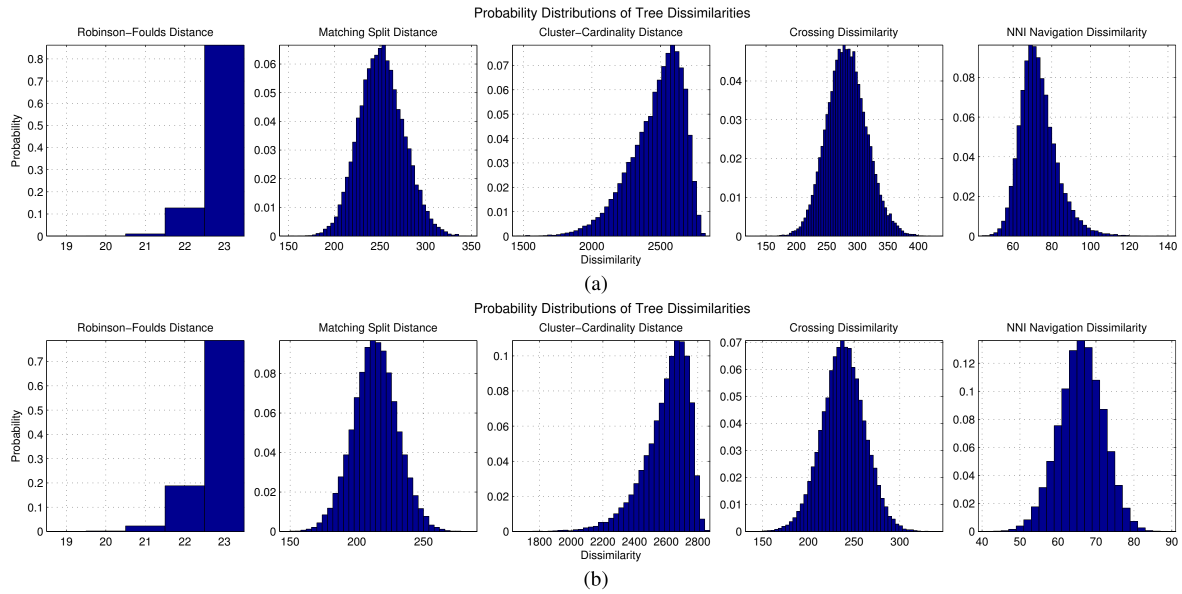

Figure 7. Empirical distribution of tree dissimilarities in $\mathcal{BT}_{[25]}$: (from left to right) the Robinson-Foulds distance d_{RF} (14), the matching split distance d_{MS} (Def. 7), the cluster-cardinality distance d_{CC} (18), the crossing dissimilarity d_{CM} (Def. 10), and the NNI navigation dissimilarity d_{nav} (Def. 19). 100000 sample hierarchies are generated using (a) the uniform and (b)Yule model [34]. The resolutions of histograms of tree measures, from left to right, are 1, 4, 32, 4, 2 unit(s), respectively.

Table 1. Skewness and Kurtosis Values for the Distributions of Tree Measures in $\mathcal{BT}_{[25]}$

	Skewness		Kurtosis	
	Uniform	Yule	Uniform	Yule
d_{RF} (14)	−2.6162	−2.0740	9.8609	7.3998
d_{MS} (Def. 7)	0.1293	−0.0117	3.0060	3.1136
d_{CC} (18)	−0.9294	−1.2507	3.8601	5.2724
d_{CM} (Def. 10)	0.1390	−0.0405	3.1275	3.2103
d_{nav} (Def. 19)	0.8809	−0.1195	4.8707	3.0746

6. Conclusion

This paper presents three new tree measures for efficient discriminative comparison of trees. First, using the well known relation between trees and ultrametrics, the cluster-cardinality metric d_{CC} is constructed as the pullback of matrix norms along an embedding of trees into the space of matrices. Second, we present the crossing dissimilarity d_{CM} that counts the pairwise incompatibilities of trees. Third, the NNI navigation dissimilarity d_{nav} while presented in closed form is constructed as the length of a navigation path in the space of trees.

All of our dissimilarities can be computed in $O(n^2)$ with the number of leaves n, and they generalize to degenerate trees as well. Moreover, we provide a closed form expression for each proposed dissimilarity and present an ordering relation between these tree dissimilarities and related tree metrics in the literature (Theorem 3). Our numerical studies, summarized in Figure 7, suggest that the proposed tree measures are significantly more informative and discriminative than the Robinson-Foulds distance d_{RF}, while maintaining a computational advantage over other distances such as the matching-split distance [3, 4].

Finally, the system of projector graphs (Theorem 1) and navigation graphs (Theorem 2) seems to play a fundamental role in the geometry of the NNI graph, realizing many of the intuitive desiderata of tree dissimilarity measures that have accumulated in the literature over the years. Consequently, NNI navigation paths are likely of some significance for consensus/average models or statistical analysis of trees.

Acknowledgements

This work was funded in part by the Air Force Office of Science Research under the MURI FA9550-10-1-0567.

Appendix A. Proofs

Appendix A.1. Proof of Lemma 1

Proof. Sufficiency is directly evident from Definition 4 because the cluster sets of a pair of nondegenerate hierarchies differ exactly by one cluster if and only if they are NNI-adjacent. To verify necessity, let the move (σ, P), $P \in \mathcal{G}(\sigma)$ join σ to τ, and $R = P^{-\sigma}$ and $Q = \mathrm{Pr}^2(P, \sigma) \setminus \mathrm{Pr}(P, \sigma)$. By Definition 4, $\{\mathrm{Pr}(P, \sigma)\} = \{P \cup R\} = \mathcal{C}(\sigma) \setminus \mathcal{C}(\tau)$ and $\left\{\mathrm{Pr}^2(P, \sigma) \setminus P\right\} = \{R \cup Q\} = \mathcal{C}(\tau) \setminus \mathcal{C}(\sigma)$. Further, (P, R, Q) is the only ordered triple of common clusters of σ and τ with the property that $\{P \cup R\} = \mathcal{C}(\sigma) \setminus \mathcal{C}(\tau)$ and $\{R \cup Q\} = \mathcal{C}(\tau) \setminus \mathcal{C}(\sigma)$ since the cluster sets of any two NNI-adjacent hierarchies differ exactly by one element. ■

Appendix A.2. Proof of Lemma 2

Proof. To observe that $res_K(\mathcal{BT}_S) \supseteq \mathcal{BT}_K$, consider any two nondegenerate trees $\sigma \in \mathcal{BT}_K$ and $\gamma \in \mathcal{BT}_{S \setminus K}$, and let $\tau \in \mathcal{BT}_S$ be the nondegenerate tree with cluster set $\mathcal{C}(\tau) = \mathcal{C}(\sigma) \cup \{S\} \cup \mathcal{C}(\gamma)$. Note that $\mathrm{Ch}(S, \tau) = \{K, S \setminus K\}$. Hence, we have from Remark 2 that $\sigma = res_K(\tau)$. To prove that $res_K(\mathcal{BT}_S) \subseteq \mathcal{BT}_K$, let $\tau \in \mathcal{BT}_S$ and $I \in \mathcal{C}(\tau)$ with the property that $|I \cap K| \geq 2$. Note that $I \cap K$ is an interior cluster of $\tau\big|_K$. We shall show that the cluster $I \cap K \in \mathcal{C}\left(\tau\big|_K\right)$ always admits a bipartition in $\tau\big|_K$. That is to say, there exist a cluster $A \in \mathcal{C}(\tau)$ with children $\{A_L, A_R\} = \mathrm{Ch}(A, \tau)$ such that $A \cap K = I \cap K$ and $A_L \cap K \neq \varnothing$ and $A_R \cap K \neq \varnothing$. Hence, $\mathrm{Ch}\left(I \cap K, \tau\big|_K\right) = \{A_L \cap K, A_R \cap K\}$. Now observe that either $I_L \cap K \neq \varnothing$ and $I_R \cap K \neq \varnothing$ for $\{I_L, I_R\} = \mathrm{Ch}(I, \tau)$, or there exists one and only one descendant $D \in \mathrm{Des}(I, \tau)$ with $\{D_L, D_R\} = \mathrm{Ch}(D, \tau)$ such that $I \cap K = D \cap K$ and $D_L \cap K \neq \varnothing$ and $D_R \cap K \neq \varnothing$. Thus, all the interior clusters of $\tau\big|_K$ have exactly two children, which completes the proof. ■

Appendix A.3. Proof of Lemma 3

Proof. The proof of the sufficiency for being an ultrametric is as follows. Positive definiteness and symmetry of d_τ are evident from (15) and Lemma 3.(a)-(b). To show the strong triangle inequality, let $i \neq j \neq k \in S$ and $I = (i \wedge j)_\tau$, and so $d_\tau(i, j) = h_\tau(I)$. Accordingly, let $\{I_i, I_j\} \subseteq \mathrm{Ch}(I, \tau)$ with the property that $i \in I_i$ and $j \in I_j$.

If $k \in I$, without loss of generality, let $k \in I_i$, and so $k \notin I_j$. Then, using (15) and Lemma 3.(a), one can verify that $d_\tau(i, k) \leq h_\tau(I_i) \leq h_\tau(I)$ and $d_\tau(j, k) = h_\tau(I)$ because $(i \wedge k)_\tau \subseteq I_i$ and $(j \wedge k)_\tau = I$. Also note that if neither $k \in I_i$ nor $k \in I_j$ (but still $k \in I$), then $d_\tau(i, k) = d_\tau(j, k) = h_\tau(I)$ since $(i \wedge k)_\tau = (j \wedge k)_\tau = I$. Similarly, if $k \notin I$, then $d_\tau(i, k) \geq h_\tau(I)$ and $d_\tau(j, k) \geq h_\tau(I)$ because only some ancestors of I in τ might contain all i, j, k. Therefore, overall, one always has $d_\tau(i, j) \leq \max(d_\tau(i, k), d_\tau(k, j))$, which completes the proof of the sufficiency.

Let us continue with the necessity for being an ultrametric. Note that Lemma 3.(b) directly follows from positive definiteness of d_τ. Let $I \in \mathcal{C}(\tau) \setminus \{S\}$ be any non-singleton cluster of τ and $i \neq j \in I$ with the property that $(i \wedge j)_\tau = I$. For any $k \in I^{-\tau}$, we always have $(i \wedge k)_\tau = (j \wedge k)_\tau = \mathrm{Pr}(I, \tau)$. Now, using the ultrametric inequality of d_τ, one deduces Lemma 3.(a) from

$$h_\tau(I) = d_\tau(i, j) \leq \max(d_\tau(i, k), d_\tau(j, k)) = h_\tau(\mathrm{Pr}(I, \tau)), \tag{A.1}$$

which completes the proof. ■

References

[1] D. F. Robinson, L. R. Foulds, Comparison of phylogenetic trees, Mathematical Biosciences 53 (1-2) (1981) 131 – 147.
[2] W. H. E. Day, Optimal algorithms for comparing trees with labeled leaves, Journal of Classification 2 (1985) 7–28.
[3] D. Bogdanowicz, K. Giaro, Matching split distance for unrooted binary phylogenetic trees, Computational Biology and Bioinformatics, IEEE/ACM Transactions on 9 (1) (2012) 150–160.
[4] Y. Lin, V. Rajan, B. Moret, A metric for phylogenetic trees based on matching, Computational Biology and Bioinformatics, IEEE/ACM Transactions on 9 (4) (2012) 1014–1022.
[5] D. Robinson, Comparison of labeled trees with valency three, Journal of Combinatorial Theory, Series B 11 (2) (1971) 105 – 119.
[6] G. Moore, M. Goodman, J. Barnabas, An iterative approach from the standpoint of the additive hypothesis to the dendrogram problem posed by molecular data sets, Journal of Theoretical Biology 38 (3) (1973) 423 – 457.

[7] B. L. Allen, M. Steel, Subtree transfer operations and their induced metrics on evolutionary trees, Annals of Combinatorics 5 (2001) 1–15.
[8] J. Felsenstein, Inferring Phylogenies, Sinauer Associates, Suderland, USA, 2004.
[9] B. DasGupta, X. He, T. Jiang, M. Li, J. Tromp, L. Zhang, On distances between phylogenetic trees, in: Proceedings of the eighth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 1997, pp. 427–436.
[10] O. Arslan, D. P. Guralnik, D. E. Koditschek, Hierarchically clustered navigation of distinct Euclidean particles, in: Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on, 2012, pp. 946–953.
[11] O. Arslan, D. P. Guralnik, D. E. Koditschek, Navigation of distinct Euclidean particles via hierarchical clustering, in: Algorithmic Foundations of Robotics XI, Vol. 107 of Springer Tracts in Advanced Robotics, 2015, pp. 19–36.
[12] O. Arslan, D. P. Guralnik, D. E. Koditschek, Coordinated robot navigation via hierarchical clustering, IEEE Transactions on Robotics 32 (2) (2016) 352–371.
[13] N. Ayanian, V. Kumar, D. Koditschek, Synthesis of controllers to create, maintain, and reconfigure robot formations with communication constraints, in: Robotics Research, Vol. 70 of Springer Tracts in Advanced Robotics, Springer Berlin Heidelberg, 2011, pp. 625–642.
[14] R. R. Burridge, A. A. Rizzi, D. E. Koditschek, Sequential composition of dynamically dexterous robot behaviors, The International Journal of Robotics Research 18 (6) (1999) 535–555.
[15] G. Carlsson, F. Mémoli, Characterization, stability and convergence of hierarchical clustering methods, Journal of Machine Learning Research 11 (2010) 1425–1470.
[16] M. Li, J. Tromp, L. Zhang, On the nearest neighbour interchange distance between evolutionary trees, Journal of Theoretical Biology (1996) 463–467.
[17] K. CulikII, D. Wood, A note on some tree similarity measures, Information Processing Letters 15 (1) (1982) 39 – 42.
[18] E. K. Brown, W. H. E. Day, A computationally efficient approximation to the nearest neighbor interchange metric, Journal of Classification 1 (1984) 93–124.
[19] D. Bryant, A classification of consensus methods for phylogenetics, DIMACS series in Discrete Mathematics and Theoretical Computer Science 61 (2003) 163–184.
[20] T. Margush, F. R. McMorris, Consensus n-trees, Bulletin of Mathematical Biology 43 (2) (1981) 239–244.
[21] J.-P. Barthélemy, F. McMorris, The median procedure for n-trees, Journal of Classification 3 (2) (1986) 329–334.
[22] F. James Rohlf, Consensus indices for comparing classifications, Mathematical Biosciences 59 (1) (1982) 131–144.
[23] K. Bremer, Combinable component consensus, Cladistics 6 (4) (1990) 369–372.
[24] L. J. Billera, S. P. Holmes, K. Vogtmann, Geometry of the space of phylogenetic trees, Advances in Applied Mathematics 27 (4) (2001) 733 – 767.
[25] A. Schrijver, Combinatorial optimization: polyhedra and efficiency, Vol. 24, Springer, 2003.
[26] K. Vogtmann, Geodesics in the space of trees, accessed May 22, 2015. (2007).
URL `www.math.cornell.edu/~vogtmann/papers/TreeGeodesicss`
[27] M. Dale, J. Moon, The permuted analogues of three Catalan sets, Journal of Statistical Planning and Inference 34 (1) (1993) 75–87.
[28] D. Callan, A combinatorial survey of identities for the double factorial, arXiv preprint arXiv:0906.1317.
[29] A. K. Jain, R. C. Dubes, Algorithms for clustering data, Prentice-Hall, Inc., 1988.
[30] R. Rammal, G. Toulouse, M. A. Virasoro, Ultrametricity for physicists, Reviews of Modern Physics 58 (3) (1986) 765.
[31] D. D. Sleator, R. E. Tarjan, W. P. Thurston, Rotation distance, triangulations, and hyperbolic geometry, in: Proceedings of the eighteenth annual ACM symposium on Theory of computing, STOC '86, 1986, pp. 122–135.
[32] D. Epstein, M. Paterson, J. Cannon, D. Holt, S. Levy, W. P. Thurston, Word processing in groups, AK Peters, Ltd., 1992.
[33] M. Waterman, T. Smith, On the similarity of dendrograms, Journal of Theoretical Biology 73 (4) (1978) 789 – 800.
[34] C. Semple, M. Steel, Phylogenetics, Vol. 24, Oxford University Press, 2003.
[35] J. A. Rice, Mathematical statistics and data analysis, Cengage Learning, 2007.