Discriminative Measures for Comparison of Phylogenetic Trees

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

Abstract—Efficient and informative comparison of trees is a common essential interest of both computational biology and pattern classification. In this paper, we introduce a novel dissimilarity measure on non-degenerate hierarchies (rooted binary trees), called *the NNI navigation distance*, that counts the steps along the trajectory of a discrete dynamical system defined over the Nearest Neighbor Interchange(NNI) graph of binary hierarchies. The NNI navigation distance has a unique unifying nature of combining both edge comparison methods and edit operations for comparison of trees and is an efficient approximation to the (NP-hard) NNI distance. It is given by a closed form expression which simply generalizes to nondegenerate hierarchies as well. A "relaxation" on the closed form of the NNI navigation distance results a simpler dissimilarity measure on all trees, named *the crossing dissimilarity*, counts pairwise cluster incompatibilities of trees. Both of our dissimilarity measures on nondegenerate hierarchies are positive definite (vanishes only between identical trees) and symmetric but are not a true metric because they do not satisfy the triangle inequality. Although they are not true metrics, they are both linearly bounded below by the widely used Robinson-Foulds metric and above by a new tree metric, called *the cluster-cardinality* distance — the pullback metric of a matrix norm along an embedding of hierarchies into the space of matrices. All of these proposed tree measures can be efficiently computed in time $O(n^2)$ in the number of leaves, n.

Index Terms—Evolutionary trees, Nearest Neighbor Interchange, Comparison of Classifications, Tree Space, Tree Metric, Robinson-Foulds Distance, Consensus Tree, Median Tree, Rotation Distance, Diagonal-Flip Distance.

1 INTRODUCTION

A fundamental classification problem common to both computational biology and engineering is the efficient and informative comparison of hierarchical structures. Typically, in bioinformatics settings, these take the form of phylogenetic trees representing evolutionary relationships within a set of taxa. Typically, in pattern recognition or data mining settings, hierarchical trees encode nested sequences of groupings of a set of observations. Dissimilarity between combinatorial trees has been computed in the past literature largely by recourse to one of two approaches: either comparing edges or counting edit distances. Representing the former approach, a widely used tree metric is the Robinson-Foulds (RF) or symmetric difference distance, d_{RF}, [1] whose count of the disparate edges between trees requires linear time, $\mathrm{O}(n)$, 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, $\mathrm{O}\left(n^{2.5}\log n\right)$. 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].

In this paper we introduce a new dissimilarity measure on nondegenerate hierarchies, the *NNI navigation distance* d_{Nav}, that can be computed in time $O(n^2)$ for trees with n leaves. Although d_{Nav} is positive definite (vanishes only between identical trees) and symmetric, it is not a true metric because it does not satisfy the triangle inequality. We provide tight bounds by showing it is at least linear and at most quadratic,

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

with respect to the Robinson-Foulds distance. Our measure, a closed form expression comprising a weighted count of special incompatible clusters of two trees, is derived as a kind of discrete "path integral". Namely, it counts the steps along the trajectory of a discrete dynamical system defined over the NNI-graph of tree space that seeks to reduce the number of incompatible clusters level by level at each chosen NNI operation. In this sense, d_{Nav} seems distinguished in the large and still rapidly growing tree distance literature by offering a compromise between the two traditional approaches. On the one hand, d_{Nav} heuristically (but efficiently) approximates d_{NNI}, while on the other, it is designed to be sensitive to the edge (i.e. tree clusters - as opposed to NNI-graph edges) properties as well.

We find it useful to introduce a "relaxation" of d_{Nav} yielding a simpler dissimilarity measure, the crossing dissimilarity d_{CM}. This function simply counts all the pairwise cluster incompatibilities between two trees, hence it still incurs a $\mathrm{O}\left(n^2\right)$ computational cost. Like d_{Nav}, d_{CM} is a positive definite, symmetric function of nondegenerate trees whose failure to respect the triangle inequality is mitigated by tight upper (quadratic) and

* *Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104*
This work was funded in part by the Air Force Office of Science Research under the MURI FA9550-10-1-0567.

lower (linear) bounds respecting d_{RF},

$$d_{RF} \leq d_{CM} \leq d_{RF}^2. \quad (2)$$

Because of their close relationship, d_{CM} defines a linear bound on d_{Nav},

$$d_{Nav} \leq \frac{3}{2} d_{CM}. \quad (3)$$

But unlike d_{Nav}, d_{CM} is simple enough to work with that it can be linearly bounded from above by a true metric. Namely, we introduce yet a new tree metric, *the cluster cardinality distance*, d_{CC}, — the pullback of a matrix norm along an embedding of hierarchies into the space of matrices, which is computable in $\mathrm{O}(n^2)$ time. Thus, although neither d_{Nav} nor d_{CM} are true metrics, in addition to the RF bounds just stated, they can also be linearly bounded above by d_{CC},

$$\frac{2}{3} d_{RF} \leq \frac{2}{3} d_{Nav} \leq d_{CM} \leq d_{CC}. \quad (4)$$

We have mentioned 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 [10]–[12], d_{Nav} has some resemblance to an early NNI graph navigation algorithm, d_{ra} [12] which used a divide-and-conquer approach with a balancing strategy to achieve an $\mathrm{O}(n \log n)$ 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 [10]–[12], does not admit of a closed form expression (and, likely in consequence, enjoys no reported metric upper bound). Perhaps more significantly for potential applications, d_{Nav} and our related measures are, like d_{MS} [3], sensitive to the tree depth at which disparity occurs. In the definition of d_{Nav} edges closer to the root have greater influence on the total cost. As pointed out in [3], this is intuitively consistent with many agglomerative hierarchical clustering and distance-based phylogenetic reconstruction methods whose ”bottom-up” nature generally implies that cluster merging cost, a measure of cluster dissimilarity, increases at each iteration towards the root.

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 [13]. For instance, the majority rule tree [14] of a set of trees is a median tree respecting the RF distance and provides statistics on the central tendency of trees [15]. 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 strict [16] and loose (semi-strict) [17] consensus trees behave as median trees with respect to both our dissimilarities. In fact, the loose consensus tree is the maximal (finest) median tree for both d_{Nav} and d_{CM}.

A final observation of significant interest in some application settings is that the computation of d_{Nav} derives from an exact path in tree space that can be explicitly computed with the same $O(n^2)$ computational cost. Such paths are motivated by independent problems related to particle swarm coordination [18], [19] , but may likely hold value for researchers interested in tree consensus and averaging as well.

2 Background & Notation

We now introduce our basic notation used throughout the paper and recall several standard notions of hierarchies, such as cluster compatibility, hierarchical relations of clusters and tree operations, from a set theoretical perspective.

2.1 Hierarchies

A hierarchy τ over a fixed finite index set S, say $S = [n] := \{1, 2, \ldots, n\}$, uniquely determines (and henceforth will be interchangeably identified with) a rooted semi-labelled tree : a directed acyclic graph $G_\tau = (V_\tau, E_\tau)$, whose leaves, vertices of degree one, are bijectively labelled by S and interior vertices have out-degree at least two, and all the edges in E_τ are directed away from a vertex, designated to be the *root*, with the property that all of its other vertices are reachable from the root through a directed path in τ [20]. The *cluster* $\mathcal{C}(v)$ of a vertex $v \in V_\tau$ is defined to be the set of leaves reachable from v by a directed path in τ. Singletons and the root cluster S belong to all trees, so we refer to them as the *trivial clusters*. The cluster set $\mathcal{C}(\tau)$ of τ is defined to be the set of all its vertex clusters,

$$\mathcal{C}(\tau) := \left\{ \mathcal{C}(v) \,\middle|\, v \in V_\tau \right\} \subseteq \mathcal{P}(S), \quad (5)$$

where $\mathcal{P}(S)$ denotes the power set of S. It is convenient to have $\mathcal{C}_{int}(\tau)$ denote the nontrivial clusters of τ,

$$\mathcal{C}_{int}(\tau) := \left\{ I \in \mathcal{C}(\tau) \setminus \{S\} \,\middle|\, |I| \geq 2 \right\}. \quad (6)$$

2.1.1 Cluster Compatibility

Definition 1 ([8], [21]) *Let A, B be finite sets, then A and B are said to be* compatible, *$A \bowtie B$, if they are disjoint or one is a subset of the other,*

$$A \cap B = \emptyset \ \vee \ A \subseteq B \ \vee \ B \subseteq A. \quad (7)$$

If A and B are incompatible, $A \not\bowtie B$, then they are said to cross.

If $\mathcal{A}$ and $\mathcal{B}$ are subsets of $\mathcal{P}(S)$, we say that $\mathcal{A}$ and $\mathcal{B}$ are compatible if $A \bowtie B$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. In particular, a cluster $I \in \mathcal{P}(S)$ is said to be compatible with a tree $\tau \in \mathcal{T}_S$ if $\{I\} \bowtie \mathcal{C}(\tau)$, and two trees $\sigma, \tau \in \mathcal{T}_S$ are compatible if $\mathcal{C}(\sigma) \bowtie \mathcal{C}(\tau)$.

It is easy to observe that any two elements of $\mathcal{C}(\tau)$ are compatible for any tree τ, which motivates the following definition:

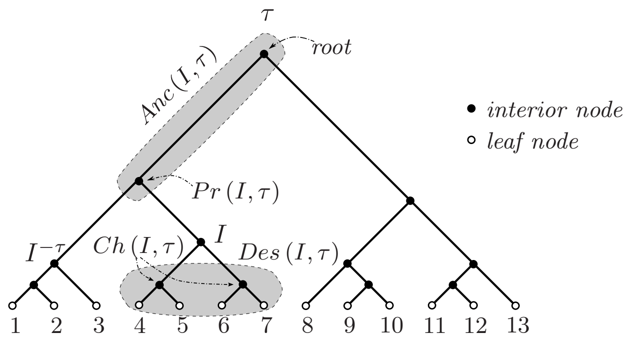


Fig. 1. Hierarchical Relations: ancestors - $\mathrm{Anc}\,(I,\tau)$, parent - $\mathrm{Pr}\,(I,\tau)$, children - $\mathrm{Ch}(I,\tau)$, descendants - $\mathrm{Des}\,(I,\tau)$, and local complement- $I^{-\tau}$ of cluster I of a rooted binary phylogenetic tree, $\tau \in \mathcal{BT}_S$, where $S = \{1,2,\ldots,13\}$. Filled and unfilled circles represent interior and leaf nodes, respectively. An interior node is referred by its cluster, the list of leaves below it; for example, $I = \{4,5,6,7\}$.

Definition 2 *A subset $\mathcal{A}$ of $\mathcal{P}(S)$ is said to be* nested *(also referred to in the literature as a* laminar family *[21]) if any two elements of $\mathcal{A}$ are compatible.*

Note that $\mathcal{C}\,(\tau)$ is also known as the laminar family associated with τ [21].

2.1.2 Hierarchical Relations

The cluster set $\mathcal{C}\,(\tau)$ of a hierarchy τ determines its representation as a graph $G_\tau = (V_\tau, E_\tau)$ completely: we have that $\mathcal{C}\,(\tau)$ stands in bijective correspondence with V_τ, and $(v,v') \in E_\tau$ if and only if $\mathcal{C}\,(v) \supset \mathcal{C}\,(v')$ and there is no $\tilde{v} \in V_\tau$ such that $\mathcal{C}\,(v) \supset \mathcal{C}\,(\tilde{v}) \supset \mathcal{C}\,(v')$. In particular, we adopt the following notation

$$\mathrm{Anc}\,(I,\tau) = \left\{ V \in \mathcal{C}\,(\tau) \,\middle|\, I \subsetneq V \right\}, \tag{8a}$$
$$\mathrm{Pr}\,(I,\tau) \in \mathrm{Anc}\,(I,\tau) \setminus \bigcup_{A \in \mathrm{Anc}(I,\tau)} \mathrm{Anc}\,(A,\tau), \tag{8b}$$
$$\mathrm{Ch}(I,\tau) = \left\{ V \in \mathcal{C}\,(\tau) \,\middle|\, \mathrm{Pr}\,(V,\tau) = I \right\}, \tag{8c}$$
$$\mathrm{Des}\,(I,\tau) = \left\{ V \in \mathcal{C}\,(\tau) \,\middle|\, V \subsetneq I \right\}, \tag{8d}$$

for the standard notions of, respectively, the set of ancestors, parents, children and descendants of every cluster $I \in \mathcal{C}\,(\tau)$. For the trivial case we set $\mathrm{Pr}\,(S,\tau) = \emptyset$. Because the children comprise a partition of each parent, we find it useful to define the *local complementary* cluster $I^{-\tau}$ of cluster $I \in \mathcal{C}\,(\tau)$ as

$$I^{-\tau} := \mathrm{Pr}\,(I,\tau) \setminus I, \tag{9}$$

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 $\mathrm{Pr}^2(I,\tau) := \mathrm{Pr}\left(\mathrm{Pr}\,(I,\tau),\tau\right)$ in τ. We denote the set of all grandchildren in τ by $\mathcal{G}(\tau)$,

$$\mathcal{G}(\tau) := \left\{ G \in \mathcal{C}\,(\tau) \,\middle|\, \mathrm{Pr}^2(G,\tau) \neq \emptyset \right\}. \tag{10}$$

If A, B are either elements of S or clusters of τ, it is convenient to have $(A \wedge B)_\tau$ denote the smallest common ancestor of A and B in τ,

$$(A \wedge B)_\tau := \underset{\substack{I \in \mathcal{C}(\tau) \\ A,B \subseteq I}}{\arg\min} |I|. \tag{11}$$

Finally, the depth function (or level), $\ell_\tau : \mathcal{C}\,(\tau) \to \mathbb{N}$, of hierarchy τ is defined by: [1]

$$\ell_\tau(I) := \left|\mathrm{Anc}\,(I,\tau)\right|, \quad \forall I \in \mathcal{C}\,(\tau). \tag{12}$$

Thus the depth of a cluster in the tree τ equals the number of its ancestors in τ.

2.1.3 Nondegeneracy & Certain Types of Trees

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 sets of rooted trees, over a fixed finite leaf set S, by $\mathcal{T}_S$ and the set of non-degenerate rooted trees 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. This leads to the well-known result:

Remark 1 ([21], [22]) *Let $\tau \in \mathcal{T}_S$. Then τ has at most $2|S| - 1$ vertices, with equality if and only if τ is non-degenerate, if and only if $\mathcal{C}\,(\tau)$ is a maximal laminar family in $\mathcal{P}(S)$ with respect to inclusion[2].*

Definition 3 ([23]) *Hierarchies $\sigma, \tau \in \mathcal{T}_S$ are said to be* disjoint *if they have no non-trivial clusters in common.*

Definition 4 ([16], [17]) *For any set of trees T in $\mathcal{T}_S$, the strict consensus tree T_* of T is defined to be the tree consisting of all common clusters of trees in T, i.e.*

$$\mathcal{C}\,(T_*) = \bigcap_{\tau \in T} \mathcal{C}\,(\tau), \tag{13}$$

and the loose consensus tree T^* of T is the tree each of whose clusters is a cluster of at least one tree in T and is compatible with all trees in T, i.e.

$$\mathcal{C}\,(T^*) = \left\{ I \in \bigcup_{\tau \in T} \mathcal{C}\,(\tau) \,\middle|\, \forall \sigma \in T \quad I \bowtie \mathcal{C}\,(\sigma) \right\}. \tag{14}$$

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

1. Here, $|V|$ denotes the cardinality of V.
2. In this paper we adopt the convention that a laminar family does not contain the empty set (as an element).

2.2 Tree Operations

2.2.1 NNI Moves

A convenient restatement of the standard definition of NNI walks of unrooted binary trees [5], [6] for rooted binary trees, illustrated in Figure 2, is:

Definition 5 *Let $\tau \in \mathcal{BT}_S$. We say that $\sigma \in \mathcal{BT}_S$ is the result of performing a* Nearest Neighbor Interchange, *or* NNI move, *on τ at a grandchild $G \in \mathcal{G}(\tau)$ if*

$$\mathcal{C}(\sigma) = \big(\mathcal{C}(\tau) \setminus \{\Pr(G,\tau)\}\big) \cup \big(\Pr^2(G,\tau) \setminus G\big). \qquad (15)$$

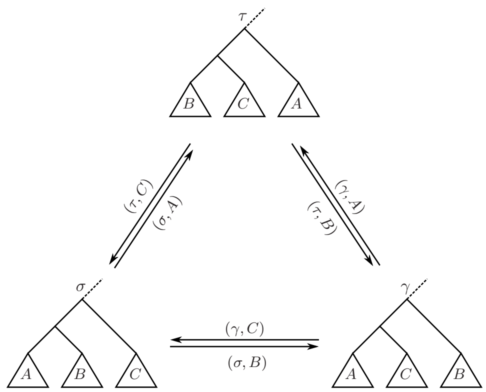


Fig. 2. An illustration of NNI moves between binary trees, each arrow is labeled by a source tree and associated grandchild defining the move.

Note that after an NNI move at cluster G of τ, grandchild G of grandparent $P = \Pr^2(G,\tau)$ with respect to τ becomes child G of parent $P = \Pr(G,\sigma)$ with respect to the adjacent tree σ.

2.2.2 The NNI-Graph

We define the NNI-graph $\mathcal{N}_S = (\mathcal{BT}_S, \mathcal{E})$ to have vertex set $\mathcal{BT}_S$, with two trees connected by an edge if and only if one can be obtained from the other by a single NNI move, see Figure 3. The NNI-graph on n leaves is a regular graph of degree $2(n-2)$ [5] [3] and the number of nondegenerate trees in the NNI-graph grows super exponentially with the number of leaves, n, [20],

$$\begin{aligned} \left|\mathcal{BT}_{[n]}\right| &= (2n-3)!! = (2n-3)(2n-5)\ldots 3, \\ &= \frac{(2n-2)!}{2^{n-1}(n-1)!}, \quad \text{for } n \geq 2. \end{aligned} \qquad (16)$$

As a result, an exploration of the entire NNI-graph (for example, searching for the shortest path between hierarchies or an optimal phylogenetic tree model) becomes rapidly more impractical and costly with increasing number of leaves.

A useful observation for NNI-adjacent trees is:

Lemma 1 *An ordered pair of hierarchies (σ,τ) is an edge in the NNI-graph $\mathcal{N}_S = (\mathcal{BT}_S, \mathcal{E})$ if and only if there exists one and only one 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 $\{B \cup C\} = \mathcal{C}(\tau) \setminus \mathcal{C}(\sigma)$. Call (A,B,C) the "NNI-triplet" of (σ,τ).*

3. It is clear that $|\mathcal{G}_\tau| = 2(|S|-2)$ for any $\tau \in \mathcal{BT}_S$.

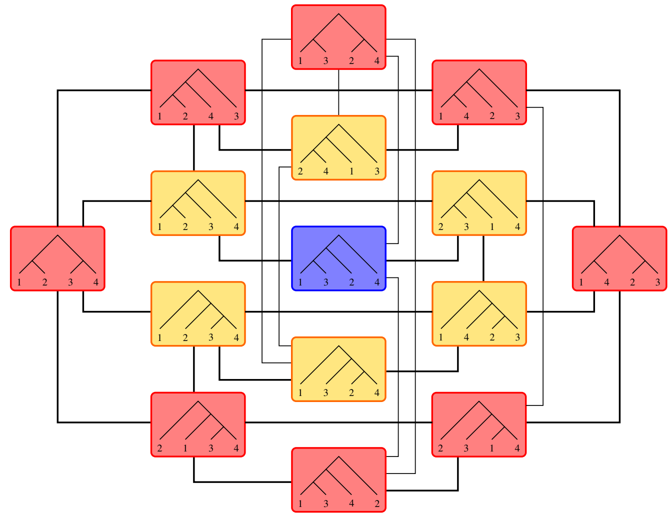


Fig. 3. The NNI Graph: a representation of the space of rooted binary trees, $\mathcal{BT}_S$, with NNI connectivity, for $S = [n] = \{1,2,3,4\}$.

Proof: See Appendix A.1 and Figure 2. ☐

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.

2.2.3 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*

$$res_K(\mathcal{A}) := \left\{ A \cap K \,\middle|\, A \in \mathcal{A},\ A \cap K \neq \emptyset \right\} \qquad (17)$$

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:

$$\sigma = res_K(\tau) \iff \mathcal{C}(\sigma) = \mathcal{C}(\tau)\big|_K. \qquad (18)$$

Remark 2 *Let $\tau \in \mathcal{BT}_S$ and $\{S_L, S_R\} = \mathrm{Ch}(S,\tau)$. Then,*

$$\mathcal{C}(\tau) = \mathcal{C}\left(\tau|_{S_L}\right) \cup \{J\} \cup \mathcal{C}\left(\tau|_{S_R}\right). \qquad (19)$$

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

By a *dissimilarity measure* on X, or simply a *dissimilarity*, we mean $d : X \times X \to \mathbb{R}_{\geq 0}$ on a space X is a nonnegative, symmetric function. A dissimilarity d on X is said to be positive definite if it satisfies $d(x,y) = 0 \iff x = y$ for any $x, y \in X$. For instance, many approximations of the

(NP-hard) NNI metric are positive definite dissimilarities [10]–[12].

A dissimilarity d is a metric if it satisfies the triangle inequality,

$$d(x,y) \leq d(x,z) + d(z,y), \quad \forall x,y,z \in X. \tag{20}$$

For example, we recall the definition of the commonly-used Robinson-Foulds metric on $X = \mathcal{T}_S$:

Definition 7 ([1]) *The Robinson-Foulds distance (or simply* symmetric difference distance*) $d_{RF} : \mathcal{T}_S \times \mathcal{T}_S \to \mathbb{R}_{\geq 0}$ between a pair of hierarchies $\sigma, \tau \in \mathcal{T}_S$ is defined to be*[4]

$$d_{RF}(\sigma,\tau) = \frac{1}{2}\left|\mathcal{C}(\sigma) \ominus \mathcal{C}(\tau)\right|. \tag{21}$$

Recently a more discriminative metric was introduced:

Definition 8 ([3], [4]) *(The Matching Split metric) Let $\sigma, \tau \in \mathcal{BT}_S$ and $G_S(\sigma,\tau)$ denote the complete bipartite graph with sides $\mathcal{C}_{int}(\sigma)$ and $\mathcal{C}_{int}(\tau)$, where each edge (I,J) carries the weight*[5]

$$A_S(I,J) = \min\left(\left|I \ominus J\right|, \left|I \ominus J^C\right|\right), \tag{22}$$

for all $I \in \mathcal{C}_{int}(\sigma)$ and $J \in \mathcal{C}_{int}(\tau)$.

The matching split distance $d_{MS} : \mathcal{BT}_S \times \mathcal{BT}_S \to \mathbb{R}_{\geq 0}$ *between a pair of hierarchies σ and τ is defined to be the value of a minimum-weighted perfect matching in $G_S(\sigma,\tau)$.*

Remark that $d_{RF} \leq d_{MS} \leq \frac{|S|+1}{2} d_{RF}$ [3], which explains the improvement in discriminative power over d_{RF}. However, the cost of computing a minimum weighted perfect matching in any $G_S(\sigma,\tau)$ is $\mathrm{O}\left(|S|^{2.5}\log|S|\right)$, which motivates the search for dissimilarities bounding d_{RF} from above but having a lower computational cost.

An ultrametric $d : X \times X \to \mathbb{R}$ is a metric on X satisfying the following strong form of the triangle inequality:

$$d(x,y) \leq \max\left(d(x,z), d(z,y)\right). \tag{23}$$

A restatement of a well-known fact [24]–[26] revealing the relation between hierarchies and ultrametrics is:

Lemma 3 ([24]) *Let $\tau \in \mathcal{T}_S$ and $h_\tau : \mathcal{C}(\tau) \to \mathbb{R}_{\geq 0}$. Then the following dissimilarity on S associated with τ,*

$$d_\tau(i,j) := h_\tau\left((i \wedge j)_\tau\right), \quad \forall i,j \in S, \tag{24}$$

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

i) *if $I \subseteq J$, then $h_\tau(I) \leq h_\tau(J)$,*
ii) *$h_\tau(I) = 0$ if and only if $|I| = 1$.*

Proof: See Appendix A.3. □

4. Here, $\ominus$ denotes the symmetric set difference, i.e. $A \ominus B = (A \setminus B) \sqcup (B \setminus A)$ for any sets A and B.
5. This corresponds to the hamming distance of clusters.

In general, a set X can inherit a metric from a metric space (Y, d_Y) by embedding X into Y through an injective function $f : X \to Y$; that is to say,

$$d_X(x_1,x_2) := d_Y(f(x_1), f(x_2)), \quad \forall x_1,x_2 \in X, \tag{25}$$

is a metric on X. d_X is known as the pullback metric of d_Y along f. In fact, the RF metric is a pullback metric: 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 $\tau \mapsto \mathcal{C}(\tau)$.

Lemma 3 provides us with a tool for constructing various embeddings of the space of trees in the space of (ultra-)metrics on S, which, in turn, can be identified with a subspace of $\mathbb{R}^{|S|\times|S|}$. A metric on $\mathcal{T}_S$ constructed by pullback of matrix norms under this embedding is introduced in the next section.

3 Discriminative Comparison of Edges

In this section, we shall introduce a new tree metric based on ultrametric representation of hierarchies and a dissimilarity measure counting pairwise cluster compatibilities of trees.

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 9 *The* ultrametric representation *is the map $\mathbf{U} : \mathcal{T}_S \to \mathbb{R}^{|S|\times|S|}$, defined by*

$$\mathbf{U}(\tau)_{ij} := h\left((i \wedge j)_\tau\right), \tag{26}$$

where $h : \mathcal{P}(S) \to \mathbb{N}$ is set as

$$h(I) := |I| - 1, \quad \forall I \subseteq S. \tag{27}$$

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

Proof: By Lemma 3, $\mathbf{U}(\tau)$ is an ultrametric representation of any $\tau \in \mathcal{T}_S$ since h (27) satisfies both conditions of Lemma 3.

To see the injectivity of $\mathbf{U}$ (26), we shall show that $\mathbf{U}(\sigma) \neq \mathbf{U}(\tau)$ for any $\sigma \neq \tau \in \mathcal{T}_S$.

Recall that 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 $\Pr(I,\sigma) \neq \Pr(I,\tau)$. Depending on the cardinality of parent clusters:

- If $|\Pr(I,\sigma)| = |\Pr(I,\tau)|$, then observe that there exists some $j \in \Pr(I,\sigma)$ and $j \notin \Pr(I,\tau)$ since $\Pr(I,\sigma) \neq \Pr(I,\tau)$. In fact, notice that $j \in I^{-\sigma}$ and $j \notin I^{-\tau}$ (recall (9)). Hence, we have $(i \wedge j)_\sigma =$

$\Pr(I,\sigma)$ and $\Pr(I,\tau) \subsetneq (i \wedge j)_\tau$ for any $i \in I$. Thus, it is clear that

$$\mathbf{U}(\sigma)_{ij} = |\Pr(I,\sigma)| - 1 < \mathbf{U}(\tau)_{ij} = \left|(i \wedge j)_\tau\right| - 1, \quad (28)$$

for any $i \in I$.

- Otherwise ($|\Pr(I,\sigma)| \neq |\Pr(I,\tau)|$), without loss of generality, let $|\Pr(I,\sigma)| < |\Pr(I,\tau)|$. Then, one can easily observe that

$$\mathbf{U}(\sigma)_{ij} = |\Pr(I,\sigma)| - 1 < \mathbf{U}(\tau)_{ij} = \left|(i \wedge j)_\tau\right| - 1, \quad (29)$$

for any $i \in I$ and $j \in I^{-\sigma}$ since $(i \wedge j)_\tau \supseteq \Pr(I,\tau)$.

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

Using the embedding $\mathbf{U}$ of $\mathcal{T}_S$ into $\mathbb{R}^{|S| \times |S|}$ (26), we can define different notions of tree metrics as the pullback metrics of various matrix norms as one below:

Definition 10 The cluster-cardinality metric, $d_{CC} : \mathcal{T}_S \times \mathcal{T}_S \to \mathbb{R}_{\geq 0}$, *in* $\mathcal{T}_S$ *is defined to be* [6]

$$d_{CC}(\sigma,\tau) := \frac{1}{2}\left\|\mathbf{U}(\sigma) - \mathbf{U}(\tau)\right\|_1, \quad \forall \sigma,\tau \in \mathcal{T}_S. \quad (30)$$

Here, we find the 1-norm convenient to illustrate the relation between d_{CC} and our dissimilarity measures d_{CM} and d_{Nav} later.

Proposition 1 *The cluster cardinality distance* $d_{CC} : \mathcal{T}_S \times \mathcal{T}_S \to \mathbb{R}_{\geq 0}$ *between a pair of hierarchies over a finite fixed leaf set* S *can be computed in* $\mathrm{O}\left(|S|^2\right)$ *time.*

Proof: The 1-norm of difference of a pair of $|J| \times |J|$ matrices obviously requires $\mathrm{O}\left(|S|^2\right)$ time to compute, which is the lower bound of the computation cost of d_{CC}. Now, we shall show that the embedding $\mathbf{U}$ (26) can also be efficiently obtained with the same computational cost, $\mathrm{O}\left(|S|^2\right)$, by post-order traversal, visiting children first and then the parent, of trees.

To show the result we follow a proof by induction based on the post order tree traversal. For any $\tau \in \mathcal{T}_S$:

- (Base case) For the two-leaf tree $\tau \in \mathcal{BT}_{[2]}$, i.e. $|S| = 2$, the result simply follows since $\mathbf{U}(\tau) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.
- (Induction) Otherwise ($|S| \geq 3$), let $\{S_k\}_{1 \leq k \leq K} = \mathrm{Ch}(S,\tau)$, where $K \geq 2$ is the number of children of the root S in τ. Note that for any $|S_k| = 1$, say $S_k = \{i\}$ for some $i \in S$, $\mathbf{U}(\tau)_{ii} = 0$. Hence, each element of $\mathbf{U}(\tau)$ associated with a singleton child, $|S_k| = 1$, can be updated in constant, $\mathrm{O}(1)$, time. Suppose that the relevant elements of $\mathbf{U}(\tau)$ associated with the subtree rooted at S_k can be computed in $\mathrm{O}\left(|S_k|^2\right)$ time for any $1 \leq k \leq K$ and $|S_k| \geq 2$. 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 $1 \leq k,l \leq K$. Here, remark that one can obtain cluster of each vertex of τ and its size in linear time during the post-order traversal of τ using $S = \bigcup_{k=1}^{K} S_k$ and $|S| = \sum_{k=1}^{K}|S_k|$. Hence, all clusters and their size can be obtained in $\mathrm{O}\left(|J|^2\right)$ time by a single post-order traversal. Thus, the total cost of obtaining $\mathbf{U}(\tau)$ by a single post-order traversal is $\sum_{k=1}^{K}\mathrm{O}\left(|S|_k^2\right) + \sum_{k=1}^{K}\sum_{l=1}^{K}|S_k|\,|S_l| + \mathrm{O}\left(|S|^2\right) = \mathrm{O}\left(|S|^2\right)$.

Thus, the result follows. □

It is always of interest to know the diameter, $\mathrm{diam}(X,d)$, of a finite metric space (X,d),

$$\mathrm{diam}(X,d) := \max\left\{d(x,y)\,\middle|\,x,y \in X\right\}. \quad (31)$$

Accordingly, the diameters of the set of hierarchies over a finite leaf set S in various metrics [3], [4], [27] are

$$\mathrm{diam}(\mathcal{T}_S, d_{RF}) = |S| - 2, \quad (32)$$

$$\mathrm{diam}(\mathcal{BT}_S, d_{MS}) = \mathrm{O}\left(|S|^2\right), \quad (33)$$

$$\mathrm{diam}(\mathcal{BT}_S, d_{NNI}) = \mathrm{O}(|S|\log|S|), \quad (34)$$

and the diameter of the set of all trees in the cluster-cardinality distance is:

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

Proof: It is evident from (27) that the minimum and maximum ultrametric distances between two different elements of S are, respectively, 1 and $|S| - 1$. Hence, the maximum elementwise difference of zero-diagonal $|S| \times |S|$ matrices in (30) is $|S| - 2$. Moreover, using the tight upper bound on the change of the cluster cardinality distance after a single NNI move in Proposition 3, the range of the diameter of $\mathcal{T}_S$ in the cluster cardinality metric can be obtained as

$$\left\lfloor \frac{2}{27}|S|^3 \right\rfloor \leq \mathrm{diam}(\mathcal{T}_S, d_{CC}) \leq \frac{1}{2}|S|(|S|-1)(|S|-2), \quad (35)$$

which completes the proof. □

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

$$d_{NNI}(\sigma,\tau) = 1 \Longleftrightarrow d_{RF}(\sigma,\tau) = 1, \quad (36)$$

$$d_{NNI}(\sigma,\tau) = 1 \Longrightarrow 2 \leq d_{MS}(\sigma,\tau) \leq \left\lfloor \frac{|S|}{2} \right\rfloor. \quad (37)$$

Accordingly, an important observation relating d_{CC} to the NNI moves is:

6. $\|.\|_1$ denotes elementwise 1-norm of a matrix, i.e. $\|\mathbf{U}\|_1 := \sum_{i=1}^{n}\sum_{j=1}^{n}|\mathbf{U}_{ij}|$ for all $\mathbf{U} \in \mathbb{R}^{n \times n}$.

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. Then, the cluster-cardinality distance between σ and τ satisfies*

$$2 \leq d_{CC}(\sigma, \tau) = 2\,|A|\,|B|\,|C| \leq \left\lfloor \frac{2}{27}\,|S|^3 \right\rfloor , \qquad (38)$$

and both bounds are tight.[7]

Proof: Let $P = A \cup B \cup C$ and, recall that $A \cup B \in \mathcal{C}(\sigma)$ and $B \cup C \in \mathcal{C}(\tau)$. Here, 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

$$\begin{aligned}
d_{CC}(\sigma, \tau) &= \frac{1}{2}\left\|\mathbf{U}(\sigma) - \mathbf{U}(\tau)\right\|_1 , && (39)\\
&= \sum_{\substack{i \in A\\ j \in B}} \left|\mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij}\right| + \sum_{\substack{i \in A\\ j \in C}} \left|\mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij}\right| \\
&\quad + \sum_{\substack{i \in B\\ j \in C}} \left|\mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij}\right| , && (40)\\
&= \sum_{\substack{i \in A\\ j \in B}} \underbrace{|h(A \sqcup B) - h(P)|}_{=|C|} + \sum_{\substack{i \in A\\ j \in C}} \underbrace{|h(P) - h(P)|}_{=0} \\
&\quad + \sum_{\substack{i \in B\\ j \in C}} \underbrace{|h(P) - h(B \cup C)|}_{=|A|}, && (41)\\
&= 2\,|A|\,|B|\,|C| . && (42)
\end{aligned}$$

The lower bound in (38) is clearly realized when $|A| = |B| = |C| = 1$. Moreover, one can simply verify that the maximum product of three numbers whose sum is constant occurs when all the numbers are equal. In our case, $|A| + |B| + |C| \leq |S|$, and so $|A|\,|B|\,|C| \leq \left\lfloor \frac{|S|^3}{27} \right\rfloor$ since $|.|$ is integer-valued. Thus, the result follows. ☐

Note that neither d_{MS} and nor d_{CC} completely capture the NNI-adjacency as d_{RF} does since neither of them provide a linear lower bound on the NNI distance d_{NNI}:

Lemma 5 *Let (X, d_X) be a metric space and $G = (X, E, A)$ be a weighted undirected connected graph comprising the set X of vertices, a set E of edges and $A : E \rightarrow \mathbb{R}_{\geq 0}$ a weight function on E. Let $d_G : X \times X \rightarrow \mathbb{R}$ denote the sum of edge weights along the shortest path in G joining a pair of elements of X.*

If $d_X(x, y) \leq A(x, y)$ for all $(x, y) \in E$, then $d_X(x, y) \leq d_G(x, y)$ for all $x, y \in X$.

Proof: The result simply follows from the triangle inequality of d_X, see Appendix A.4. ☐

Here are some useful applications of Lemma 5:

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

Corollary 1 *For all nondegenerate hierarchies $\sigma, \tau \in \mathcal{BT}_S$,*

$$d_{RF}(\sigma, \tau) \leq d_{NNI}(\sigma, \tau) . \qquad (43)$$

Corollary 2 *Let d be a dissimilarity on $\mathcal{BT}_S$ with the property that $d(\sigma, \tau) \leq 1$ for all $\sigma, \tau \in \mathcal{BT}_S$ and $d_{NNI}(\sigma, \tau) = 1$.*

If $d(\sigma, \tau) > d_{NNI}(\sigma, \tau)$ for some $\sigma, \tau \in \mathcal{BT}_S$, then d is not a metric.

Corollary 3 *The crossing dissimilarity, d_{CM} (46), and NNI navigation dissimilarity, d_{Nav} (105), are not metrics.*

3.2 The Crossing Dissimilarity

We define *the compatibility matrix*, $\mathbf{C}(\sigma, \tau)$, of a pair of hierarchies $\sigma, \tau \in \mathcal{T}_S$ to be the binary matrix[8]

$$\mathbf{C}(\sigma, \tau)_{I,J} := \mathbb{1}(I \bowtie J) , \qquad (44)$$

for all $\forall I \in \mathcal{C}(\sigma)$ and $J \in \mathcal{C}(\tau)$, where $\mathbb{1}(.)$ is the standard indicator function which returns unity if its argument is true; otherwise returns zero. Their *crossing matrix*, $\mathbf{X}(\sigma, \tau)$, is defined to be

$$\mathbf{X}(\sigma, \tau) := \mathbf{1} - \mathbf{C}(\sigma, \tau) , \qquad (45)$$

where $\mathbf{1}$ is the matrix of all ones of the appropriate size.

We propose a new dissimilarity, called *the crossing dissimilarity* d_{CM}, between a pair of nondegenerate hierarchies, $\sigma, \tau \in \mathcal{T}_S$, to be

$$d_{CM}(\sigma, \tau) := \left\|\mathbf{X}(\sigma, \tau)\right\|_1 , \qquad (46)$$

that counts pairwise cluster incompatibilities of trees. We find it convenient to use the 1-norm of the crossing matrix to easily reveal possible relations between d_{CM} and d_{CC} (30), but one can use any appropriate matrix norms to define alternative dissimilarities.

We now continue with a list of significant properties of d_{CM} including its relation with certain tree rearrangements and alternative tree metrics:

Remark 3 *The crossing dissimilarity d_{CM} (46) in $\mathcal{BT}_S$ is positive definite and symmetric, but it is not a metric (Corollary 3).*[9]

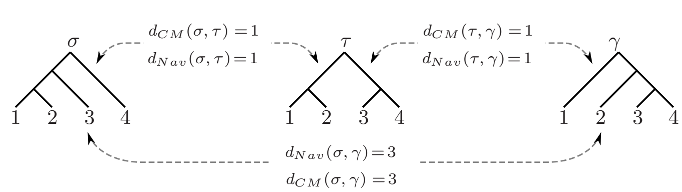


Fig. 4. d_{CM} and d_{Nav} are not metrics: an example of the triangle inequality failing for both dissimilarities.

8. $\mathbf{C}(\sigma, \tau)$ and $\mathbf{X}(\sigma, \tau)$ can be defined only in terms of nontrivial clusters of σ and τ since any two trivial clusters are always compatible with each other. However, this might result with matrices of zero dimension since a tree might have no nontrivial clusters.

9. Note that $d_{CM} : \mathcal{T}_S \times \mathcal{T}_S \rightarrow \mathbb{R}$ is nonnegative and symmetric.

Proposition 4 *The crossing dissimilarity $d_{CM} : \mathcal{T}_S \times \mathcal{T}_S \rightarrow \mathbb{R}_{\geq 0}$ between any pair of hierarchies over a finite fixed leaf set S can be computed in $\mathrm{O}\left(|S|^2\right)$ time.*

Proof: The crossing-matrix $\mathbf{X}(\sigma, \tau)$ (45) 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 a known crossing matrix $\mathbf{X}(\sigma,\tau)$ requires $\mathrm{O}\left(|S|^2\right)$ time to compute, which bounds the cost of d_{CM} from below. We shall show that $\mathbf{X}(\sigma,\tau)$ can be obtained in $\mathrm{O}\left(|S|^2\right)$ time by post-order traversals of trees.

Recall from Definition 1 that for any sets A and B

$$A \bowtie B \Longleftrightarrow (A \subseteq B) \vee (B \subseteq A) \vee (A \cap B = \emptyset)\,. \tag{47}$$

Let $\sigma, \tau \in \mathcal{T}_S$, and $I \in \mathcal{C}(\sigma)$ and $J \in \mathcal{C}(\tau)$. If at least one of 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 $|J| \geq 2$, observe the following recursions

$$I \subseteq J \Longleftrightarrow D \subseteq J, \qquad \forall D \in \mathrm{Ch}(I,\sigma)\,, \tag{48}$$

$$I \cap J = \emptyset \Longleftrightarrow D \cap J = \emptyset, \qquad \forall D \in \mathrm{Ch}(I,\sigma)\,. \tag{49}$$

Hence, for any cluster $J \in \mathcal{C}(\tau)$, one can easily check whether J is disjoint with or a superset of each cluster I of σ by a post-order traversal of σ in linear, $\mathrm{O}(|S|)$, time, and vice versa. Thus, all the pairwise inclusions and disjointness of clusters of σ and τ can be computed in $\mathrm{O}\left(|S|^2\right)$ time. As a result, using (47) $\mathbf{X}(\sigma,\tau)$ can be obtained with the same cost, $\mathrm{O}\left(|S|^2\right)$, which completes the proof. ☐

Proposition 5 *The diameter $\mathrm{diam}\left(\mathcal{T}_S, d_{CM}\right)$ (31) of the set of hierarchies over a fixed finite index set S with respect to the crossing dissimilarity d_{CM} (46) is*

$$\mathrm{diam}\left(\mathcal{T}_S, d_{CM}\right) = \left(|S| - 2\right)^2. \tag{50}$$

Proof: Note that the number of nontrivial clusters of a tree in $\mathcal{T}_S$ is at most $|S| - 2$ (Remark 1). Hence, an upper bound on $\mathrm{diam}\left(\mathcal{T}_S, d_{CM}\right)$ is $\left(|S| - 2\right)^2$. To observe that this upper bound is realized, see Figure 5. ☐

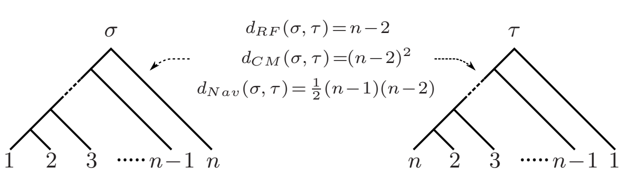


Fig. 5. A pair of nondegenerate hierarchies realizing $\mathrm{diam}\left(\mathcal{T}_{[n]}, d_{CM}\right) = (n-2)^2$ and $\mathrm{diam}\left(\mathcal{BT}_{[n]}, d_{Nav}\right) = \frac{1}{2}(n-1)(n-2)$.

Lemma 6 *Two non-degenerate hierarchies are NNI-adjacent if and only if they are crossing-adjacent, that is:*[10]

$$d_{NNI}(\sigma,\tau) = 1 \Longleftrightarrow d_{CM}(\sigma,\tau) = 1, \tag{51}$$

for all $\sigma, \tau \in \mathcal{BT}_S$

Proof: The result is simply evident from Remark 1 and Definition 5 on page 3. ☐

Here, note that d_{CM} and d_{NNI} define the same adjacency in $\mathcal{BT}_S$ (Lemma 6), but d_{CM} does not provide a linear lower bound on d_{NNI} since $\mathrm{diam}\left(\mathcal{BT}_S, d_{NNI}\right) = \mathrm{O}(|S|\log|S|) < \mathrm{diam}\left(\mathcal{BT}_S, d_{CM}\right) = \mathrm{O}\left(|S|^2\right)$ (Proposition 5). This inequality provides us with an additional, more conceptual, argument that d_{CM} is not a metric, by applying Lemma 5.

Proposition 6 *The crossing dissimilarity between any pair of nondegenerate hierarchies, $\sigma, \tau \in \mathcal{BT}_S$, has tight (linear) lower and (quadratic) upper bounds with respect to the RF distance,*

$$d_{RF}(\sigma,\tau) \leq d_{CM}(\sigma,\tau) \leq d_{RF}(\sigma,\tau)^2\,. \tag{52}$$

Proof: The lower bound directly follows from Remark 1 since a pair of distinct hierarchies always have uncommon clusters and an unshared cluster of one tree crosses at least one unshared cluster of the other tree. This bound is tight since

$$d_{RF}(\sigma,\tau)\!=\!1 \Leftrightarrow d_{NNI}(\sigma,\tau)\!=\!1 \Leftrightarrow d_{CM}(\sigma,\tau)\!=\!1, \tag{53}$$

for any $\sigma, \tau \in \mathcal{BT}_S$.

For any $\sigma, \tau \in \mathcal{BT}_S$, the columns and rows of $\mathbf{X}(\sigma,\tau)$ (45) associated with common clusters of trees have all zeros. Hence, each nonzero element of $\mathbf{X}(\sigma,\tau)$ is associated with a pair of unshared clusters of σ and τ. The number of unshared cluster pairs of nondegenerate trees σ and τ is equal to $d_{RF}(\sigma,\tau)^2$, which defines the upper bound in (52). To observe that this bound is also tight, see Figure 5. ☐

Proposition 7 *The crossing dissimilarity is bounded from above by the cluster-cardinality distance,*

$$d_{CM}(\sigma,\tau) \leq d_{CC}(\sigma,\tau)\,, \quad \forall \sigma, \tau \in \mathcal{T}_S. \tag{54}$$

Proof: Let $\sigma, \tau \in \mathcal{T}_S$. Throughout this proof, the symbols I and J denote clusters in $\mathcal{C}(\sigma)$ and $\mathcal{C}(\tau)$, respectively, and $i, j \in S$.

We shall introduce a function $q : \mathcal{C}(\sigma) \times \mathcal{C}(\tau) \rightarrow S \times S$ with the following properties:

i) for any I and J, $I \bowtie J \Longleftrightarrow i = j$ whenever $(i,j) = q(I,J)$,

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

10. In a finite metric space (X, d), two elements $x \neq y \in X$ are said to be adjacent if their distance is equal to the minimum nonnegative value of d in X, i.e. $d(x,y) = \min\left\{d(x',y')\,|x' \neq y' \in X\right\}$.

Hence, one can easily observe that

$$\bigcup_{i\neq j\in S} q^{-1}_{\sigma,\tau}(i,j) = \left\{(I,J)\,\middle|\, I \not\bowtie J\right\}. \tag{55}$$

As a result, it is evident from (55) that

$$d_{CM}(\sigma,\tau) = \sum_{i\neq j\in S}\left|q^{-1}_{\sigma,\tau}(i,j)\right| \leq d_{CC}(\sigma,\tau)\,. \tag{56}$$

Observe that if $I \not\bowtie J$, then there exists $i \in I\cap J$ and $j \in I\setminus J$ with the property that $(i\wedge j)_\sigma = I$. Accordingly, define

$$Q(I,J) := \left\{(i,j)\,\middle|\, i\in I\cap J,\ j\in I\setminus J,\ (i\wedge j)_\sigma = I\right\}, \tag{57}$$

$$R(I,J) := \left\{(i,j)\,\middle|\, i\in I\cap J,\ j\in J\setminus I,\ (i\wedge j)_\tau = J\right\}. \tag{58}$$

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

$$q(I,J) := \begin{cases}\left(\min(I\cup J)\,,\min(I\cup J)\right), & \text{if } I\bowtie J,\\ \min Q(I,J) & \text{, if } I\not\bowtie J, |I|\leq|J|\,,\\ \min R(I,J) & \text{, if } I\not\bowtie J, |I|>|J|\,.\end{cases} \tag{59}$$

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

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

$$q^{-1}(i,j) \subseteq A(i,j)\cup B(i,j)\,, \tag{60}$$

where

$$A(i,j) := \left\{(I,J)\,\middle|\, I\not\bowtie J,\ |I|\leq|J|\,,(i,j)\in Q(I,J)\right\}, \tag{61}$$

$$B(i,j) := \left\{(I,J)\,\middle|\, I\not\bowtie J,\ |I|\geq|J|\,,(i,j)\in R(I,J)\right\}. \tag{62}$$

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

$$A(i,j)\neq\emptyset \Longrightarrow B(i,j)=\emptyset. \tag{63}$$

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

$$I = (i\wedge j)_\sigma\,, J\subsetneq(i\wedge j)_\tau\,, |I|\leq|J| \text{ and } J\in \mathrm{Anc}\left(\{i\},\tau\right). \tag{64}$$

Hence, one can conclude that

$$|A(i,j)| \leq |(i\wedge j)_\tau| - |(i\wedge j)_\sigma| = \left(\mathbf{U}(\tau)_{ij} - \mathbf{U}(\sigma)_{ij}\right). \tag{65}$$

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

$$|B(i,j)| \leq |(i\wedge j)_\sigma| - |(i\wedge j)_\tau| = \left(\mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij}\right). \tag{66}$$

Thus, overall, using (60) and (63), one can obtain the second property of q as follows: for any $i\neq j\in S$

$$\left|q^{-1}_{\sigma,\tau}(i,j)\right| \leq |A(i,j)| + |B(i,j)| \leq \left|\mathbf{U}(\tau)_{ij} - \mathbf{U}(\sigma)_{ij}\right|, \tag{67}$$

which completes the proof. $\square$

4 NAVIGATION IN THE SPACE OF TREES

We now introduce an abstract discrete dynamical system in the NNI graph $\mathcal{N}_S = (\mathcal{BT}_S,\mathcal{E})$ of binary hierarchies over a fixed finite leaf set S. First, we shall propose a new NNI control policy to navigate toward any desired goal hierarchy $\tau\in\mathcal{BT}_S$ from any arbitrary hierarchy $\sigma\in\mathcal{BT}_S$ with provable convergence guarantees. Next, we will introduce a new dissimilarity between a pair of trees based on NNI navigation paths joining them.

4.1 A Discrete-Time Dynamical System Perspective

Recall that in order to define a control policy for navigating in $\mathcal{N}_S = (\mathcal{BT}_S,\mathcal{E})$, we need to construct an input bundle capturing all possible transitions (edges) in $\mathcal{N}_S$. Let $\widehat{\mathcal{E}}$ denote the set of *directed* edges of the NNI graph,

$$\widehat{\mathcal{E}} := \bigsqcup_{\sigma\in\mathcal{BT}_S}\widehat{\mathcal{E}}_\sigma = \bigcup_{\sigma\in\mathcal{BT}_S}\{\sigma\}\times\widehat{\mathcal{E}}_\sigma, \quad \widehat{\mathcal{E}}_\sigma := \mathcal{G}(\sigma)\cup\{\emptyset\}. \tag{68}$$

Every directed edge in $\widehat{\mathcal{E}}$ is referenced by a source tree and an associated grandchild in that tree. Consequently, let $\mathrm{NNI}:\widehat{\mathcal{E}}\to\mathcal{BT}_S$ denote the NNI move on a non-degenerate hierarchy $\sigma\in\mathcal{BT}_S$ at a grandchild cluster $G\in\widehat{\mathcal{E}}_\sigma$. Here, note that the NNI move at the empty cluster corresponds to the identity map in $\mathcal{BT}_S$, i.e. $\sigma = \mathrm{NNI}(\sigma,\emptyset)$ for all $\sigma\in\mathcal{BT}_S$. Therefore, the notion of identity map in $\mathcal{BT}_S$ slightly extends the NNI graph by adding self-loops at every vertex (it is necessary for a discrete-time dynamical system in $\mathcal{BT}_S$ to have fixed points).

Accordingly, one can consider an abstract discrete-time dynamical system in $\mathcal{BT}_S$ using NNI moves described as

$$\sigma^{k+1} = \mathrm{NNI}\left(\sigma^k, G^k\right), \tag{69a}$$

$$G^k = \mathbf{u}(\sigma^k), \tag{69b}$$

where $\mathbf{u}$ is a control policy of $\sigma^k\in\mathcal{BT}_S$ and returns a grandchild $G^k\in\widehat{\mathcal{E}}_{\sigma^k}$. Abusing notation, we shall denote the closed-loop dynamical system as

$$\sigma^{k+1} = \left(\mathrm{NNI}\circ\mathbf{u}\right)\left(\sigma^k\right). \tag{70}$$

4.2 Special Crossings of Clusters

Definition 11 *For any $\sigma,\tau\in\mathcal{BT}_J$, let $\mathcal{K}(\sigma,\tau)$ denote the set of common clusters of σ and τ with crossing splits,*

$$\mathcal{K}(\sigma,\tau) := \left\{K\in\mathcal{C}(\sigma)\cap\mathcal{C}(\tau)\,\middle|\,\mathrm{Ch}(K,\sigma)\neq\mathrm{Ch}(K,\tau)\right\}. \tag{71}$$

Note that two trees are distinct if and only if they have a common cluster with different splittings into children. That is to say for any $\sigma,\tau\in\mathcal{BT}_S$

$$\sigma=\tau \Longleftrightarrow \mathcal{K}(\sigma,\tau)=\emptyset. \tag{72}$$

Let $\sigma,\tau\in\mathcal{BT}_S$ be two distinct trees and $K\in\mathcal{K}(\sigma,\tau)$. Note that every cluster $I\in\mathcal{C}(\sigma)\setminus\mathrm{Des}(K,\sigma)$ either contains K or is disjoint with K, and so I is compatible with $\mathrm{Ch}(K,\tau)$. Hence, any cluster $I \in \mathcal{C}(\sigma)$ incompatible with $\mathrm{Ch}(K,\tau)$ is always a descendant of K in σ, which motivates the following definition:

Definition 12 *For $\sigma \neq \tau \in \mathcal{BT}_S$ and $K \in \mathcal{K}(\sigma,\tau)$, let*

$$\mathcal{I}(\sigma,\tau;K) := \left\{ I \in \mathrm{Des}\,(K,\sigma) \,\middle|\, I \not\bowtie \mathrm{Ch}(K,\tau) \right\}, \tag{73}$$

and denote the subset of deep *clusters incompatible with* $\mathrm{Ch}(K,\tau)$ *as (see Figure 6)*

$$\mathcal{D}(\sigma,\tau;K) := \Big\{ I \in \mathcal{I}(\sigma,\tau;K) \,\Big|\, \mathrm{Ch}(I,\sigma) \bowtie \mathrm{Ch}(K,\tau)\,, \\ \mathrm{Ch}\big(I^{-\sigma},\sigma\big) \bowtie \mathrm{Ch}(K,\tau) \Big\}. \tag{74}$$

Note that $\mathcal{I}(\sigma,\tau;K)$ and $\mathcal{D}(\sigma,\tau;K)$ are nonempty since at least an element of $\mathrm{Ch}(K,\sigma)$ is incompatible with $\mathrm{Ch}(K,\tau)$, and vice versa, which is evident from the following lemma:

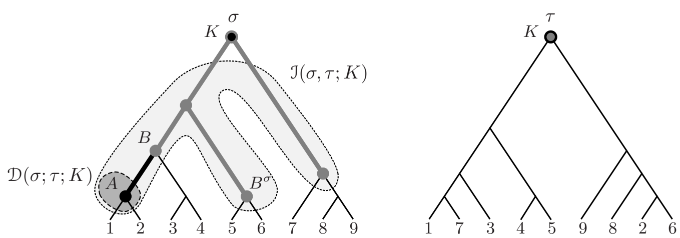


Fig. 6. An illustration of $\mathcal{I}(\sigma,\tau;K)$ and $\mathcal{D}(\sigma,\tau;K)$ of $\sigma,\tau \in \mathcal{BT}_{[9]}$, and $K = [9] \in \mathcal{K}(\sigma;\tau)$. The vertices and edges associated with clusters of σ incompatible with $\mathrm{Ch}(K,\tau)$ are thickened. The only deep cluster of σ incompatible with $\mathrm{Ch}(K,\tau)$ is $A = \{1,2\}$ which is also Type 1. B and $B^{-\sigma}$ are examples of Type 2 clusters incompatible with $\mathrm{Ch}(K,\tau)$.

Lemma 7 *Let $\{K_L, K_R\}$ be a bipartition of set K and $I \subsetneq K$. Then, the following equivalences hold*

$$I \bowtie \{K_L, K_R\} \Longleftrightarrow (I \subseteq K_L) \vee (I \subseteq K_R)\,. \tag{75}$$

Proof: See Appendix A.5. □

Here is a useful application of Lemma 7:

Corollary 4 *For any $\sigma \neq \tau \in \mathcal{BT}_S$ and $K \in \mathcal{K}(\sigma,\tau)$, if $I \in \mathcal{I}(\sigma,\tau;K)$, then $\mathrm{Anc}\,(I,\sigma) \cap \mathrm{Des}\,(K,\sigma) \subseteq \mathcal{I}(\sigma,\tau;K)$.*

Definition 13 *For any $\sigma \neq \tau \in \mathcal{BT}_S$ and $K \in \mathcal{K}(\sigma,\tau)$, a cluster $I \in \mathcal{I}(\sigma,\tau;K)$ is said to be* Type 1 *if $I^{-\sigma} \bowtie \mathrm{Ch}(K,\tau)$, and otherwise it is said to be* Type 2 *(see Figure 6).*

Finally, we find it useful to remark a property of Type 2 incompatibilities:

Lemma 8 *Let $\sigma \neq \tau \in \mathcal{BT}_S$ and $K \in \mathcal{K}(\sigma,\tau)$. Siblings $I, I^{-\sigma} \in \mathcal{I}(\sigma,\tau;K)$ are both Type 2 if and only if they are both incompatible with each child D of K in τ. That is to say, for any $I \in \mathrm{Des}\,(K,\sigma)$ and $D \in \mathrm{Ch}(K,\tau)$,*

$$\left\{ \begin{array}{l} I \not\bowtie \mathrm{Ch}(K,\tau)\,, \\ I^{-\sigma} \not\bowtie \mathrm{Ch}(K,\tau) \end{array} \right. \Longleftrightarrow \left\{ \begin{array}{l} I \not\bowtie D, \\ I^{-\sigma} \not\bowtie D. \end{array} \right. \tag{76}$$

Proof: See Appendix A.6. □

4.3 NNI Control Law

To navigate from an arbitrary hierarchy $\sigma \in \mathcal{BT}_S$ towards any selected desired hierarchy $\tau \in \mathcal{BT}_S$ in the NNI-graph, we propose an NNI control policy $\mathbf{u}_\tau$ that returns an NNI move on σ at a grandchild $G \in \mathcal{G}(\sigma) \cup \{\emptyset\}$, see Figure 7, as follows:

1) If $\sigma = \tau$, then just return the identity move, $G = \emptyset$.
2) Otherwise,
   a) Select a common cluster $K \in \mathcal{K}(\sigma,\tau)$ (71).
   b) Find a deep cluster $I \in \mathcal{D}(\sigma,\tau;K)$ (74) incompatible with $\mathrm{Ch}(K,\tau)$.
   c) Return a proper NNI navigation move on σ at grandchild $G \in \mathrm{Ch}(I,\sigma)$ selected as follows:
      i) If $I^{-\sigma} \bowtie \mathrm{Ch}(K,\tau)$ (Type 1, see Figure 8(a)), then return $G \in \mathrm{Ch}(I,\sigma)$ with the property that $G^{-\sigma}, I^{-\sigma} \subset J$ for some $J \in \mathrm{Ch}(K,\tau)$.
      ii) Otherwise (Type 2, see Figure 8(b)), return an arbitrary NNI move at a child of I in σ.

This NNI control law preserves common clusters of hierarchies. As a result, the navigation problem of trees can be divided into subproblems of disjoint trees which then may be solved in parallel. This is known as the decomposability property [28].

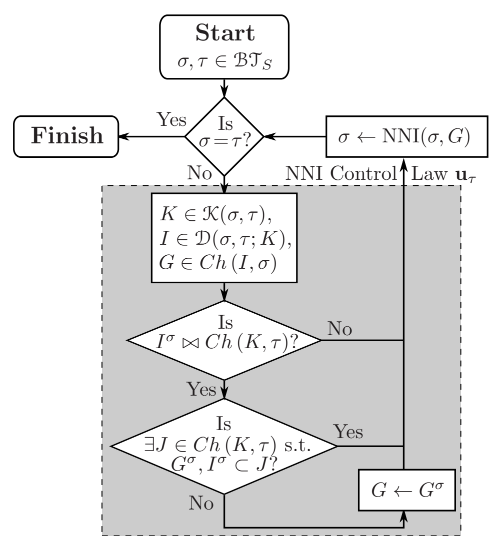


Fig. 7. A flowchart of navigating from $\sigma \in \mathcal{BT}_S$ towards $\tau \in \mathcal{BT}_S$ based on the NNI control law $\mathbf{u}_\tau$.

In brief, our NNI control scheme resolves incompatibilities between clusters of σ and τ level by level, depending on the selected common cluster K and one of its deep clusters I in Step 2. More precisely, for a fixed $K \in \mathcal{K}(\sigma,\tau)$, the clusters of σ incompatible with $\mathrm{Ch}(K,\tau)$

are replaced by compatible ones using NNI moves associated with deep clusters in $\mathcal{D}(\sigma,\tau;K)$ in a bottom to top fashion. If desired, one can choose the highest common cluster, $K = \arg\min_{J\in\mathcal{K}(\sigma,\tau)} \ell_\sigma(J) + \ell_\tau(J)$ – a top-down strategy, to obtain common splits at higher levels first, yielding higher priority resolution of incompatibilities for clusters closer to the root.

By construction, the NNI control law $\mathbf{u}_\tau$ is nondeterministic, and therefore generates multiple choices of paths from any given source σ to the target τ. All such paths will be referred to as *NNI navigation paths*. We leave the making of systematic selections of K and I in step (2) to a future discussion of specific implementations of the control policy [13], [29], [30]. Here we only mean to focus on properties of this control scheme pertinent to its application as a means for constructing NNI navigation paths and computing their length: since any two NNI navigation paths joining a given pair of vertices turn out (Theorem 2) to have equal lengths, the NNI control scheme gives rise to a new dissimilarity, d_{Nav}, which we now proceed to.

A final remark related to the behaviour of the NNI control law at deep clusters incompatible with $\mathrm{Ch}(K,\tau)$ is:

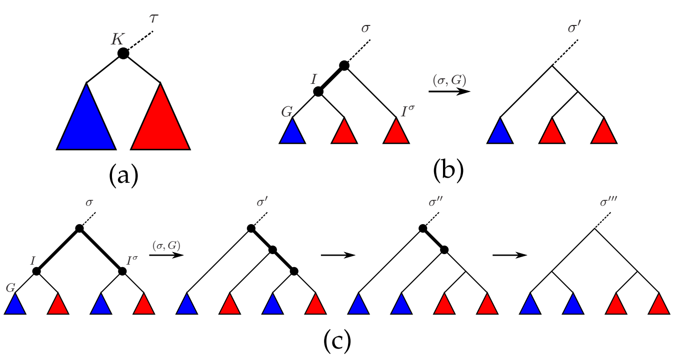


Fig. 8. An illustration of deep clusters incompatible with $\mathrm{Ch}(K,\tau)$: Type 1 (b) and Type 2 (c) incompatibilities with $\mathrm{Ch}(K,\tau)$ (a) of a common cluster $K \in \mathcal{K}(\sigma,\tau)$, and the associated NNI navigation moves until resolving incompatibilities with $\mathrm{Ch}(K,\tau)$. Clusters are colored according to their inclusion relation, and the thickened vertices show a portion of incompatible clusters in $\mathcal{I}(\sigma,\tau;K)$.

Remark 4 *Let $\sigma,\tau \in \mathcal{BT}_S$ be two distinct trees, and $K \in \mathcal{K}(\sigma,\tau)$ and $I \in \mathcal{D}(\sigma,\tau;K)$ be the cluster selected by the NNI control law $\mathbf{u}_\tau$. After a single NNI move,*

- *If I is Type 1, $\mathbf{u}_\tau$ replaces I by a cluster compatible with $\mathrm{Ch}(K,\tau)$ (see Figure 8(a)).*
- *If I is Type 2, $\mathbf{u}_\tau$ replaces I by a Type 1 cluster incompatible with $\mathrm{Ch}(K,\tau)$ and its sibling $I^{-\sigma}$, a Type 2 cluster in $\mathcal{D}(\sigma,\tau;K)$, becomes Type 1 incompatible with $\mathrm{Ch}(K,\tau)$ in the next hierarchy (see Figure 8(b)).*

Meanwhile, the incompatibility type (Type 1, Type 2 or none) of the rest of clusters in $\mathcal{C}(\sigma)\setminus\{I, I^{-\sigma}\}$ are kept unchanged in the next hierarchy.

Proposition 8 *An output of the NNI control law $\mathbf{u}_\tau$ to navigate in the NNI graph $\mathcal{N}_S = (\mathcal{BT}_S,\mathcal{E})$ towards a desired hierarchy $\tau \in \mathcal{BT}_S$ can be obtained in $\mathrm{O}(|S|)$ time.*

Proof: Let $\sigma \in \mathcal{BT}_S$. The common clusters and equality of σ and τ can be determined in linear time, $\mathrm{O}(|S|)$, using the algorithm in [2].

If $\sigma=\tau$, then $\mathbf{u}_\tau$ returns the identity move. Otherwise, $\mathcal{K}(\sigma,\tau)\neq\emptyset$ and, given the common clusters of σ and τ, a common cluster $K\in\mathcal{K}(\sigma,\tau)$ with crossing splits can be found in $\mathrm{O}(|S|)$ time by a traversal of σ. Given $K \in \mathcal{K}(\sigma,\tau)$, as discussed in the proof of Proposition 4, the clusters of σ incompatible with $\mathrm{Ch}(K,\tau)$, i.e. $\mathcal{I}(\sigma,\tau;K)$, can be determined in $\mathrm{O}(|K|)$ time using Lemma 7 and post-order traversal of subtree of σ rooted at K. Given $\mathcal{I}(\sigma,\tau;K)$, a deep cluster $I\in\mathcal{D}(\sigma,\tau;K)$ and a proper NNI move on σ at $G\in\mathrm{Ch}(I,\sigma)$ can be found in $\mathrm{O}(|K|)$ time by a traversal of the subtree of σ rooted at K.

Thus, the overall cost of computing the NNI control law τ is $\mathrm{O}(|S|)$, which completes the proof. □

In order to find a path from any given vertex $\sigma\in\mathcal{BT}_S$ to τ one simply obeys the controller $\mathbf{u}_\tau$. The rest of this section is dedicated to discussing the termination time complexity of this algorithm.

4.3.1 Stability Properties

For a desired nondegenerate hierarchy $\tau\in\mathcal{BT}_S$ over a fixed finite label set S, a candidate Lyapunov function [31] $V_\tau : \mathcal{BT}_S \to \overline{\mathbb{R}}_+$ can be defined using the crossing matrix $\mathbf{X}$ (45) as

$$V_\tau(\sigma) := \mathbf{p}_\sigma^{\mathrm{T}}\,\mathbf{X}(\sigma,\tau)\,\mathbf{p}_\tau, \tag{77}$$

$$= \sum_{\substack{I\in\mathcal{C}(\sigma)\\ J\in\mathcal{C}(\tau)}} \frac{1}{\rho^{\ell_\sigma(I)+\ell_\tau(J)}}\mathbb{1}(I \not\bowtie J), \tag{78}$$

where $\mathbf{p}_\tau \in \mathbb{R}_+^{(2|J|-1)\times 1}$ is the *hierarchical attenuation vector* associated with τ,

$$\mathbf{p}_\tau := \left(\frac{1}{\rho^{\ell_\tau(J)}}\right)_{J\in\mathcal{C}(\tau)}. \tag{79}$$

Here, $\rho\geq 1$ is a *hierarchical attenuation constant* and ℓ_τ (12) returns the level or depth of a cluster of τ. Note that since each nondegenerate hierarchy corresponds to a unique set of compatible clusters of maximum cardinality (Remark 1), it is clear that $V_\tau(\tau)=0$ and $V_\tau(\sigma)>0$ for all $\sigma\in\mathcal{BT}_S\setminus\{\tau\}$. Also, observe that $V_\tau(\sigma)$ is a weighted version (a continuous one-parameter deformation) of $d_{CM}(\sigma,\tau)$ where equality holds for $\rho=1$ (46).

One can conclude that:

Theorem 1 *The NNI control law $\mathbf{u}_\tau$ defines an abstract closed loop discrete dynamical system (70) in the NNI-graph $\mathcal{N}_S = (\mathcal{BT}_S,\mathcal{E})$ on a fixed finite leaf set S. There exist $\epsilon > 0$*

such that for every $\rho \geq 10+4\sqrt{5}$ and for any $\sigma \in \mathcal{BT}_S \setminus \{\tau\}$ one has

$$V_\tau\big((NNI \circ \mathbf{u}_\tau)(\sigma)\big) - V_\tau(\sigma) < -\epsilon < 0. \tag{80}$$

Hence all paths generated by the control policy $\mathbf{u}_\tau$ terminate at the goal, τ.

Proof: See Appendix A.7 ☐

4.3.2 Tree Metrics and the NNI Control Law

The NNI control law is compatible with d_{RF} (21) and d_{CC} (30) in the sense that:

Proposition 9 *The Robinson-Foulds, d_{RF} (21), and cluster-cardinality, d_{CC} (30), distances to any desired hierarchy $\tau \in \mathcal{BT}_S$ are non-increasing at each evolution of the closed loop discrete dynamical system (70) obeying the NNI control law $\mathbf{u}_\tau$, i.e. for any $d \in \{d_{RF}, d_{CC}\}$ and $\sigma \in \mathcal{BT}_S$*

$$d\big(NNI \circ \mathbf{u}_\tau(\sigma)\,, \tau\big) - d(\sigma, \tau) \leq 0. \tag{81}$$

Proof: For d_{RF}, the result is evident from that the NNI control law $\mathbf{u}_\tau$ preserves the common clusters of the current and goal hierarchies.

For d_{CC}, the statement holds trivially for $\sigma = \tau$. If $\sigma \neq \tau$, then let $K \in \mathcal{K}(\sigma,\tau)$ with $\{K_L, K_R\} = \mathrm{Ch}(K,\tau)$ and $I \in \mathcal{D}(\sigma,\tau;K)$ be the selected clusters by the NNI control law while determining the NNI move on σ at $G \in \mathrm{Ch}(I,\sigma)$ yielding $\gamma = (\mathrm{NNI} \circ \mathbf{u}_\tau)(\sigma)$. Note that this restructuring of σ only changes relative relations of G, $G^{-\sigma}$ and $I^{-\sigma}$ below $P = \mathrm{Pr}^2(G,\sigma) \subseteq K$. Further, by Definition 13, $G \subseteq K_A$ and $G^{-\sigma} \subseteq K_B$ for some $A \neq B \in \{L, R\}$, and so $(i \wedge j)_\tau = K$ for any $i \in G$ and $j \in G^{-\sigma}$. Accordingly, the change in d_{CC} with respect to τ after the transition from σ to γ can be written as[11]

$$\begin{aligned}
d_{CC}(\gamma,\tau) - d_{CC}(\sigma,\tau) &= \frac{1}{2}\|\mathbf{U}(\gamma) - \mathbf{U}(\tau)\| - \frac{1}{2}\|\mathbf{U}(\sigma) - \mathbf{U}(\tau)\|, \\
&= \sum_{\substack{i \in G \\ j \in G^{-\sigma}}} \Big|\mathbf{U}(\gamma)_{ij} - \mathbf{U}(\tau)_{ij}\Big| - \Big|\mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij}\Big| \\
&\quad + \sum_{\substack{i \in G \\ j \in I^{-\sigma}}} \Big|\mathbf{U}(\gamma)_{ij} - \mathbf{U}(\tau)_{ij}\Big| - \Big|\mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij}\Big| \\
&\quad + \sum_{\substack{i \in G^{-\sigma} \\ j \in I^{-\sigma}}} \Big|\mathbf{U}(\gamma)_{ij} - \mathbf{U}(\tau)_{ij}\Big| - \Big|\mathbf{U}(\sigma)_{ij} - \mathbf{U}(\tau)_{ij}\Big|, \qquad (82) \\
&= \sum_{\substack{i \in G \\ j \in G^{-\sigma}}} \underbrace{\big|h(P) - h(K)\big| - \big|h(I) - h(K)\big|}_{= -h(P) + h(I) = -|I^{-\sigma}|} \\
&\quad + \sum_{\substack{i \in G \\ j \in I^{-\sigma}}} \Big|h(P) - \mathbf{U}(\tau)_{ij}\Big| - \Big|h(P) - \mathbf{U}(\tau)_{ij}\Big| \\
&\quad + \sum_{\substack{i \in G^{-\sigma} \\ j \in I^{-\sigma}}} \underbrace{\Big|h(G^{-\sigma} \sqcup I^{-\sigma}) - \mathbf{U}(\tau)_{ij}\Big| - \Big|h(P) - \mathbf{U}(\tau)_{ij}\Big|}_{\in [h(G^{-\sigma} \sqcup I^{-\sigma}) - h(P),\, h(P) - h(G^{-\sigma} \sqcup I^{-\sigma})] = [-|G|, |G|]}, \qquad (83) \\
&\leq -\Big|G\Big|\,\big|G^{-\sigma}\big|\,\big|I^{-\sigma}\big| + \Big|G\Big|\,\big|G^{-\sigma}\big|\,\big|I^{-\sigma}\big| = 0. \qquad (84)
\end{aligned}$$

11. Here, one can easily verify that $d_{CC}(\gamma,\tau) - d_{CC}(\sigma,\tau) < 0$ if I is jointly incompatible with split $\mathrm{Ch}(K,\tau)$.

Note that equality in (81) can hold for both d_{RF} and d_{CC} as illustrated in Figure 9. ☐

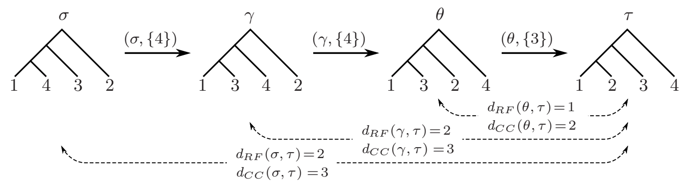


Fig. 9. An NNI navigation path joining σ to τ and associated NNI navigation moves. The NNI move from σ to $\gamma = (\mathrm{NNI} \circ \mathbf{u}_\tau)(\sigma)$ illustrates that d_{CC} and d_{RF} to the desired hierarchy τ might stay the same after an NNI navigation transition.

4.4 The NNI Navigation Distance

Definition 14 *An* NNI navigation path *from $\sigma \in \mathcal{BT}_S$ to $\tau \in \mathcal{BT}_S$ is a path in the NNI-graph $\mathcal{N}_S = (\mathcal{BT}_S, \mathcal{E})$ consistent with the (nondeterministic) closed-loop dynamical system (70) obeying the NNI control law $\mathbf{u}_\tau$.*

The NNI navigation distance *$d_{Nav}(\sigma,\tau)$ between the trees σ, τ is the length of an NNI navigation path joining them.*

Here we prove (Theorem 2) that all NNI navigation paths joining a pair of trees have the same length, which makes d_{Nav} into a well-defined dissimilarity on $\mathcal{BT}_S$. Furthermore, we will provide an explicit expression for the NNI navigation distance, d_{Nav}, and show that it can be computed in $O\Big(|S|^2\Big)$ time (Proposition 10).

4.4.1 Resolving Incompatibilities with the Root Split

Let $\{S_L, S_R\}$ be a bipartition of the leaf set S, and $\mathcal{BT}_{\{S_L,S_R\}}$ denote the subset of $\mathcal{BT}_S$ containing nondegenerate hierarchies with the root split of $\{S_L, S_R\}$,

$$\mathcal{BT}_{\{S_L,S_R\}} := \Big\{\tau \in \mathcal{BT}_S \Big| \mathrm{Ch}(S,\tau) = \{S_L, S_R\}\Big\}. \tag{85}$$

Now, instead of the original problem of navigating from any $\sigma \in \mathcal{BT}_S$ towards a single desired hierarchy $\tau \in \mathcal{BT}_S$ in the NNI-graph, consider a related and simpler problem of navigating hierarchies towards the set $\mathcal{BT}_{\{S_L,S_R\}}$. One can easily observe that the NNI control law can be used to solve this new problem by selecting any desired hierarchy $\tau \in \mathcal{BT}_{\{S_L,S_R\}}$ and fixing the common cluster $K \in \mathcal{K}(\sigma,\tau)$, in Step 2a of the NNI control policy, as $K = S$. We denote this version of the NNI control policy by $\mathbf{u}_{\{S_L,S_R\}}$.

In general, for any bipartition $\{M_L, M_R\}$ of subset $M \subseteq S$ let $\mathbf{u}_{\{M_L,M_R\},S}$ be the local controller, whose domain is

$$\mathcal{BT}_{M,S} := \big\{\sigma \in \mathcal{BT}_S \big| M \in \mathcal{C}(\sigma)\big\}, \tag{86}$$

that coincides with the NNI control law $\mathbf{u}_\tau$ for $\tau \in \mathcal{BT}_{M,S}$ if for any $\sigma \in \mathcal{BT}_{M,S}$ the common cluster

$K \in \mathcal{K}(\sigma,\tau)$, selected by the NNI control policy in in Step 2a, is fixed as $K = M$. Observe that $\mathbf{u}_{\{M_L,M_R\},S}$ terminates at

$$\mathcal{BT}_{\{M_L,M_R\},S} := \left\{\sigma \in \mathcal{BT}_{M,S} \middle| \mathrm{Ch}(M,\sigma) = \{M_L, M_R\}\right\}. \tag{87}$$

In fact, from a hybrid systems perspective, for any desired hierarchy $\tau \in \mathcal{BT}_S$ the NNI control law $\mathbf{u}_\tau$ consists of local controllers $\mathbf{u}_{\mathrm{Ch}(K,\tau),S}$ associated with nonsingular clusters $K \in \mathcal{C}(\tau)$. For any hierarchy $\sigma \in \mathcal{BT}_S \setminus \{\tau\}$, by selecting $K \in \mathcal{K}(\sigma,\tau)$, the NNI control law $\mathbf{u}_\tau$ arbitrarily selects one, $\mathbf{u}_{\mathrm{Ch}(K,\tau),S}$, of the local controllers $\left\{\mathbf{u}_{\mathrm{Ch}(J,\tau),S}\right\}_{\substack{J\in\mathcal{C}(\tau)\\ |J|>1}}$ whose domain contains σ.

Here, a critical observation about where an NNI navigation path consistent with $\mathbf{u}_{\{S_L,S_R\}}$ ends in $\mathcal{BT}_{\{S_L,S_R\}}$ is:

Lemma 9 *Let $\sigma \in \mathcal{BT}_S$ and $\{S_L, S_R\}$ be a bipartition of S. The closed loop dynamical system (70) obeying the NNI control rule $\mathbf{u}_{\{S_L,S_R\}}$ converges in finite time to the nondegenerate hierarchy $\gamma \in \mathcal{BT}_{\{S_L,S_R\}}$ with cluster set*

$$\mathcal{C}(\gamma) = \mathcal{C}\left(\sigma|_{S_L}\right) \cup \{S\} \cup \mathcal{C}\left(\sigma|_{S_R}\right). \tag{88}$$

Proof: If $\mathrm{Ch}(S,\sigma) = \{S_L, S_R\}$, then the results directly holds with $\gamma = \sigma$.

Otherwise ($\sigma \notin \mathcal{BT}_{\{S_L,S_R\}}$), let $\tau \in \mathcal{BT}_{\{S_L,S_R\}}$ and $I \in \mathcal{D}(\sigma,\tau;S)$ (74) be the deep cluster selected by the NNI control law $\mathbf{u}_{\{S_L,S_R\}}$. Observe from Remark 4 (also see Figure 8) that, after a certain number of proper NNI moves resulting from $\mathbf{u}_\tau$, clusters I and $I^{-\sigma}$ are replaced by $\mathrm{Pr}(I,\sigma) \cap S_L$ and $\mathrm{Pr}(I,\sigma) \cap S_R$ while the rest of clusters of σ are kept unchanged. Also note that a cluster $J \in \mathcal{C}(\sigma)$ compatible with $\{S_L, S_R\}$ is a proper subset of either S_L or S_R (Lemma 7). Accordingly, one can conclude that $\mathbf{u}_{\{S_L,S_R\}}$ successively replaces all cluster of σ incompatible with $\{S_L, S_R\}$ as described above yielding a hierarchy $\gamma \in \mathcal{BT}_S$ with cluster set

$$\begin{aligned}\mathcal{C}(\gamma) = &\left\{I \cap S_L \middle| I \in \mathcal{C}(\sigma),\ I \cap S_L \neq \emptyset\right\} \cup \{S\} \\ &\cup \left\{I \cap S_R \middle| I \in \mathcal{C}(\sigma),\ I \cap S_R \neq \emptyset\right\}, \quad (89)\\ = &\ \mathcal{C}\left(\sigma|_{S_L}\right) \cup \{S\} \cup \mathcal{C}\left(\sigma|_{S_R}\right). \quad (90)\end{aligned}$$

Finally, the finite time convergence of the system directly follows from Theorem 1, and this completes the proof. □

An NNI navigation path from $\sigma \in \mathcal{BT}_S$ to the set $\mathcal{BT}_{\{S_L,S_R\}}$ is not necessarily unique since $\mathbf{u}_{\{S_L,S_R\}}$ is nondeterministic. However:

Lemma 10 *For any bipartition $\{S_L, S_R\}$ of S, the lengths of all NNI navigation paths from $\sigma \in \mathcal{BT}_S$ to $\mathcal{BT}_{\{S_L,S_R\}}$ (85) are the same, and the NNI navigation distance $d_{Nav}(\sigma, \mathcal{BT}_{\{S_L,S_R\}})$ is given by*

$$d_{Nav}\left(\sigma, \mathcal{BT}_{\{S_L,S_R\}}\right) = \sum_{I \in \mathcal{C}(\sigma)} \left(\eta \circ \kappa_{\sigma,\{S_L,S_R\}}\right)(I), \tag{91}$$

where $\kappa_{\sigma,\{S_L,S_R\}} : \mathcal{C}(\sigma) \to \{0,1,2\}$ returns the number of children of $I \in \mathcal{C}(\sigma)$ incompatible with $\{S_L, S_R\}$,

$$\kappa_{\sigma,\{S_L,S_R\}}(I) := \sum_{A \in \mathrm{Ch}(I,\sigma)} \mathbb{1}\left(A \not\bowtie \{S_L, S_R\}\right), \tag{92}$$

and

$$\eta : \{0,1,2\} \to \{0,1,3\} \tag{93a}$$
$$x \mapsto \frac{1}{2}\left(x^2 + x\right) \tag{93b}$$

encodes the required number of NNI moves to replace children of I incompatible with $\{S_L, S_R\}$ (Remark 4).

Proof: Proof by induction. If $\sigma \in \mathcal{BT}_{\{S_L,S_R\}}$ (base case), the result simply follows.

Otherwise (induction), let $\{\sigma_0 = \sigma, \sigma_1, \sigma_2, \ldots, \sigma_k\}$ be an NNI navigation path, consistent with the NNI control law $\mathbf{u}_{\{S_L,S_R\}}$, starting from σ and ending at $\sigma_k \in \mathcal{BT}_{\{S_L,S_R\}}$ such that for all $0 \leq i < k$ $\sigma_i \notin \mathcal{BT}_{\{S_L,S_R\}}$.

Observe that $(\sigma_1, \sigma_2, \ldots, \sigma_k)$ is also an NNI navigation path, consistent with the NNI control law $\mathbf{u}_{\{S_L,S_R\}}$, starting from σ and ending in $\mathcal{BT}_{\{S_L,S_R\}}$.

For any $\gamma \in \mathcal{BT}_S$ let $\alpha(\gamma)$ and $\beta(\gamma)$, respectively, denote the numbers of Type 1 and Type 2 clusters of γ incompatible with $\{S_L, S_R\}$.

Now suppose (91) holds, and so one can rewrite $d_{Nav}(\sigma_1, \mathcal{BT}_{\{S_L,S_R\}})$ as

$$d_{Nav}\left(\sigma_1, \mathcal{BT}_{\{S_L,S_R\}}\right) = \alpha(\sigma_1) + \frac{3}{2}\beta(\sigma_1). \tag{94}$$

Hence,

$$d_{Nav}\left(\sigma, \mathcal{BT}_{\{S_L,S_R\}}\right) = \alpha(\sigma_1) + \frac{3}{2}\beta(\sigma_1) + 1. \tag{95}$$

Let $I \in \mathcal{C}(\sigma)$ be the selected cluster by $\mathbf{u}_{\{S_L,S_R\}}$ so that the NNI move on σ at a child $G \in \mathrm{Ch}(I,\sigma)$ yields σ_1.

- <u>Case 1:</u> If I is Type 1 incompatible with $\{S_L, S_R\}$, then using Remark 4, one can verify that

$$\alpha(\sigma) = \alpha(\sigma_1) + 1, \text{ and } \beta(\sigma) = \beta(\sigma_1). \tag{96}$$

  As a result, (95) yields

$$d_{Nav}\left(\sigma, \mathcal{BT}_{\{S_L,S_R\}}\right) = \alpha(\sigma) + \frac{3}{2}\beta(\sigma). \tag{97}$$

- <u>Case 2:</u> If I is Type 2 incompatible with $\{S_L, S_R\}$, then similarly, using Remark 4, one can observe that

$$\alpha(\sigma) = \alpha(\sigma_1) - 2, \text{ and } \beta(\sigma) = \beta(\sigma_1) + 2, \tag{98}$$

  and so (95) becomes

$$d_{Nav}\left(\sigma, \mathcal{BT}_{\{S_L,S_R\}}\right) = \alpha(\sigma) + \frac{3}{2}\beta(\sigma). \tag{99}$$

Therefore, the result follows. □

Lemma 11 *For any bipartition $\{S_L, S_R\}$ of S and $\sigma \in \mathcal{BT}_S$, the NNI navigation distance $d_{Nav}(\sigma, \mathcal{BT}_{\{S_L,S_R\}})$ can be computed in linear time,* $\mathrm{O}(|S|)$.

Proof: As discussed in the proof of Proposition 4, all cluster compatibilities of σ with $\{S_L, S_R\}$ can be determined in $\mathrm{O}(|S|)$ time using Lemma 7 and by a post order traversal of σ. Therefore, the NNI navigation distance $d_{Nav}(\sigma, \mathcal{BT}_{\{S_L,S_R\}})$ in (91) can be computed in $\mathrm{O}(|S|)$ by a complete traversal of σ. □

Recall from Lemma 10 that $d_{Nav}(\sigma, \mathcal{BT}_{\{S_L,S_R\}})$ is a weighted count of cluster incompatibilities of σ with $\{S_L, S_R\}$. Hence, a tight bound on $d_{Nav}(\sigma, \mathcal{BT}_{\{S_L,S_R\}})$ can be found using the maximal cluster incompatibilities of σ with $\{S_L, S_R\}$ as follows:

Lemma 12 *Let $\{S_L, S_R\}$ be a bipartition of S and $\sigma \in \mathcal{BT}_S$. The NNI navigation distance $d_{Nav}(\sigma, \mathcal{BT}_{\{S_L,S_R\}})$ (91) is tightly bounded from above as*

$$d_{Nav}(\sigma, \mathcal{BT}_{\{S_L,S_R\}}) \leq |S| + \min(|S_L|, |S_R|) - 3, \quad (100)$$

where the numbers of clusters of σ incompatible with $\{S_L, S_R\}$ is at most $|S| - 2$ and the number of clusters of σ both of whose children are incompatible with $\{S_L, S_R\}$ is at most $\min(|S_L|, |S_R|) - 1$.

Proof: One can easily observe from (91) that the NNI navigation distance $d_{Nav}(\sigma, \mathcal{BT}_{\{S_L,S_R\}})$ is maximized if all nontrivial clusters of σ are incompatible with $\{S_L, S_R\}$ and the number of cluster siblings incompatible with $\{S_L, S_R\}$ is maximized.

Recall that the number of nontrivial clusters of a nondegenerate hierarchy $\sigma \in \mathcal{BT}_S$ is $|S| - 2$ (Remark 1). Hence, the maximum number of clusters of σ incompatible with $\{S_L, S_R\}$ is $|S| - 2$.

Now, consider the set of clusters of σ whose children are incompatible with $\{S_L, S_R\}$,

$$\mathcal{I}_{\{S_L,S_R\}}(\sigma) := \left\{ I \in \mathcal{C}(\sigma) \,\middle|\, \forall D \in \mathrm{Ch}(I, \sigma), D \not\bowtie \{S_L, S_R\} \right\}. \quad (101)$$

Let $I \in \mathcal{I}_{\{S_L,S_R\}}(\sigma)$. Recall from Lemma 8 that $D \cap F \neq \emptyset$ for all $D \in \mathrm{Ch}(I, \sigma)$ and $F \in \{S_L, S_R\}$. Hence, such a cluster $I \in \mathcal{I}_{\{S_L,S_R\}}(\sigma)$ with the minimal cardinality requires two elements from each block of $\{S_L, S_R\}$, and any other cluster $A \in \mathcal{I}_{\{S_L,S_R\}}(\sigma)$ requires at least one extra leaf label from each block of $\{S_L, S_R\}$ not contained in I. By successive application of this argument, one can easily observe that $\left|\mathcal{I}_{\{S_L,S_R\}}(\sigma)\right| \leq \min(|S_L|, |S_R|) - 1$.

As a result, an upper bound on (91) can be obtained as

$$d_{Nav}(\sigma, \mathcal{BT}_{\{S_L,S_R\}}) \leq \underbrace{|S| - 2}_{\substack{\text{maximum number of}\\ \text{clusters incompatible}\\ \text{with } \{S_L, S_R\}}} + \underbrace{\min(|S_L|, |S_R|) - 1}_{\substack{\text{maximum number of}\\ \text{cluster siblings}\\ \text{incompatible with } \{S_L, S_R\}}}. \quad (102)$$

One can easily verify that this upper bound is actually tight using the hierarchies $\sigma, \tau \in \mathcal{BT}_{[n]}$ in Figure 5 and the bipartition $\{S_L, S_R\} = \mathrm{Ch}([n], \tau) = \{\{1\}, \{2, 3, \ldots, n\}\}$, where $d_{Nav}(\sigma, \mathcal{BT}_{\{S_L,S_R\}}) = n - 2$. □

4.4.2 Properties of the NNI Navigation Distance

We now introduce a version of the crossing matrix $\mathbf{X}$ (45) encoding weighted special cluster crossings between trees, which will be shown to have a significant connection with d_{Nav} later in Theorem 2.

Definition 15 *The special crossing matrix $\mathbf{S}(\sigma, \tau)$ of a pair of hierarchies $\sigma, \tau \in \mathcal{T}_S$ is defined to be*

$$\mathbf{S}(\sigma, \tau)_{I,J} := (\eta \circ \kappa_{\sigma,\tau})(I, J), \quad (103)$$

where $\kappa_{\sigma,\tau} : \mathcal{C}(\sigma) \times \mathcal{C}(\tau) \to \mathbb{R}$ counts the number of restricted crossings of children of cluster I in σ with children of cluster J in τ,

$$\kappa_{\sigma,\tau}(I, J) := \sum_{A \in \mathrm{Ch}(I,\sigma)|_J} \mathbb{1}\left(A \not\bowtie \mathrm{Ch}(J, \tau)\big|_I\right), \quad (104)$$

where $\mathrm{Ch}(I, \sigma)\big|_J$ denotes the restriction of $\mathrm{Ch}(I, \sigma)$ on J (Definition 6), and $\eta : \mathbb{R} \to \mathbb{R}$ is as defined in (93b).

Recall that an NNI navigation path joining $\sigma, \tau \in \mathcal{BT}_S$ might not be unique since the NNI control law in Section 4.3 is nondeterministic. However:

Theorem 2 *The length of all NNI navigation paths joining $\sigma, \tau \in \mathcal{BT}_S$ are the same and the NNI navigation distance $d_{Nav}(\sigma, \tau)$ between σ and τ is given by*[12]

$$\begin{aligned} d_{Nav}(\sigma, \tau) &= d_{Nav}(\sigma, \mathcal{BT}_{\mathrm{Ch}(S,\tau)}) + d_{Nav}\left(\sigma|_{S_L}, \tau|_{S_L}\right) \\ &\qquad + d_{Nav}\left(\sigma|_{S_R}, \tau|_{S_R}\right), \quad (105a) \\ &= \|\mathbf{S}(\sigma, \tau)\|_1 \quad (105b) \end{aligned}$$

where $\{S_L, S_R\} = \mathrm{Ch}(S, \tau)$. Here, $d_{Nav}(\sigma, \mathcal{BT}_{\mathrm{Ch}(S,\tau)})$ (91) denotes the NNI navigation distance of σ to $\mathcal{BT}_{\mathrm{Ch}(S,\tau)}$ (85) and $\mathbf{S}(\sigma, \tau)$(103) is the special crossing matrix of σ and τ.

Proof: See Appendix A.9. □

Now, one might wonder if d_{Nav} (105) defines a dissimilarity measure in $\mathcal{BT}_S$, and may be a metric.

Lemma 13 *The NNI navigation distance $d_{Nav}(\sigma, \tau)$ between any pair of nondegenerate hierarchies $\sigma, \tau \in \mathcal{BT}_S$ is positive definite and symmetric, but it does not define a metric.*

Proof: By definition, d_{Nav} positive definite and symmetry. In addition to Corollary 3, see the counter example in Figure 4 for d_{Nav} being a metric. □

12. Note that $d_{Nav}(\sigma, \tau)$ is always zero for $|J| = 2$, which is the base case of the recursion. For any $I \subseteq S$ with $|I| = 1$ we trivially set $d_{Nav}(\sigma|_I, \tau|_I) = 0$.

Proposition 10 *The NNI navigation distance d_{Nav} in $\mathcal{BT}_S$ is computable in $\mathrm{O}\left(|S|^2\right)$ time.*

Proof: The result can be easily seen from the recursive expression of d_{Nav} (105a) and Lemma 11.

Let $\sigma, \tau \in \mathcal{BT}_S$ and $\{S_L, S_R\} = \mathrm{Ch}(S, \tau)$. The recursion in (105a) requires $d_{Nav}\left(\sigma, \mathcal{BT}_{\mathrm{Ch}(S,\tau)}\right)$, which can be obtained in $\mathrm{O}(|S|)$ time (Lemma 11), and the restrictions of σ to S_L and S_R, which can be computable by post-order traversal of σ in linear time, $\mathrm{O}(|S|)$. Hence, d_{Nav} (105a) requires a complete (depth-first) traversal of τ each of whose iteration costs linear time with the number of leaves. Thus, the recursive computation of d_{Nav} costs $\mathrm{O}\left(|S|^2\right)$ time. Also note that the special crossing matrix $\mathbf{S}(\sigma, \tau)$ (103) can be obtained with the same cost, $\mathrm{O}\left(|S|^2\right)$ using a similar construction to the crossing matrix $\mathbf{X}(\sigma, \tau)$ (45) in the proof of Proposition 4. ☐

Proposition 11 *The diameter $\mathrm{diam}\left(\mathcal{BT}_S, d_{Nav}\right)$ (31) of $\mathcal{BT}_S$ in the NNI navigation distance d_{Nav} (105) is*

$$\mathrm{diam}\left(\mathcal{BT}_S, d_{Nav}\right) = \frac{1}{2}(|S|-1)(|S|-2)\,. \tag{106}$$

Proof: For $|S| = 2$, d_{Nav} is always zero since $\left|\mathcal{BT}_{[2]}\right| = 1$, and so (106) holds.

For $|S| \geq 3$, let $\sigma, \tau \in \mathcal{BT}_S$. Using the recursive formula of d_{Nav} in (105a), we can verify the result as follows

$$\begin{aligned}
d_{Nav}(\sigma, \tau) = {} & \underbrace{d_{Nav}\left(\sigma, \mathcal{BT}_{\mathrm{Ch}(S,\tau)}\right)}_{\substack{\text{by Lemma 12} \\ \leq |S| + \min(|S_L|, |S_R|) - 3}} + \underbrace{d_{Nav}\left(\sigma\big|_{S_L}, \tau\big|_{S_L}\right)}_{\leq \frac{1}{2}(|S_L|-1)(|S_L|-2)} \\
& + \underbrace{d_{Nav}\left(\sigma\big|_{S_R}, \tau\big|_{S_R}\right)}_{\leq \frac{1}{2}(|S_R|-1)(|S_R|-2)}, && (107) \\
\leq {} & \frac{1}{2}\left(|S_L|^2 + |S_R|^2\right) - \frac{3}{2}\underbrace{(|S_L| + |S_R|)}_{|S|} + |S| + \min(|S_L|, |S_R|) - 1, && (108) \\
= {} & \frac{1}{2}\Big(|S|^2 - \underbrace{2\,|S_L|\,|S_R|}_{=\min(|S_L|,|S_R|)(|S| - \min(|S_L|,|S_R|))}\Big) - \frac{1}{2}|S| + \min(|S_L|, |S_R|) - 1, && (109) \\
= {} & \frac{1}{2}(|S|-1)(|S|-2) + \underbrace{\left(1 - \min(|S_L|, |S_R|)\right)(|S|-1)}_{\leq 0, \ \forall |S| \geq 1}, && (110) \\
\leq {} & \frac{1}{2}(|S|-1)(|S|-2)\,, && (111)
\end{aligned}$$

where $\{S_L, S_R\} = \mathrm{Ch}(S, \tau)$. Here, recall that we set $d_{Nav}\left(\sigma\big|_I, \tau\big|_I\right) = 0$ for any $I \subset S$ with $|I| = 1$.

One can easily observe that this upper bound is tight by considering the trees in Figure 5. ☐

Although d_{Nav} is not a true metric, like d_{CM} (Proposition 6), it can be tightly bounded from below and above in terms of d_{RF} as follows:

Proposition 12 *For any $\sigma, \tau \in \mathcal{BT}_S$,*

$$d_{RF}(\sigma, \tau) \leq d_{Nav}(\sigma, \tau) \leq \frac{1}{2} d_{RF}(\sigma, \tau)^2 + \frac{1}{2} d_{RF}(\sigma, \tau), \tag{112}$$

where both bounds are tight.

Proof: The lower bound simply follows from the construction of $d_{Nav}(\sigma, \tau)$ since, by Definition 5, an NNI move on a hierarchy can only replace one cluster at a time. Hence, replacements of all unshared clusters take a number of NNI moves at least as much as the number of uncommon clusters, which is the Robinson-Foulds distance. The tightness of the bound is simply evident from

$$d_{RF}(\sigma, \tau) = 1 \Leftrightarrow d_{NNI}(\sigma, \tau) = 1 \Leftrightarrow d_{Nav}(\sigma, \tau) = 1, \tag{113}$$

for any $\sigma, \tau \in \mathcal{BT}_S$.

If σ and τ have no nontrivial common clusters, then $d_{RF}(\sigma, \tau) = |S| - 2$ and the result follows from Proposition 11. Otherwise, to prove the upper bound in (112), we shall use an induction. The result clearly holds for $|S| = 2$. For $|S| \geq 3$, let $I \in \mathcal{C}(\sigma) \cap \mathcal{C}(\tau)$ be a nontrivial common cluster such that we split σ and τ just above I yielding subtrees $\sigma_I = \sigma\big|_I$ and $\tau_I = \tau\big|_I$. Let $\sigma_{\neg I}$ and $\tau_{\neg I}$, respectively, denote the remaining parts of σ and τ after pruning subtrees rooted at cluster I and each of them contains a new leaf as a representative of associated pruned subtree.

Since the NNI control law preserves the common edges, the length of the NNI navigation path can be written as

$$d_{Nav}(\sigma, \tau) = d_{Nav}(\sigma_I, \tau_I) + d_{Nav}(\sigma_{\neg I}, \tau_{\neg I})\,. \tag{114}$$

Similarly, we have

$$d_{RF}(\sigma, \tau) = d_{RF}(\sigma_I, \tau_I) + d_{RF}(\sigma_{\neg I}, \tau_{\neg I})\,. \tag{115}$$

Let $\alpha = d_{RF}(\sigma_I, \tau_I)$ and $\beta = d_{RF}(\sigma_{\neg I}, \tau_{\neg I})$, and so $d_{RF}(\sigma, \tau) = \alpha + \beta$. Suppose that

$$d_{Nav}(\sigma_I, \tau_I) \leq \frac{1}{2}\alpha(\alpha+1)\,, \tag{116}$$

$$d_{Nav}(\sigma_{\neg I}, \tau_{\neg I}) \leq \frac{1}{2}\beta(\beta+1)\,. \tag{117}$$

It is evident from (114) that

$$\begin{aligned}
d_{Nav}(\sigma, \tau) &\leq \frac{1}{2}\alpha(\alpha+1) + \frac{1}{2}\beta(\beta+1)\,, && (118) \\
&\leq \frac{1}{2}(\alpha+\beta)(\alpha+\beta+1)\,, && (119) \\
&= \frac{1}{2} d_{RF}(\sigma, \tau)\left(d_{RF}(\sigma, \tau) + 1\right). && (120)
\end{aligned}$$

This is also a tight upper bound due to Proposition 11. Thus, the lemma follows. ☐

Alternatively, d_{Nav} can be bounded from above by the dissimilarity measure d_{CM} (46) counting the pairwise cluster crossings of trees:

Proposition 13 *For any $\sigma, \tau \in \mathcal{BT}_S$,*

$$d_{Nav}(\sigma, \tau) \leq \frac{3}{2} d_{CM}(\sigma, \tau)\,. \tag{121}$$

Proof: Consider the closed form expression of d_{Nav} (105b) in terms of the special crossing matrix $\mathbf{S}(\sigma, \tau)$ (103),

$$d_{Nav}(\sigma, \tau) = \|\mathbf{S}(\sigma, \tau)\|_1 = \sum_{\substack{I \in \mathcal{C}(\sigma) \\ J \in \mathcal{C}(\tau)}} (\eta \circ \kappa_{\sigma,\tau})(I, J)\,. \tag{122}$$

Note that $\kappa_{\sigma,\tau}(I, J) \in \{0, 1, 2\}$ for all nondegenerate trees in $\mathcal{BT}_S$, and $\eta(x) = \frac{1}{2}x^2 + \frac{1}{2}x \leq \frac{3}{2}x$ for all $x \in [0, 2]$. Hence, d_{Nav} can be bounded from above as

$$d_{Nav}(\sigma, \tau) \leq \frac{3}{2} \sum_{\substack{I \in \mathcal{C}(\sigma) \\ J \in \mathcal{C}(\tau)}} \kappa_{\sigma,\tau}(I, J)\,, \tag{123}$$

$$= \frac{3}{2} \sum_{\substack{I \in \mathcal{C}(\sigma) \\ J \in \mathcal{C}(\tau)}} \sum_{A \in \mathrm{Ch}(I,\sigma)|_J} \underbrace{\mathbb{1}\left(A \not\bowtie \mathrm{Ch}(J, \tau)\big|_I\right)}_{\leq \sum_{B \in \mathrm{Ch}(J,\tau)|_I} \mathbb{1}(A \not\bowtie B)}, \tag{124}$$

$$\leq \frac{3}{2} \sum_{\substack{I \in \mathcal{C}(\sigma) \\ J \in \mathcal{C}(\tau)}} \mathbb{1}\Big(\big(I \cap \mathrm{Pr}\,(J, \tau)\big) \not\bowtie \big(J \cap \mathrm{Pr}\,(I, \sigma)\big)\Big). \tag{125}$$

Now, let A, B, C be sets with $B \subseteq C$, then one can easily observe that

$$A \bowtie B \Longrightarrow (A \cap C) \bowtie B, \tag{126}$$

but the reverse is not necessarily true since $A \cap C \subseteq B$ does not imply $A \subseteq B$. In other words,

$$\mathbb{1}(A \bowtie B) \leq \mathbb{1}((A \cap C) \bowtie B)\,. \tag{127}$$

Thus, using the contra-positive of (126), one can rewrite the upper bound (125) on d_{Nav} as

$$d_{Nav}(\sigma, \tau) \leq \frac{3}{2} \sum_{\substack{I \in \mathcal{C}(\sigma) \\ J \in \mathcal{C}(\tau)}} \mathbb{1}(I \not\bowtie J) = \frac{3}{2} \|\mathbf{X}(\sigma, \tau)\|_1\,, \tag{128}$$

$$= \frac{3}{2} d_{CM}(\sigma, \tau)\,, \tag{129}$$

which completes the proof. □

The overall ordering of tree dissimilarities in Corollary 1, Proposition 7 and Proposition 13 can be combined as:

Theorem 3 *For nondegenerate hierarchies,*

$$\frac{2}{3} d_{RF} \leq \frac{2}{3} d_{NNI} \leq \frac{2}{3} d_{Nav} \leq d_{CM} \leq d_{CC}\,. \tag{130}$$

Finally, remark that the NNI navigation distance d_{Nav} can be generalized to a pair of trees, σ and τ, in $\mathcal{T}_S$ as

$$d_{Nav}(\sigma, \tau) = \frac{1}{2}\left(\|\mathbf{S}(\sigma, \tau)\|_1 + \|\mathbf{S}(\tau, \sigma)\|_1 \right), \tag{131}$$

which is nonnegative and symmetric. Note that for nondegenerate trees $\sigma, \tau \in \mathcal{BT}_S$, $\mathbf{S}(\sigma, \tau) = \mathbf{S}(\tau, \sigma)^{\mathrm{T}}$,[13] so d_{Nav} in (131) simplifies back to (105b).

13. $\mathbf{A}^{\mathrm{T}}$ denotes the transpose of a matrix $\mathbf{A}$.

Proposition 14 *For any $\sigma, \tau \in \mathcal{T}_S$,*

$$d_{Nav}(\sigma, \tau) \leq \left(\frac{1}{8} |S|^2 + \frac{1}{4} |S| \right) d_{CM}(\sigma, \tau)\,. \tag{132}$$

Proof: Note that the number of nontrivial children of a cluster of σ and τ can be at most $\frac{1}{2}|S|$. Hence one can verify the result following similar steps as in the proof of Proposition 13. □

5 Discussion and Numerical 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, the problem of 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 15 *Both the strict and loose consensus trees, T_* and T^*, of any set of trees T in $\mathcal{T}_S$ (Definition 4) are median trees in the sense of the crossing, d_{CM}(46), and NNI navigation, d_{Nav} (105), dissimilarities i.e. they both minimize the sum of distances to the sample trees with the total distance of zero, for any $d \in \{d_{CM}, d_{Nav}\}$*

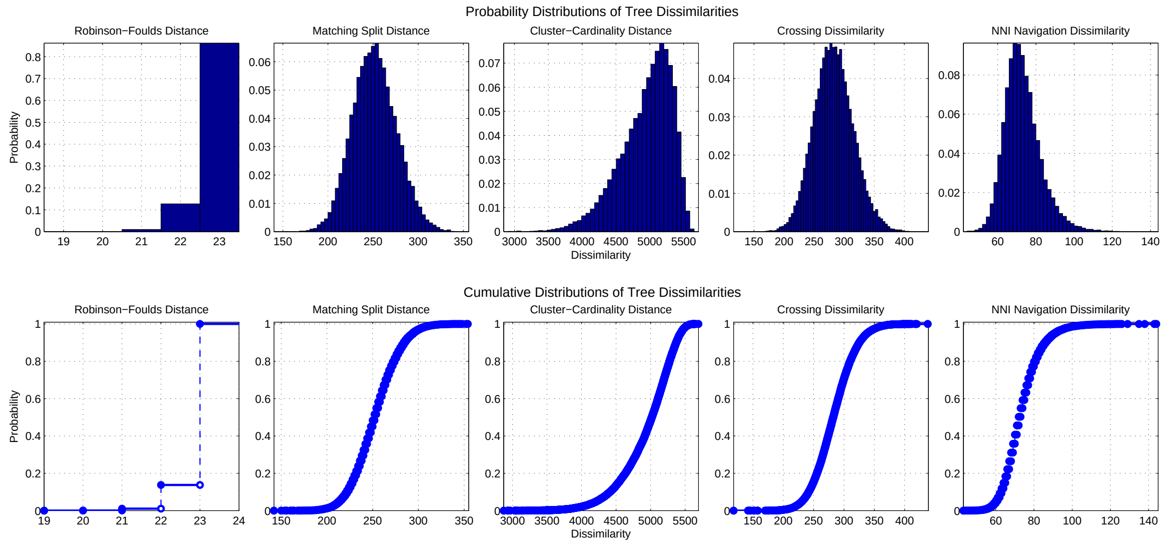
$$\sum_{\tau \in T} d(\tau, T_*) = \sum_{\tau \in T} d(\tau, T^*) = 0. \tag{133}$$

Moreover, the loose consensus tree is the maximal (finest) median tree sharing each of its clusters with at least one sample tree.

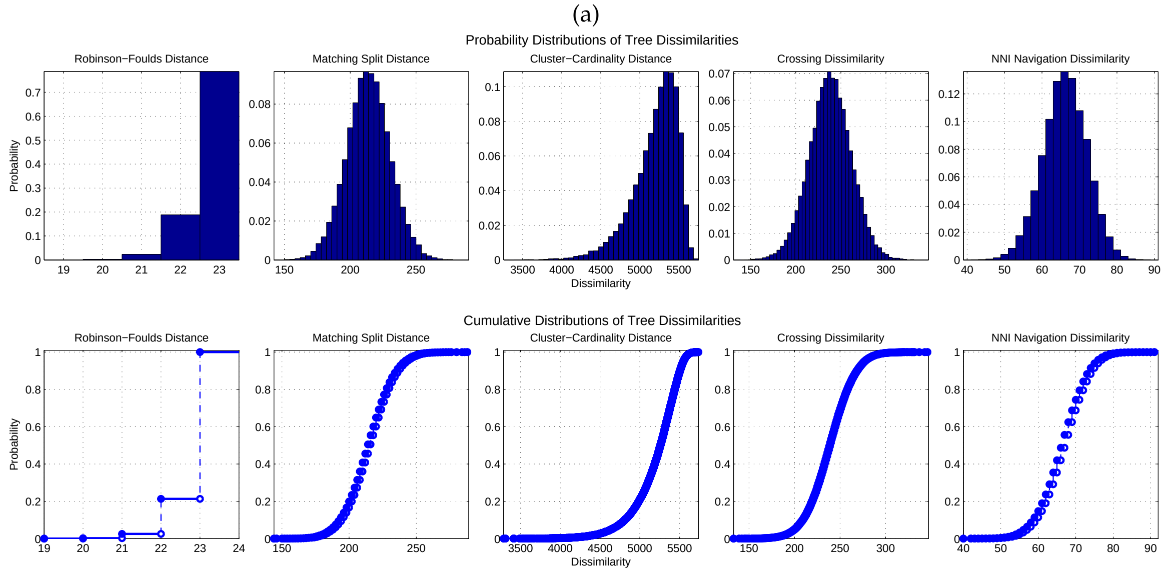
Proof: By Definition 4, both strict and loose consensus trees only contain compatible clusters with the clusters of any tree in T, and the loose consensus tree is the finest median tree containing only clusters from the sample trees. Thus, the result follows for both d_{CM} and d_{Nav} due their relation in Proposition 14. □

5.2 Sample Distribution of Dissimilarities

As states in [4], there is no available biologically motivated benchmark dataset for the comparison of different tree measures. As a result, a standard way of comparing tree dissimilarities is the statistical analysis of their distribution. 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 a discrete metric as in the case of the Robinson-Foulds distance — see Figure 10. 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



(a)



(b)

Fig. 10. Empirical distribution of tree dissimilarities in $\mathcal{BT}_{[25]}$: (from left to right) the Robinson-Foulds distance dRF (21) , the matching split distance d_{MS} (Def.8), the cluster-cardinality distance d_{CC} (30), the crossing dissimilarity d_{CM} (46), and the NNI navigation dissimilarity d_{Nav} (105). 100000 sample hierarchies are generated using (a) the uniform and (b)Yule model [32]. The resolutions of histograms, from left to right, are 1, 4, 64, 4, 1 unit(s), respectively.

model [32] 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 10. Moreover, in Table 1 we present two commonly used statistical measures, skewness and kurtosis, for describing the shapes of the probability distribution 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. In addition to their computational advantage over d_{MS}, as illustrated in both Figure 10 and Table 1, like d_{MS}, our tree measures, d_{CC}, d_{CM} and d_{Nav}, are significantly more informative and discriminative, with wider ranges of values and symmetry, than d_{RF}.

TABLE 1
Skewness and Kurtosis Values for the Distributions of Tree Measures

	Skewness		Kurtosis	
	Uniform	Yule	Uniform	Yule
d_{RF} (21)	-2.6162	-2.0740	9.8609	7.3998
d_{MS} (Def. 8)	0.1293	-0.0117	3.0060	3.1136
d_{CC} (30)	-0.9294	-1.2507	3.8601	5.2724
d_{CM} (46)	0.1390	-0.0405	3.1275	3.2103
d_{Nav} (105)	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, we construct the cluster-cardinality metric d_{CC} 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, we construct the NNI navigation dissimilarity d_{Nav} based on a kind of discrete path integral of (i.e. counting the steps along) the trajectory of an abstract discrete dynamical system in the space of trees. All of our tree measures can be efficiently computed in time $\mathrm{O}(n^2)$ in the number of leaves, n. Moreover, we provide a closed form expression for each proposed tree measure and present an ordering relation between these tree dissimilarities and other related tree metrics. In addition to being computationally more efficient than d_{MS}, using extensive numerical studies we demonstrate that the proposed tree measures are significantly more informative and discriminative than d_{RF}.

Appendix A
Proofs

A.1 Proof of Lemma 1

Proof: The sufficiency is simply evident from Definition 5 since the cluster sets of a pair of hierarchies differ exactly by one cluster if and only if they are joined by an NNI edge.

To see the necessity, let the NNI move on σ at $P \in \mathcal{G}(\sigma)$ joins σ to τ, and $R = P^{-\sigma}$ and $Q = \mathrm{Pr}^2(P,\sigma) \setminus \mathrm{Pr}\,(P,\sigma)$. By Definition 5, $\{\mathrm{Pr}\,(P,\sigma)\} = \{P \cup R\} = \mathcal{C}\,(\sigma) \setminus \mathcal{C}\,(\tau)$ 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 NNI-adjacent hierarchies differs exactly by one element. $\square$

A.2 Proof of Lemma 2

Proof: To show that $res_K(\mathcal{BT}_S) \supseteq \mathcal{BT}_K$, consider 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

$$\mathcal{C}\,(\tau) = \mathcal{C}\,(\sigma) \cup \{S\} \cup \mathcal{C}\,(\gamma)\,. \tag{134}$$

Now, observe that $\sigma = res_K(\tau)$ by Remark 2 where $\mathrm{Ch}(S,\tau) = \{K, S\setminus K\}$.

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 $A \in \mathcal{C}\,(\tau)$ such that $A \cap K = I \cap K$ and $A_L \cap K \neq \emptyset$ and $A_R \cap K \neq \emptyset$ for $\{A_L, A_R\} = \mathrm{Ch}(A,\tau)$. Hence, $\mathrm{Ch}\big(I \cap K, \tau\big|_K\big) = \{A_L \cap K, A_R \cap K\}$.

Now observe that either $I_L \cap K \neq \emptyset$ and $I_R \cap K \neq \emptyset$ for $\{I_L, I_R\} = \mathrm{Ch}(I,\tau)$, or there exists one and only one descendant $D \in \mathrm{Des}\,(I,\tau)$ such that $I \cap K = D \cap K$ and $D_L \cap K \neq \emptyset$ and $D_R \cap K \neq \emptyset$ where $\{D_L, D_R\} = \mathrm{Ch}(D,\tau)$.

Thus, all the interior clusters of $\tau\big|_K$ have exactly two children, which completes the proof. $\square$

A.3 Proof of Lemma 3

Proof: Let us start with the proof of the sufficiency. Positive definiteness and symmetry of d_τ are simply evident from (24) and Lemma 3.(ii). To observe the strong triangle inequality, let $i \neq j \neq k \in S$, and let $I \in \mathcal{C}\,(\tau)$ with the property that $i \in I_i$ and $j \in I_j$ where $\{I_i, I_j\} \subseteq \mathrm{Ch}(I,\tau)$.Hence, one can easily observe that $I = (\{i\} \wedge \{j\})_\tau$, and so $d_\tau(i,j) = h_\tau(I)$.

If $k \in I$, without loss of generality, let $k \in I_i$, and so $k \notin I_j$. Then, using (24) and Lemma 3.(i), one can simply verify that $d_\tau(i,k) \leq h_\tau(I_i) \leq h_\tau(I)$ and $d_\tau(j,k) = h_\tau(I)$ since $(\{i\} \wedge \{k\})_\tau \subseteq I_i$ and $(\{j\} \wedge \{k\})_\tau = I$. Note that if neither $k \in I_i$ nor $k \in I_j$ (but still $k \in I$), $d_\tau(i,k) = d_\tau(j,k) = h_\tau(I)$ since $(\{i\} \wedge \{k\})_\tau = (\{j\} \wedge \{k\})_\tau = I$.

Similarly, if $k \notin I$, then we have $d_\tau(i,k) \geq h_\tau(I)$ and $d_\tau(j,k) \geq h_\tau(I)$ since only some ancestors of I in τ might contain all i,j,k. Thus, we always have $d_\tau(i,j) \leq \max\big(d_\tau(i,k)\,, d_\tau(k,j)\big)$, which completes the proof of the sufficiency.

Let us continue with the necessity. Note that we have Lemma 3.(ii) from positive definiteness of d_τ. Let $I \in \mathcal{C}\,(\tau) \setminus \{S\}$ be any nonsingleton 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 strong triangle inequality of d_τ, one can easily see Lemma 3.(i) from

$$h_\tau(I) = d_\tau(i,j) \leq \max\big(d_\tau(i,k)\,, d_\tau(j,k)\big) = h_\tau\big(\mathrm{Pr}\,(I,\tau)\big), \tag{135}$$

which completes the proof. $\square$

A.4 Proof of Lemma 5

Proof: The result simply follows from the triangle inequality of d_X.

Let $x^0, x^* \in X$ and $(x_k)_{0 \leq k \leq K}$ be the sequence of points in X, where $(x_{k-1}, x_k) \in E$ for all $1 \leq k \leq K$ and $x_0 = x^0$ and $x_K = x^*$, defining the shortest path between x^0 and x^*. To put it another way,

$$d_G\big(x^0, x^*\big) = \sum_{k=1}^{K} \underbrace{A(x_{k-1}, x_k)}_{\geq d_X(x_{k-1}, x_k)}, \tag{136}$$

$$\geq \underbrace{\sum_{k=1}^{K} d_X\big(x_{k-1}, x_k\big)}_{\substack{\geq d_X\left(x^0, x^*\right),\\ \text{from the triangle inequality}}} \geq d_X\big(x^0, x^*\big)\,, \tag{137}$$

which completes the proof. $\square$

A.5 Proof of Lemma 7

Proof: By Definition 1, we have

$$I \bowtie K_L \Longleftrightarrow (I \subseteq K_L) \vee (K_L \nsubseteq I) \vee \underbrace{(I \cap K_L = \emptyset)}_{\Leftrightarrow I \subseteq K_R}, \tag{138}$$

$$\Longleftrightarrow (I \subseteq K_L) \vee (K_L \subseteq I) \vee (I \subseteq K_R)\,. \tag{139}$$

Therefore, the incompatibility of cluster I with $\{K_L, K_R\}$ can be rewritten as

$$\begin{aligned} I \bowtie \{K_L, K_R\} &= (I \bowtie K_L) \wedge (I \bowtie K_R)\,, && (140)\\ &= (I \subseteq K_L) \vee (I \subseteq K_R) \vee \underbrace{\Big((K_L \subseteq I) \wedge (K_R \subseteq I)\Big)}_{\text{false since } I \subsetneq K = K_L \cup K_R}, && (141)\\ &= (I \subseteq K_L) \vee (I \subseteq K_R)\,, && (142)\end{aligned}$$

and the lemma follows. □

A.6 Proof of Lemma 8

Proof of Lemma 8: The sufficiency for being Type 2 directly follows from Definition 1,

$$I \not\bowtie D, I^{-\sigma} \not\bowtie D \Longrightarrow I \not\bowtie \mathrm{Ch}(K,\tau)\,, I^{-\sigma} \not\bowtie \mathrm{Ch}(K,\tau)\,. \tag{143}$$

To see the necessity for being Type 2, recall from Lemma 7 that

$$\begin{aligned} I \bowtie \mathrm{Ch}(K,\tau) &\Longleftrightarrow (I \subseteq D) \vee (I \subseteq (K \setminus D))\,,\\ &\Longleftrightarrow (I \subseteq D) \vee (I \cap D = \emptyset)\,, \qquad (144)\\ I^{-\sigma} \bowtie \mathrm{Ch}(K,\tau) &\Longleftrightarrow \big(I^{-\sigma} \subseteq D\big) \vee \big(I^{-\sigma} \cap D = \emptyset\big)\,. \end{aligned}$$

Further, using Lemma 7, observe that

$$\begin{aligned} D \subseteq I &\Longrightarrow I^{-\sigma} \subseteq (K \setminus D) \Longrightarrow I^{-\sigma} \bowtie \mathrm{Ch}(K,\tau)\,,\\ D \subseteq I^{-\sigma} &\Longrightarrow I \bowtie \mathrm{Ch}(K,\tau)\,. \end{aligned} \tag{145}$$

As a result, using (144) and (145), one can obtain the necessity as

$$I \not\bowtie \mathrm{Ch}(K,\tau)\,, I^{-\sigma} \not\bowtie \mathrm{Ch}(K,\tau) \Longrightarrow I \not\bowtie D, I^{-\sigma} \not\bowtie D, \tag{146}$$

which completes the proof. □

A.7 Proof of Theorem 1

A useful observation characterizing the effect of hierarchical attenuation constant ρ below a certain level of a nondegenerate hierarchy is:

Lemma 14 *For any cluster $K \in \mathcal{C}(\tau)$ of a binary hierarchy $\tau \in \mathcal{BT}_S$, the hierarchical attenuation constant ρ satisfies*

$$\sum_{I \in \mathrm{Des}(K,\tau)} \frac{1}{\rho^{\ell_\tau(I)}} < \frac{2}{\rho - 2} \frac{1}{\rho^{\ell_\tau(K)}}, \quad \forall \rho > 2. \tag{147}$$

Proof: Proof by induction.

- For $|K| = 1$ (base case) : It is trivially true since $\mathrm{Des}(K,\tau) = \emptyset$.
- For $|K| \geq 2$ (induction): Using the clustering identity between descendants of cluster K and its children, $\{K_L, K_R\} = \mathrm{Ch}(K,\tau)$, in binary hierarchy τ, we can factor the left hand side and find an upper bound as

$$\begin{aligned} \sum_{I \in \mathrm{Des}(K,\tau)} \frac{1}{\rho^{\ell_\tau(I)}} &= \underbrace{\sum_{I \in \mathrm{Ch}(K,\tau)} \frac{1}{\rho^{\ell_\tau(I)}}}_{= \frac{2}{\rho^{\ell_\tau(K)+1}}} + \underbrace{\sum_{I \in \mathrm{Des}(K_L,\tau)} \frac{1}{\rho^{\ell_\tau(I)}}}_{< \frac{2}{\rho-2}\frac{1}{\rho^{\ell_\tau(K_L)}}}\\ &\quad + \underbrace{\sum_{I \in \mathrm{Des}(K_R,\tau)} \frac{1}{\rho^{\ell_\tau(I)}}}_{< \frac{2}{\rho-2}\frac{1}{\rho^{\ell_\tau(K_R)}}}, && (148)\\ &< \frac{1}{\rho^{\ell_\tau(K)}} \left(\frac{2}{\rho} + \frac{4}{\rho(\rho-2)}\right), && (149)\\ &= \frac{2}{\rho - 2} \frac{1}{\rho^{\ell_\tau(K)}}, && (150)\end{aligned}$$

  where the depth of children and parent clusters are related by $\ell_\tau(K_L) = \ell_\tau(K_R) = \ell_\tau(K) + 1$.

Thus, the result follows. □

Now, using Lemma 14, one can obtain an upper bound on the change of Lyapunov function after an NNI move as follows:

Lemma 15 *For any desired hierarchy $\tau \in \mathcal{BT}_S$ and hierarchical attenuation constant $\rho > 2$, the change in the value of Lyapunv function V_τ (77) after the NNI move on $\sigma \in \mathcal{BT}_S$ at $G \in \mathcal{G}(\sigma)$ is bounded from above as*

$$V_\tau(NNI(\sigma,G)) - V_\tau(\sigma) < \frac{1}{\rho^{\ell_\sigma(P)+\ell_\tau(K)+2}} \left(\frac{16(\rho-1)}{(\rho-2)^2} - 1\right), \tag{151}$$

where $P = \mathrm{Pr}^2(G,\sigma)$ and $K \in \mathcal{C}(\tau)$ satisfying $P \subseteq K$.[14]*.*

Proof: See Appendix A.8. □

The result in Theorem 1 directly follows from:

Corollary 5 *For any desired hierarhy $\tau \in \mathcal{BT}_S$ and hierarchical attenuation constant $\rho \geq 10 + 4\sqrt{5} (\approx 18.94)$, the discrete dynamical system in (70) obeying the NNI control law $\mathbf{u}_\tau$ in Section 4.3 strictly decreases the value of Lyapunov function V_τ (77) at any hierarchy $\sigma \in \mathcal{BT}_S$ away from τ,*

$$(V_\tau \circ NNI \circ \mathbf{u}_\tau)(\sigma) - V_\tau(\sigma) < 0. \tag{152}$$

A.8 Proof of Lemma 15

Since the root cluster S is a common cluster of all hierarchies in $\mathcal{BT}_S$, we find it convenient to rewrite V_τ (77) as

$$V_\tau(\sigma) = \sum_{J \in \mathcal{C}(\tau)} \frac{1}{\rho^{\ell_\tau(J)}} \Phi_\sigma\big(\mathrm{Des}(S,\sigma), J\big)\,, \tag{153}$$

where $\Phi_\sigma : \mathcal{P}\big(\mathcal{C}(\sigma)\big) \times \mathcal{C}(\tau) \to \mathbb{R}_{\geq 0}$ returns the total weighted incompatibilities of a subset $\mathcal{A}$ of $\mathcal{C}(\sigma)$ with a cluster B of τ,

$$\Phi_\sigma(\mathcal{A}, B) := \sum_{A \in \mathcal{A}} \frac{1}{\rho^{\ell_\sigma(A)}} \mathbb{1}(A \not\bowtie B)\,. \tag{154}$$

14. Such a cluster $K \in \mathcal{C}(\tau^*)$ always exists since $P \subseteq S$ and $S \in \mathcal{C}(\tau)$

Using the attenuation characteristic of ρ over a non-degenerate hierarchy in Lemma 14, lower and upper bounds of $\Phi_\sigma\big(\mathrm{Ch}(I,\sigma), J\big)$ and $\Phi_\sigma\big(\mathrm{Des}(I,\sigma), J\big)$ for clusters $I \in \mathcal{C}(\sigma)$ and $J \in \mathcal{C}(\tau)$ can be simply obtained as

$$0 \leq \Phi_\sigma\big(\mathrm{Ch}(I,\sigma), J\big) \leq \frac{2}{\rho^{\ell_\sigma(I)+1}}, \tag{155}$$

$$0 \leq \Phi_\sigma\big(\mathrm{Des}(I,\sigma), J\big) < \frac{2}{\rho-2}\frac{1}{\rho^{\ell_\sigma(I)}}, \tag{156}$$

for $\rho > 2$.

One can further factor out $\Phi_\sigma\big(\mathrm{Des}(I,\sigma), J\big)$ using the hierarchical relation between descendant clusters as

$$\begin{aligned}\Phi_\sigma\big(\mathrm{Des}(I,\sigma), J\big) = \Phi_\sigma\big(\mathrm{Ch}(I,\sigma), J\big) + \Phi_\sigma\big(\mathrm{Des}(I_L,\sigma), J\big) \\ + \Phi_\sigma\big(\mathrm{Des}(I_R,\sigma), J\big),\end{aligned} \tag{157}$$

where $\{I_L, I_R\} = \mathrm{Ch}(I,\sigma)$. Combining all (155), (156) and (157) yields the following tighter lower and upper bounds on $\Phi_\sigma(\mathrm{Des}(I), J)$,

$$\begin{aligned}\Phi_\sigma\big(\mathrm{Ch}(I,\sigma), J\big) &\leq \Phi_\sigma\big(\mathrm{Des}(I,\sigma), J\big) \\ &< \Phi_\sigma\big(\mathrm{Ch}(I,\sigma), J\big) + \frac{4}{\rho-2}\frac{1}{\rho^{\ell_\sigma(I)+1}}.\end{aligned} \tag{158}$$

Here, note that the levels of children and parent clusters are related to each other by $\ell_\sigma(I_L) = \ell_\sigma(I_R) = \ell_\sigma(I) + 1$.

Change in the value of Lyapunov function:

Before continue with the proof of Lemma 15, a useful observation about the local restructuring of trees after an NNI move is:

Remark 5 *Let (σ,γ) be any edge in the NNI-graph $\mathcal{N}_S = (\mathcal{BT}_S, \mathcal{E})$. Then, the NNI move on σ at $G \in \mathcal{G}(\sigma)$ joining γ changes the hierarchical organization of common clusters descending the (grand) parent cluster $P = \mathrm{Pr}^2(G,\sigma) = \mathrm{Pr}(G,\gamma)$ of (grand) child G and keeps the remaining clusters unchanged with the property that for all $I \in \mathcal{C}(\sigma) \setminus \mathrm{Des}(P,\sigma) = \mathcal{C}(\gamma) \setminus \mathrm{Des}(P,\gamma)$*

$$\ell_\sigma(I) = \ell_\gamma(I). \tag{159}$$

We now continue with the change in the value of Lyapunov function after the NNI move on σ at G towards $\gamma = NNI(\sigma, G)$,

$$\begin{aligned}V_\tau(\gamma) - V_\tau(\sigma) = \sum_{J\in\mathcal{C}(\tau)} \frac{1}{\rho^{\ell_\tau(J)}}\Big(\Phi_\gamma\big(\mathrm{Des}(J,\gamma), J\big) \\ - \Phi_\sigma\big(\mathrm{Des}(J,\sigma), J\big)\Big).\end{aligned} \tag{160}$$

Using Remark 5, one can rewrite (160) in terms of (grand)parent cluster $P = \mathrm{Pr}^2(G,\sigma) = \mathrm{Pr}(G,\gamma)$ as

$$\begin{aligned}V_\tau(\gamma) - V_\tau(\sigma) = \sum_{J\in\mathcal{C}(\tau)} \frac{1}{\rho^{\ell_\tau(J)}}\Big(\Phi_\gamma\big(\mathrm{Des}(P,\gamma), J\big) \\ - \Phi_\sigma\big(\mathrm{Des}(P,\sigma), J\big)\Big).\end{aligned} \tag{161}$$

Let $Q \in \mathcal{C}(\tau)$ with the property that $P \subseteq Q$. Notice that every cluster J in $\mathcal{C}(\tau) \setminus \mathrm{Des}(Q,\tau)$ either contains or is disjoint with clusters $I \in \mathrm{Des}(P,\sigma)$ and $K \in \mathrm{Des}(P,\gamma)$, and so they are compatible, i.e. $I \bowtie J$ and $K \bowtie J$. As a result, (161) can be further simplified as

$$\begin{aligned}V_\tau(\gamma) - V_\tau(\sigma) = \sum_{J\in\mathrm{Des}(Q,\tau)} \frac{1}{\rho^{\ell_\tau(J)}}\Big(\Phi_\gamma\big(\mathrm{Des}(P,\gamma), J\big) \\ - \Phi_\sigma\big(\mathrm{Des}(P,\sigma), J\big)\Big).\end{aligned} \tag{162}$$

Now, using the bounds on $\Phi_\sigma\big(\mathrm{Des}(P,\sigma), J\big)$ in (158) and the property of hierarchical attenuation constant in Lemma 14, an upper bound on (162) can be obtained as

$$\begin{aligned}V_\tau(\gamma) - V_\tau(\sigma) < \sum_{J\in\mathrm{Des}(Q,\tau)} \frac{1}{\rho^{\ell_\tau(J)}}\Big(\frac{4}{\rho-2}\frac{1}{\rho^{\ell_\sigma(P)+1}} \\ + \Phi_\gamma\big(\mathrm{Ch}(P,\gamma), J\big) - \Phi_\sigma\big(\mathrm{Ch}(P,\sigma), J\big)\Big),\end{aligned} \tag{163}$$

$$\begin{aligned}< \frac{8}{(\rho-2)^2}\frac{1}{\rho^{\ell_\sigma(P)+\ell_\tau(Q)+1}} + \\ + \sum_{J\in\mathrm{Des}(Q,\tau)} \frac{1}{\rho^{\ell_\tau(J)}}\Big(\Phi_\gamma\big(\mathrm{Ch}(P,\gamma), J\big) - \Phi_\sigma\big(\mathrm{Ch}(P,\sigma), J\big)\Big),\end{aligned} \tag{164}$$

where $\ell_\sigma(P) = \ell_\gamma(P)$. Further, using the hierarchical relation between descendant clusters, we might obtain a looser upper bound in terms of clusters P and Q and their children as follows:

$$\begin{aligned}V_\tau(\gamma) - V_\tau(\sigma) &< \frac{8}{(\rho-2)^2}\frac{1}{\rho^{\ell_\sigma(P)+\ell_\tau(Q)+1}} \\ &+ \frac{1}{\rho^{\ell_\tau(Q)+1}} \sum_{J\in\mathrm{Ch}(Q,\tau)} \Big(\Phi_\gamma\big(\mathrm{Ch}(P,\gamma), J\big) - \Phi_\sigma\big(\mathrm{Ch}(P,\sigma), J\big)\Big) \\ &+ \underbrace{\sum_{J\in\mathrm{Des}(Q_L,\tau)} \frac{1}{\rho^{\ell_\tau(J)}}}_{< \frac{2}{\rho-2}\frac{1}{\rho^{\ell_\tau(Q)+1}},\ \text{by Lemma 14}} \underbrace{\Big(\Phi_\gamma\big(\mathrm{Ch}(P,\gamma), J\big) - \Phi_\sigma\big(\mathrm{Ch}(P,\sigma), J\big)\Big)}_{\leq \frac{2}{\rho^{\ell_\sigma(P)+1}},\ \text{from (155) and } \ell_\sigma(P)=\ell_\gamma(P)} \\ &+ \underbrace{\sum_{J\in\mathrm{Des}(Q_R,\tau)} \frac{1}{\rho^{\ell_\tau(J)}}}_{< \frac{2}{\rho-2}\frac{1}{\rho^{\ell_\tau(Q)+1}},\ \text{by Lemma 14}} \underbrace{\Big(\Phi_\gamma\big(\mathrm{Ch}(P,\gamma), J\big) - \Phi_\sigma\big(\mathrm{Ch}(P,\sigma), J\big)\Big)}_{\leq \frac{2}{\rho^{\ell_\sigma(P)+1}},\ \text{from (155) and } \ell_\sigma(P)=\ell_\gamma(P)},\end{aligned} \tag{165}$$

$$\begin{aligned}V_\tau(\gamma) - V_\tau(\sigma) < \frac{1}{\rho^{\ell_\tau(Q)+\ell_\sigma(P)+2}}\Bigg(\frac{16(\rho-1)}{(\rho-2)^2} \\ + \sum_{\substack{K\in\mathrm{Ch}(P,\gamma)\\ J\in\mathrm{Ch}(Q,\tau)}} \mathbb{1}(K \not\bowtie J) - \sum_{\substack{I\in\mathrm{Ch}(P,\sigma)\\ J\in\mathrm{Ch}(Q,\tau)}} \mathbb{1}(I \not\bowtie J)\Bigg).\end{aligned} \tag{166}$$

Relation Between the Lyapunov Function and NNI Control Law:

For any $\sigma \neq \tau \in \mathcal{BT}_S$, let $K \in \mathcal{K}(\sigma,\tau)$ and $I \in \mathcal{D}(\sigma,\tau;K)$ be the clusters selected by the NNI control policy $\mathbf{u}_\tau$ of Section 4.3 while determining the NNI move on σ at $G \in \mathrm{Ch}(I,\sigma)$ towards $\gamma = (\mathrm{NNI} \circ \mathbf{u}_\tau)(\sigma)$. Here, note that $P = \mathrm{Pr}(I,\sigma) = \mathrm{Pr}(G,\gamma) \subseteq K$.

Let $\Psi_{\sigma,\tau}(P,K)$ denote the total number of crossings between children of P in σ and children of K in τ,

$$\Psi_{\sigma,\tau} P, K := \sum_{\substack{I\in\mathrm{Ch}(P,\sigma)\\ J\in\mathrm{Ch}(K,\tau)}} \mathbb{1}(I \not\bowtie J). \tag{167}$$

Accordingly, the upper bound in the change of Lyapunov function (166) after the aforementioned NNI move can be rewritten as a function of $\Psi_{\sigma,\tau}(P,K)$ and $\Psi_{\gamma,\tau}(P,K)$ (167):

$$V_\tau(\gamma)-V_\tau(\sigma) < \frac{1}{\rho^{\ell_\tau(K)+\ell_\sigma(P)+2}}\left(\frac{16(\rho-1)}{(\rho-2)^2} + \Psi_{\gamma,\tau}(P,K)-\Psi_{\sigma,\tau}(P,K)\right). \quad (168)$$

To complete the proof of Lemma 15, we shall show that

$$\Psi_{\gamma,\tau}(P,K)-\Psi_{\sigma,\tau}(P,K) \leq -1. \quad (169)$$

Depending on the incompatibility of I with $\mathrm{Ch}(K,\tau)$, the values of $\Psi_{\sigma,\tau}(P,K)$ and $\Psi_{\gamma,\tau}(P,K)$ can be bounded from above as follows:

- Case 1 : I is Type 1 :
  If $I \in \mathcal{D}(\sigma,\tau;K)$ is Type 1, then, by Definition 13, $I \not\bowtie \mathrm{Ch}(K,\tau)$ and $I^{-\sigma} \bowtie \mathrm{Ch}(K,\tau)$. Recall that $\{I, I^{-\sigma}\} = \mathrm{Ch}(P,\sigma)$. Hence, one can obtain that

$$\Psi_{\sigma,\tau}(P,K) = \sum_{F\in \mathrm{Ch}(K,\tau)} \underbrace{\sum_{E\in \mathrm{Ch}(P,\sigma)} \mathbb{1}(E \not\bowtie F)}_{=\mathbb{1}(I\not\bowtie F),\text{ since } I^{-\sigma}\bowtie \mathrm{Ch}(K,\tau)}, \quad (170)$$

$$= \underbrace{\sum_{F\in \mathrm{Ch}(K,\tau)} \mathbb{1}(I \not\bowtie F)}_{\geq 1,\text{ since } I\not\bowtie \mathrm{Ch}(K,\tau)} \geq 1. \quad (171)$$

  Further, the NNI control rule $\mathbf{u}_\tau$ replaces I by $J = \mathrm{Pr}^2(G,\sigma)\setminus G$ whose local complement $J^{-\gamma}$ in γ is G. Note that J and $J^{-\gamma}$ are both compatible with $\mathrm{Ch}(K,\tau)$. Note that $\{J, J^{-\gamma}\} = \mathrm{Ch}(P,\gamma)$. Therefore,

$$\Psi_{\gamma,\tau}(P,K) = \sum_{\substack{D\in \mathrm{Ch}(P,\gamma)\\ F\in \mathrm{Ch}(K,\tau)}} \underbrace{\mathbb{1}(D\not\bowtie F)}_{=0,\text{since } D\bowtie \mathrm{Ch}(K,\tau)} = 0. \quad (172)$$

  As a result, for a Type 1 cluster, we always have

$$\Psi_{\gamma,\tau}(P,K) - \Psi_{\sigma,\tau}(P,K) \leq -1. \quad (173)$$

- Case 2 - I is Type 2 :
  In this case, by Definition 13, siblings $I, I^{-\sigma} \in \mathcal{D}(\sigma,\tau;K)$ are both incompatible with $\mathrm{Ch}(K,\tau)$. In fact, by Lemma 8, for all $E \in \{I, I^{-\sigma}\} = \mathrm{Ch}(P,\sigma)$ and $F \in \mathrm{Ch}(K,\tau)$ we have $E \not\bowtie F$. Thus,

$$\Psi_{\sigma,\tau}(P,K) = \sum_{\substack{E\in \mathrm{Ch}(P,\sigma)\\ F\in \mathrm{Ch}(K,\tau)}} \underbrace{\mathbb{1}(E\not\bowtie F)}_{=1,\text{by Lemma 8}} = 4. \quad (174)$$

  On the other hand, any arbitrary NNI move $G \in \mathrm{Ch}(I,\sigma)$ replaces cluster I by $J = \mathrm{Pr}^2(G,\sigma)\setminus G$ incompatible with $\mathrm{Ch}(K,\tau)$. Note that its sibbling $J^{-\gamma}$ in γ is G and compatible with split $\mathrm{Ch}(K,\tau)$. Hence, we have $J \not\bowtie \mathrm{Ch}(K,\tau)$ and $J^{-\gamma} \bowtie \mathrm{Ch}(K,\tau)$ for children clusters $\{J, J^{-\gamma}\} = \mathrm{Ch}(P,\gamma)$, which yields

$$\Psi_{\gamma,\tau}(P,K) = \sum_{F\in \mathrm{Ch}(K,\tau)} \underbrace{\sum_{D\in \mathrm{Ch}(P,\gamma)} \mathbb{1}(D \not\bowtie F)}_{=\mathbb{1}(J\not\bowtie F),\text{ since } J^{-\gamma}\bowtie \mathrm{Ch}(K,\tau)}, \quad (175)$$

$$= \sum_{F\in \mathrm{Ch}(K,\tau)} \mathbb{1}(J \not\bowtie F) \leq 2. \quad (176)$$

  Therefore, for a Type 2 cluster, we always have

$$\Psi_{\gamma,\tau}(P,K) - \Psi_{\sigma,\tau}(P,K) \leq -2. \quad (177)$$

To sum up, the NNI control policy $\mathbf{u}_\tau$ always guarantees that $\Psi_{\gamma,\tau}(P,K) - \Psi_{\sigma,\tau}(P,K) \leq -1$ after each evolution of the dynamical system (69) at every σ away from τ. This completes the proof. □

A.9 Proof of Theorem 2

Lemma 16 *Let $\{K_L, K_R\}$ be a bipartition of a fixed finite set K, and I and A be sets with the property that $I \subseteq A, I \subsetneq K$. Then the following equivalence holds*

$$I \bowtie \{K_L, K_R\} = I \bowtie \{K_L, K_R\}\big|_A, \quad (178)$$

where $\{K_L, K_R\}\big|_A$ is the restriction of $\{K_L, K_R\}$ to A (Definition 6).

Proof: Note that for any sets $X \subseteq Z$ and Y we always have $X \subseteq Y \Leftrightarrow X \subseteq Y \cap Z$. Accordingly, one can obtain the result using Lemma 7 as follows,

$$I \bowtie \{K_L, K_R\} \Longleftrightarrow \underbrace{(I\subseteq K_L)}_{\Leftrightarrow I\subseteq K_L\cap A} \vee \underbrace{(I\subseteq K_R)}_{\Leftrightarrow I\subseteq K_R\cap A}, \quad (179)$$

$$\Longleftrightarrow I \bowtie \{K_L\cap A, K_R\cap A\}. \quad (180)$$

$$\Longleftrightarrow I \bowtie \underbrace{\left(\{K_L\cap A, K_R\cap A\}\setminus\{\emptyset\}\right)}_{=\{K_L,K_R\}|_A}, \quad (181)$$

$$\Longleftrightarrow I \bowtie \{K_L, K_R\}\big|_A. \quad (182)$$

□

Lemma 17 *Let $\{K_L, K_R\}$ and $\{K_L^*, K_R^*\}$ be two bipartitions a fixed finite set K. The sum of compatible elements of one bipartition with the other bipartition is symmetric,*

$$\sum_{I\in\{K_L,K_R\}} \mathbb{1}\left(I \bowtie \{K_L^*, K_R^*\}\right) = \sum_{I^*\in\{K_L^*,K_R^*\}} \mathbb{1}\left(I^* \bowtie \{K_L, K_R\}\right), \quad (183)$$

and is only zero when bipartitions are the same.

Proof: If the bipartitions are the same, both sides of (183) simply sum to two.

Otherwise, since $\{K_L, K_R\}$ and $\{K_L^*, K_R^*\}$ are distinct binary partitions of K, at most an element of $\{K_L, K_R\}$ is a proper subset of an element of $\{K_L^*, K_R^*\}$ and vice versa. One way to observe this is a proof by contradiction. Let each element of $\{K_L, K_R\}$ is a proper subset of an element of $\{K_L^*, K_R^*\}$, then $K_L \subsetneq K_L^* \cup K_R^*$ and $K_R \subsetneq K_L^* \cup K_R^*$. Hence, we have $K_L \cup K_R \subsetneq K_L^* \cup K_R^* = K$, which is a contradiction.

Now, if, without loss of generality, $K_L \subsetneq K_L^*$, that is to say $K_R^* \subsetneq K_R$, $K_R \not\subseteq K_L^*$ and $K_L^* \not\subseteq K_R$, then using Lemma 7 one can easily verify that both sides of (183) sum to one. Otherwise (none of elements of a bipartition are a proper subset any element of other bipartition), the summations on both side of (183) are equal to zero since every pair of elements of the bipartitions are not subset of each other and so are incompatible. This completes the proof. ☐

Lemma 18 *For any $\sigma, \tau \in \mathcal{BT}_S$, the number of special crossings $\kappa_{\sigma,\tau}(I,J)$ (104) of children of $I \in \mathcal{C}(\sigma)$ in σ with children of $J \in \mathcal{C}(\tau)$ in τ is symmetric,*

$$\kappa_{\sigma,\tau}(I,J) = \kappa_{\tau,\sigma}(J,I)\,, \tag{184}$$

Proof: Consider a special case where at least one of $\mathrm{Ch}(I,\sigma)\big|_J = \{I \cap J, \emptyset\}$ and $\mathrm{Ch}(J,\tau)\big|_I = \{J \cap I, \emptyset\}$ holds. Then, it is clear that $\kappa_{\sigma,\tau}(I,J) = \kappa_{\tau,\sigma}(J,I) = 0$ since the empty cluster and $I \cap J$ are always compatible with any cluster $A \subseteq I \cap J$.

Otherwise, observe that $\mathrm{Ch}(I,\sigma)\big|_J = \mathrm{Ch}\big(I \cap J, \sigma|_{I\cap J}\big)$ and $\mathrm{Ch}(J,\tau)\big|_I = \mathrm{Ch}\big(I \cap J, \tau|_{I\cap J}\big)$, and so (184) takes the specific form in Lemma 17, which completes the proof. ☐

Proof of Theorem 2: Consider an NNI navigation path starting at σ and ending at τ. Let $\mathcal{G} = \big(G^1, G^2, \ldots, G^{|\zeta|-1}\big)$ and $\mathcal{K} = \big(K^1, K^2, \ldots, K^{|\zeta|-1}\big)$ be the ordered set of grandchild and common clusters with crossing splits selected by the NNI control law while constructing ζ.

Now observe that one can reorder $\mathcal{G}$ and $\mathcal{K}$ such that:

- the order relation of any pair of grandchildren in $\mathcal{G}$ associated with the same common cluster $K \in \mathcal{K}$ is preserved, which is required to define another valid NNI navigation path.
- any grandchild associated with S, if $S \in \mathcal{K}$, is less than any grandchild associated with a common cluster in $\mathcal{K} \setminus \{S\}$.

For instance, consider the following ordering of elements of $\mathcal{G}$: for any $i, j \in [1, |\zeta|-1]$

$$G^i \le G^j \iff \begin{array}{c} \big(K^i = S, K^j \neq S\big) \text{ or} \\ \big(i \le j, K^i = S, K^j = S\big) \text{ or} \\ \big(i \le j, K^i \neq S, K^j \neq S\big)\,. \end{array} \tag{185}$$

Let $\tilde{\mathcal{G}} = \Big(\tilde{G}^1, \tilde{G}^2, \ldots, \tilde{G}^{|\zeta|-1}\Big)$ denotes the reordering of $\mathcal{G}$ based on the ordering relation in (185), and $\tilde{\mathcal{K}} = \Big(\tilde{K}^1, \tilde{K}^2, \ldots, \tilde{K}^{|\zeta|-1}\Big)$ is the associated set of common clusters in $\mathcal{K}$ in the appropriate order, i.e. for any $i, j \in [1, |\zeta|-1]$, $K^i \le K^j \iff \tilde{G}^i \le \tilde{G}^j$. Notice that, by decomposability property of the NNI control law, the NNI moves associated with $\tilde{\mathcal{G}}$ defines an NNI navigation path $\tilde{\zeta} \in \Gamma(\sigma,\tau)$ joining σ to τ. Remark that ζ and $\tilde{\zeta}$ might be different, but they have the same length.

Hence, all NNI navigation paths from σ to τ can be rearrange in an appropriate way so that they first solve the incompatibilities of σ with the root split $\mathrm{Ch}(S,\tau)$ of τ and continue with joining subtrees of the root. In other words, using the decomposability property of the NNI control law, the length of any NNI navigation path ζ joining σ to τ can be recursively obtained as

$$\begin{aligned} |\zeta| - 1 = d_{Nav}\big(\sigma, \mathcal{BT}_{\mathrm{Ch}(S,\tau)}\big) + d_{Nav}\Big(\sigma\big|_{S_L}, \tau\big|_{S_L}\Big) \\ + d_{Nav}\Big(\sigma\big|_{S_R}, \tau\big|_{S_R}\Big)\,, \end{aligned} \tag{186}$$

where $\{S_L, S_R\} = \mathrm{Ch}(S,\tau)$.

Moreover, using Lemma 9, we can rewrite (186) as follows,

$$|\zeta| - 1 = \sum_{\substack{J \in \mathcal{C}(\tau) \\ I \in \mathcal{C}\left(\sigma|_J\right)}} \Big(\eta \circ \kappa_{\sigma|_J, \mathrm{Ch}(J,\tau)}\Big)(I)\,, \tag{187}$$

where $\kappa_{\sigma,\mathrm{Ch}(J,\tau)}(I)$ (92) determines the number of children of $I \in \mathcal{C}(\sigma)$ in σ incompatible with $\mathrm{Ch}(J,\tau)$ and η (93) returns the required number of NNI moves to resolve these cluster incompatibilities.

Let $J \in \mathcal{C}(\tau)$ with $|J| \ge 2$. As discussed in the proof of Lemma 2, an interior cluster K of $\sigma\big|_J$ is associated with a unique cluster I of σ such that $K = I \cap J$ and $\mathrm{Ch}(I,\sigma)\big|_J \neq \{I \cap J, \emptyset\}$. Hence, observe that $\mathcal{C}\left(\sigma\big|_J\right)$ can be written as

$$\begin{aligned} \mathcal{C}\left(\sigma|_J\right) &= \Big\{ I \cap J \Big| I \in \mathcal{C}(\sigma),\ I \cap J \neq \emptyset \Big\}\,, && (188) \\ &= \Big\{ I \cap J \Big| I \in \mathcal{C}(\sigma), \mathrm{Ch}(I,\sigma)\big|_J \neq \{I \cap J, \emptyset\} \Big\} \cup \bigcup_{i \in J} \{i\}. && (189) \end{aligned}$$

Accordingly, let $\mathcal{C}_J(\sigma)$ denote the set of clusters of σ defining the interior clusters of $\sigma\big|_J$,

$$\mathcal{C}_J(\sigma) = \Big\{ I \in \mathcal{C}(\sigma) \,\Big|\, \mathrm{Ch}(I,\sigma)\big|_J \neq \{I \cap J, \emptyset\} \Big\}\,. \tag{190}$$

Now, we can rewrite (187) as

$$\begin{aligned}
|\zeta| - 1 &= \sum_{\substack{J \in \mathcal{C}(\tau) \\ I \in \mathcal{C}\left(\sigma|_J\right)}} \eta \Bigg(\sum_{A \in \mathrm{Ch}\left(I, \sigma|_J\right)} \mathbb{1}\big(\underbrace{A \not\bowtie \mathrm{Ch}(J,\tau)}_{\substack{= A \not\bowtie \mathrm{Ch}(J,\tau)|_I \\ \text{by Lemma 16}}}\big) \Bigg), && (191) \\
&= \sum_{\substack{J \in \mathcal{C}(\tau) \\ I \in \mathcal{C}\left(\sigma|_J\right)}} \eta \Bigg(\underbrace{\sum_{A \in \mathrm{Ch}\left(I, \sigma|_J\right)} \mathbb{1}\big(A \not\bowtie \mathrm{Ch}(J,\tau)\big|_I\big)}_{\substack{=0 \text{ if } |I|=1, \\ \text{otherwise there is exactly one } \bar{I} \in \mathcal{C}_J(\sigma) \text{ s.t.} \\ \mathrm{Ch}\left(I, \sigma|_J\right) = \mathrm{Ch}(\bar{I},\sigma)|_J}} \Bigg), && (192) \\
&= \sum_{\substack{J \in \mathcal{C}(\tau) \\ I \in \mathcal{C}_J(\sigma)}} \eta \Bigg(\underbrace{\sum_{A \in \mathrm{Ch}(I,\sigma)|_J} \mathbb{1}\big(A \not\bowtie \mathrm{Ch}(J,\tau)\big|_I\big)}_{\substack{=0 \text{ for all } I \in \mathcal{C}(\sigma) \setminus \mathcal{C}_J(\sigma) \\ \text{since } \mathrm{Ch}(I,\sigma)|_J = \{I \cap J, \emptyset\}}} \Bigg), && (193) \\
&= \sum_{\substack{J \in \mathcal{C}(\tau) \\ I \in \mathcal{C}(\sigma)}} \eta \Bigg(\underbrace{\sum_{A \in \mathrm{Ch}(I,\sigma)|_J} \mathbb{1}\big(A \not\bowtie \mathrm{Ch}(J,\tau)\big|_I\big)}_{= \kappa_{\sigma,\tau}(I,J) (104)} \Bigg), && (194) \\
&= \sum_{\substack{J \in \mathcal{C}(\tau) \\ I \in \mathcal{C}(\sigma)}} \big(\eta \circ \kappa_{\sigma,\tau}\big)(I,J)\,, && (195) \\
&= \left\| \mathbf{S}(\sigma,\tau) \right\|_1. && (196)
\end{aligned}$$

Thus, in addition to (186), the length of any NNI navigation path in $\Gamma(\sigma,\tau)$ from σ to τ is equivalently given by $\left\|\mathbf{S}(\sigma,\tau)\right\|_1$. Similarly, the length of any NNI navigation path in $\Gamma(\sigma,\tau)$ from τ to σ is equal to

$\|\mathbf{S}(\tau,\sigma)\|_1$. Remark that for any binary tree $\sigma,\tau \in \mathcal{BT}_S$, by Lemma 18, $\|\mathbf{S}(\sigma,\tau)\|_1 = \|\mathbf{S}(\tau,\sigma)\|_1$. Therefore, the length of any NNI navigation path joining σ and τ is equal to $\|\mathbf{S}(\sigma,\tau)\|_1 = \|\mathbf{S}(\tau,\sigma)\|_1$, and the result follows. □

APPENDIX B
ADDITIONAL PROPERTIES OF THE DISCRETE DYNAMICS FOR BINARY SEARCH TREES

Definition 16 *For a totally ordered fixed finite set S, a nondegenerate hierarchy $\tau \in \mathcal{BT}_S$ is a* binary search tree (BST)[15] *if every cluster $I \in \mathcal{C}(\tau)$ with the property that $\min(I) < \min(I^{-\tau})$ satisfies $\max(I) < \min(I^{-\tau})$.*

We shall denote the set of binary search trees by $\mathcal{BST}_S$, which is, by definition, a subset of $\mathcal{BT}_S$.[16]

A distinctive feature of BSTs is:

Remark 6 *A nondegenerate hierarchy $\tau \in \mathcal{BT}_S$ over a fixed finite ordered set S is a BST if and only if its clusters are intervals of S (see Figure 1).*

We now continue with relations between the NNI control law of Section 4.3 and binary search trees. Let us start with an observation stating possible types of navigation moves in $\mathcal{BST}_S$:

Lemma 19 *(No Type 2 Crossings for BSTs) Let $\tau \in \mathcal{BST}_S$ be a BST over a fixed finite ordered index set S, and $\{S_L, S_R\}$ be a bipartition of S whose elements are intervals of S.*

If any cluster $I \in \mathcal{C}(\tau)$ is incompatible with $\{S_L, S_R\}$, then its sibling $I^{-\tau}$ is always compatible with $\{S_L, S_R\}$. That is to say, there exist no cluster of τ Type 2 incompatible with $\{S_L, S_R\}$.

Proof: Proof by contradiction.

Recall that I and $I^{-\tau}$ are disjoint intervals of S (Remark 6) and, without loss of generality, let $\max(S_L) < \min(S_R)$.

Now, suppose that both clusters $I, I^{-\tau} \in \mathcal{C}(\tau)$ are incompatible with $\{S_L, S_R\}$. Hence, by Lemma 7, $I \cap S_L \neq \emptyset$, $I \cap S_R \neq \emptyset$ and $I^{-\tau} \cap S_L \neq \emptyset$, $I^{-\tau} \cap S_R \neq \emptyset$. Thus, $\max(I) > \min(I^{-\tau})$ and $\min(I) < \max(I^{-\tau})$ which is a contradiction and completes the proof. □

Lemma 20 *The subspace $\mathcal{BST}_S$ of $\mathcal{BT}_S$ is positive invariant for the closed loop discrete dynamical system (70) obeying the NNI control law of Section 4.3.*

Proof: For distinct initial and desired hierarchies $\sigma \neq \tau \in \mathcal{BST}_S$, let $K \in \mathcal{K}(\sigma,\tau)$ be a common cluster of σ and τ with crossing splits and $I \in \mathcal{D}(\sigma,\tau;K)$ be a deep incompatible cluster of σ with $\{K_L, K_R\} = \mathrm{Ch}(K,\tau)$.

Now, the NNI move result from the control law $\mathbf{u}_\tau$ on σ at $G \in \mathrm{Ch}(I,\sigma)$ replaces $I = \mathrm{Pr}(G,\sigma)$ by $\mathrm{Pr}^2(G,\sigma) \setminus G$ in the next hierarchy $\gamma = \mathrm{NNI}(\sigma, G) \in \mathcal{BT}_S$. Note that I is Type 1 since there is no Type 2 incompatibility for BSTs (Lemma 19). Hence, $G \subseteq K_A$ and $G^{-\sigma}, I^{-\sigma} \subsetneq K_B$ for some $A \neq B \in \{L, R\}$. Thus, $\mathrm{Pr}^2(G,\sigma) \setminus G = G^{-\sigma} \cup I^{-\sigma} = \mathrm{Pr}^2(G,\sigma) \cap K_B$ which is another nonempty interval of S since clusters of σ and τ are intervals of the index set S (Remark 6). Therefore, the clusters of γ are also intervals of S, and so it is a BST, which completes the proof. □

APPENDIX C
AN EFFICIENTLY COMPUTABLE NNI NAVIGATION PATH

Lemma 21 *Let $\{S_L, S_R\}$ be a bipartition of S. An NNI navigation path starting from $\sigma \in \mathcal{BT}_S$ ending in $\mathcal{BT}_{\{S_L,S_R\}}$ can be computed in $\mathrm{O}(|S|)$ time.*

Proof: Let $\tau \in \mathcal{BT}_{\{S_L,S_R\}}$. All cluster of σ incompatible with $\{S_L, S_R\}$, i.e. $\mathcal{I}(\sigma,\tau;S)$, can be determined in $\mathrm{O}(|S|)$ time as discussed in the proof of Lemma 11. Hence, if $\sigma \notin \mathcal{BT}_{\{S_L,S_R\}}$, then a deep cluster $I \in \mathcal{D}(\sigma,\tau;S)$ incompatible with $\{S_L, S_R\}$ can be found in $\mathrm{O}(|S|)$ by a post-order traversal of σ.

Recall from Remark 4 that I and $I^{-\sigma}$ can be replaced by $\mathrm{Pr}(I,\sigma) \cap S_L$ and $\mathrm{Pr}(I,\tau) \cap S_R$, compatible with $\{S_L, S_R\}$, after a certain number of NNI moves as illustrated in Figure 8. Note that the rest of the clusters of σ are kept the same.

Accordingly, let $\gamma \in \mathcal{BT}_S$ denote the resulting intermediate tree with cluster set

$$\mathcal{C}(\gamma) = \mathcal{C}(\sigma) \setminus \{I, I^{-\sigma}\} \cup \{\mathrm{Pr}(I,\sigma) \cap S_L, \mathrm{Pr}(I,\sigma) \cap S_R\}. \tag{197}$$

Remark that $\mathcal{I}(\gamma,\tau;S) = \mathcal{I}(\sigma,\tau;S) \setminus \{I, I^{-\sigma}\}$. Now, instead of searching for a deep cluster in γ starting from the root S, using Corollary 4, one can continue the post-order search for a deep cluster in γ at $\mathrm{Pr}(I,\sigma)$ whose children, $\mathrm{Pr}(I,\sigma) \cap S_L$ and $\mathrm{Pr}(I,\tau) \cap S_R$, in γ are compatible with $\{S_L, S_R\}$.

In fact, observe that $\mathcal{I}(\sigma,\tau;S) \cup \{S\}$ defines a tree-like data structure (see Figure 6). Therefore, one can conclude that the overall construction of $\Gamma_\tau(\sigma)$ only requires complete post-order traversal of σ for $\mathcal{I}(\sigma,\tau;S)$ in $\mathrm{O}(|S|)$ time. □

Lemma 22 *An NNI navigation path joining $\sigma \in \mathcal{BT}_S$ to $\tau \in \mathcal{BT}_S$ consistent with the NNI control $\mathbf{u}_\tau$ can be computed in $\mathrm{O}\left(|S|^2\right)$ time.*

Proof: Similar to the recursive expression of d_{Nav} (105a), 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 $\mathcal{BT}_{\{S_L,S_R\}}$ in $\mathrm{O}(|S|)$

15. Note that the sequence of leaves visited during any depth-first tree traversal of a BST is the same as the order of leaf set S.

16. The subgraph of the NNI-graph, defined in Section 2.2.2, containing only BSTs is known as the rotation graph, and the associated operations between BSTs are called *rotations* [27].

(Lemma 21) and then find NNI navigation paths between subtrees. Hence, this requires a 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 $O\left(|S|^2\right)$ time, which completes the proof. $\square$

ACKNOWLEDGMENTS

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

REFERENCES

[1] D. F. Robinson and L. R. Foulds, "Comparison of phylogenetic trees," *Mathematical Biosciences*, vol. 53, no. 1-2, pp. 131 – 147, 1981. 1, 5

[2] W. H. E. Day, "Optimal algorithms for comparing trees with labeled leaves," *Journal of Classification*, vol. 2, pp. 7–28, 1985, 10.1007/BF01908061. 1, 11

[3] D. Bogdanowicz and K. Giaro, "Matching split distance for unrooted binary phylogenetic trees," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 9, no. 1, pp. 150–160, jan.-feb. 2012. 1, 2, 5, 6

[4] Y. Lin, V. Rajan, and B. Moret, "A metric for phylogenetic trees based on matching," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 9, no. 4, pp. 1014–1022, july-aug. 2012. 1, 5, 6, 16

[5] D. Robinson, "Comparison of labeled trees with valency three," *Journal of Combinatorial Theory, Series B*, vol. 11, no. 2, pp. 105 – 119, 1971. 1, 4

[6] G. Moore, M. Goodman, and 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*, vol. 38, no. 3, pp. 423 – 457, 1973. 1, 4

[7] B. L. Allen and M. Steel, "Subtree transfer operations and their induced metrics on evolutionary trees," *Annals of Combinatorics*, vol. 5, pp. 1–15, 2001, 10.1007/s00026-001-8006-8. 1

[8] J. Felsenstein, *Inferring Phylogenies*. Suderland, USA: Sinauer Associates, 2004. 1, 2

[9] B. Dasgupta, X. He, T. Jiang, M. Li, J. Tromp, and L. Zhang, "On computing the nearest neighbor interchange distance," in *In Proc. DIMACS Workshop on Discrete Problems with Medical Applications*. Press, 1997, pp. 125–143. 1

[10] M. Li, J. Tromp, and L. Zhang, "On the nearest neighbour interchange distance between evolutionary trees," *J. Theor. Biol*, pp. 463–467, 1996. 2, 5

[11] K. CulikII and D. Wood, "A note on some tree similarity measures," *Information Processing Letters*, vol. 15, no. 1, pp. 39 – 42, 1982. 2, 5

[12] E. K. Brown and W. H. E. Day, "A computationally efficient approximation to the nearest neighbor interchange metric," *Journal of Classification*, vol. 1, pp. 93–124, 1984, 10.1007/BF01890118. 2, 5

[13] D. Bryant, "A classification of consensus methods for phylogenetics," *DIMACS series in discrete mathematics and theoretical computer science*, vol. 61, pp. 163–184, 2003. 2, 11

[14] T. Margush and F. R. McMorris, "Consensus n-trees," *Bulletin of Mathematical Biology*, vol. 43, no. 2, pp. 239–244, 1981. 2

[15] J.-P. Barthélemy and F. McMorris, "The median procedure for n-trees," *Journal of Classification*, vol. 3, no. 2, pp. 329–334, 1986. 2

[16] F. James Rohlf, "Consensus indices for comparing classifications," *Mathematical Biosciences*, vol. 59, no. 1, pp. 131–144, 1982. 2, 3

[17] K. Bremer, "Combinable component consensus," *Cladistics*, vol. 6, no. 4, pp. 369–372, 1990. 2, 3

[18] O. Arslan, D. P. Guralnik, and 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. [Online]. Available: http://kodlab.seas.upenn.edu/Main/Allerton2012 2

[19] N. Ayanian, V. Kumar, and D. Koditschek, "Synthesis of controllers to create, maintain, and reconfigure robot formations with communication constraints," in *Robotics Research*, ser. Springer Tracts in Advanced Robotics, C. Pradalier, R. Siegwart, and G. Hirzinger, Eds. Springer Berlin / Heidelberg, 2011, vol. 70, pp. 625–642. 2

[20] L. J. Billera, S. P. Holmes, and K. Vogtmann, "Geometry of the space of phylogenetic trees," *Advances in Applied Mathematics*, vol. 27, no. 4, pp. 733 – 767, 2001. 2, 4

[21] A. Schrijver, *Combinatorial optimization: polyhedra and efficiency*. Springer, 2003, vol. 24. 2, 3

[22] K. Vogtmann, "Geodesics in the space of trees." 2007, accessed May. 05, 2013. [Online]. Available: www.math.cornell.edu/~vogtmann/papers/TreeGeodesicss/index.html 3

[23] M. Owen and J. S. Provan, "A fast algorithm for computing geodesic distances in tree space," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 8, no. 1, pp. 2 –13, jan.-feb. 2011. 3

[24] G. Carlsson and F. Mémoli, "Characterization, Stability and Convergence of Hierarchical Clustering methods," *Journal of Machine Learning Research*, vol. 11, pp. 1425–1470, 2010. 5

[25] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*. Prentice-Hall, Inc., 1988. 5

[26] R. Rammal, G. Toulouse, and M. A. Virasoro, "Ultrametricity for physicists," *Reviews of Modern Physics*, vol. 58, no. 3, p. 765, 1986. 5

[27] D. D. Sleator, R. E. Tarjan, and W. P. Thurston, "Rotation distance, triangulations, and hyperbolic geometry," in *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, ser. STOC '86. New York, NY, USA: ACM, 1986, pp. 122–135. 6, 23

[28] M. Waterman and T. Smith, "On the similarity of dendrograms," *Journal of Theoretical Biology*, vol. 73, no. 4, pp. 789 – 800, 1978. 10

[29] A. Feragen, M. Owen, J. Petersen, M. Wille, L. Thomsen, A. Dirksen, and M. Bruijne, "Tree-space statistics and approximations for large-scale analysis of anatomical trees," in *Information Processing in Medical Imaging*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, vol. 7917, pp. 74–85. 11

[30] J. Chakerian and S. Holmes, "Computational tools for evaluating phylogenetic and hierarchical clustering trees," *Journal of Computational and Graphical Statistics*, vol. 21, no. 3, pp. 581–599, 2012. 11

[31] H. Khalil, *Nonlinear systems, 3rd*. New Jewsey, Prentice Hall, 2002. 11

[32] C. Semple and M. Steel, *Phylogenetics*. Oxford University Press, 2003, vol. 24. 17