Recognizing Treelike $k$-Dissimilarities

Sven Herrmann
University of East Anglia, UK
Katharina T. Huber
University of East Anglia, UK
Vincent Moulton
University of East Anglia, UK
Andreas Spillner
University of Greifswald, Germany

Abstract: A $k$-dissimilarity $D$ on a finite set $X$, $|X| \geq k$, is a map from the set of size $k$ subsets of $X$ to the real numbers. Such maps naturally arise from edge-weighted trees $T$ with leaf-set $X$: Given a subset $Y$ of $X$ of size $k$, $D(Y)$ is defined to be the total length of the smallest subtree of $T$ with leaf-set $Y$. In case $k = 2$, it is well-known that 2-dissimilarities arising in this way can be characterized by the so-called “4-point condition”. However, in case $k > 2$ Pachter and Speyer (2004) recently posed the following question: Given an arbitrary $k$-dissimilarity, how do we test whether this map comes from a tree? In this paper, we provide an answer to this question, showing that for $k \geq 3$ a $k$-dissimilarity on a set $X$ arises from a tree if and only if its restriction to every $2k$-element subset of $X$ arises from some tree, and that $2k$ is the least possible subset size to ensure that this is the case. As a corollary, we show that there exists a polynomial-time algorithm to determine when a $k$-dissimilarity arises from a tree. We also give a 6-point condition for determining when a 3-dissimilarity arises from a tree, that is similar to the aforementioned 4-point condition.

Keywords: $k$-dissimilarity; Phylogenetic tree; Dissimilarity; Metric; 4-point condition; Ultrametric condition; Equidistant tree.

K.T.H., S.H. and V.M. thank Charles Semple and Mike Steel for their hospitality at University of Canterbury, where part of this work was undertaken. V.M. also thanks the Royal Society for its support. S.H. also thanks the German Academic Exchange Service (DAAD) for its support by a fellowship within its Postdoc-Programme. The authors want to thank the anonymous referees for their helpful comments.

Authors' Addresses: S. Herrmann, K.T. Huber, and V. Moulton, School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ UK; Andreas Spillner, Institut für Mathematik und Informatik, Universität Greifswald, Walther-Rathenau-Straße 47, 17487 Greifswald, Germany; e-mail: sherrmann@mathematik.tu-darmstadt.de.

Published online 4 September 2012
1. Introduction

In phylogenetics, as well as other areas making use of classification techniques, many distance-based methods for constructing trees are based on the following fundamental observation. For $X$ a non-empty finite set, a graph-theoretical tree $T = (V, E)$ with leaf-set $X \subseteq V$ and non-negative edge-weighting $\omega : E \to \mathbb{R}$ can be encoded in terms of the restriction of the pairwise dissimilarity $d_{(T,\omega)}$ to $X$, where $d_{(T,\omega)}(u, v)$ denotes the length of the shortest path in $T$ between $u$ and $v$ ($u, v \in V$). In other words, the tree $T$ can be completely recovered from the matrix of pairwise values $(d_{(T,\omega)})_{x,y\in X}$. Such dissimilarities are commonly called “tree metrics” and there is an extensive literature concerning their properties (see e.g. Semple and Steel 2003 and Gordon 1987 for overviews).

Various methods have been proposed for constructing trees that exploit this observation. These essentially work by projecting an arbitrary pairwise dissimilarity onto some “nearby” tree metric (see e.g. Felsenstein 2003; de Soete 1983). Even so, it is well-known that such methods can suffer from the fact that pairwise distance estimates involve some loss of information (see e.g. p. 176 in Felsenstein 2003). As a potential solution to this problem, Pachter and Speyer (2004) proposed using $k$-wise distance estimates, $k \geq 3$, to reconstruct trees, an approach which they subsequently implemented in Levy, Yoshida, and Pachter (2006) (see also Grishin 1999 where a related idea was investigated). Their rationale was that $k$-wise estimates are potentially more accurate since they can capture more information than pairwise distances, a point that was also made in Chapter 12 of Felsenstein (2003).

To describe Pachter and Speyer’s approach, recall that a phylogenetic tree (on $X$) is a graph-theoretical tree $T = (V, E)$ in which every non-leaf vertex has degree at least three and whose leaf-set is $X$ (cf. Figure 1(a)). In case a real-valued weight $\omega(e)$ is associated to every edge $e$ of $T$, we call $T$ a weighted phylogenetic tree, and we usually denote such a tree by $(T, \omega)$. Now, for any $k$-element subset $Y \subseteq X$, $k \geq 2$, let $D_{(T,\omega)}^k(Y)$ denote the total edge-weight of the smallest subtree of $T$ with leaf-set $Y$ (cf. Figure 1(b)). Note that this quantity is sometimes called the “phylogenetic diversity” of $Y$ (see e.g. Faith 1992 and Steel 2005) and that, for $k = 2$, $D_{(T,\omega)}^2(\{x,y\}) = d_{(T,\omega)}(x, y)$ for all $x, y \in X$.

In Pachter and Speyer (2004), the following result is proven:

**Theorem 1.** Let $(T, \omega)$ be a weighted phylogenetic tree on $X$ with $\omega$ non-negative and $\omega(e) > 0$ for every edge $e$ of $T$ that is not incident to a leaf, $|X| = n$, and $k \geq 2$ be some integer. If $n \geq 2k - 1$, then $(T, \omega)$ is determined by the map $D_{(T,\omega)}^k$ (and it is not if $2k - 2 = n > 2$).
Recognizing Treelike $k$-Dissimilarities

In other words, just as in the case $k = 2$, for $k \geq 3$ it is possible to recover $(T, \omega)$ from the function $D^k_{(T, \omega)}$ that maps the set of subsets of $X$ of size $k$ (denoted $\binom{X}{k}$) to $\mathbb{R}$. Here we call any map $D : \binom{X}{k} \rightarrow \mathbb{R}$ a $k$-dissimilarity. Note that 3-dissimilarities have been investigated, for example, in Hayashi (1972), Joly and Le Calvé (1995) and Heiser and Bennani (1997), and arbitrary $k$-dissimilarities in Deza and Rosenberg (2000) and Warrens (2010), under names such as $k$-way dissimilarities, $k$-way distances and $k$-semimetrics (see also Bandelt and Dress 1994 for related work).

In this paper we shall provide a solution to the following problem raised in Pachter and Speyer (2004):

“However, if we are simply given a $k$-dissimilarity map $D : \binom{X}{k} \rightarrow \mathbb{R}$, we do not know how to test whether this map comes from a phylogenetic tree.”

Note that Dress and Steel (2007) study the related problem of characterizing when a map $D$ from the set of subsets of $X$ of size at most $k$ into some Abelian group $G$ can be represented by a phylogenetic tree on $X$ whose edges are assigned elements from $G$. However, we consider subsets of $X$ of size precisely $k$, leading to a quite different characterization.

In order to state the main result of this paper, we first recall some more definitions concerning phylogenetic trees. A rooted phylogenetic tree (on $X$), is a tree $T = (V, E)$ with (i) a distinguished vertex $\rho$, called the root of $T$, that has degree at least 2, (ii) leaf-set $X$ and (iii) no vertex in $V \setminus (X \cup \{\rho\})$ with degree less than 3. In case a real-valued weight $\omega(e)$ is associated to every edge $e \in E$, we call $T$ a weighted, rooted phylogenetic tree, and denote it by $(T, \rho, \omega)$. Note that, for such a tree, we define the maps $d_{(T, \omega)}$ and $D^k_{(T, \omega)}$ in the same way as for (unrooted) phylogenetic trees (cf. Figure 1(c)). In addition, we call an edge-weighting $\omega$ of $T$ equidistant if (i) $d_{(T, \omega)}(x, \rho) = d_{(T, \omega)}(x', \rho)$ for all $x, x' \in X$, and (ii) $d_{(T, \omega)}(x, u) \leq
$d_{(T,\omega)}(x,v)$ for all $x \in X$ and any $u, v \in V$ that lie on the path from $x$ to $\rho$ in $T$ which first meets $u$ and then $v$ (cf. Figure 1(c)). Such weightings commonly arise when modeling sequence evolution assuming a molecular clock (see e.g. Felsenstein 2003).

Now, we call a $k$-dissimilarity $D$ treelike if there exists a weighted phylogenetic tree $(T,\omega)$ with $\omega$ non-negative such that $D = D^k_{(T,\omega)}$ holds, and we call $D$ equidistant if there exists a weighted, rooted phylogenetic tree $(T,\rho,\omega)$ on $X$ with $\omega$ equidistant such that $D = D^k_{(T,\omega)}$ holds. In this paper, we shall prove the following:

**Theorem 2.** Let $k \geq 2$ and $D$ be a $k$-dissimilarity map on a set $X$ with $|X| \geq 2k$. Then $D$ is treelike/equidistant if and only if the restriction of $D$ to every $2k$-element subset of $X$ is treelike/equidistant. Moreover, for all $k \geq 3$ there exist $k$-dissimilarity maps whose restrictions to every $(2k-1)$-element subset of $X$ are treelike/equidistant but that are not treelike/equidistant.

Note that for the case $k = 2$ this result is well-known (see e.g. Semple and Steel 2003, Theorem 7.2.5 and Corollary 7.2.7.)

After presenting some preliminaries in the next section, we prove Theorem 2 in Sections 3 and 4. As a corollary of Theorem 2, we also show that, for fixed $k \geq 2$, there is an algorithm with run-time that is bounded by a polynomial in $|X|$ to decide if an arbitrary $k$-dissimilarity $D$ is treelike (Corollary 1). It would be interesting to know if such algorithms can be found that have good run-time bounds for $k \geq 3$, such as those that have been devised for $k = 2$ (see e.g. Culberson and Rudnicki 1989, Bandelt 1990). More generally, it might also be of interest to use Theorem 2 to help devise new methods to construct trees from $k$-dissimilarities such as the one described in Levy, Yoshida, and Pachter (2006).

Note that for $k = 2$ the bound $2k = 4$ given in the second sentence of Theorem 2 is sharp for treelike dissimilarities, but that it can be improved to $2k - 1 = 3$ for equidistant dissimilarities (see e.g. Semple and Steel 2003, Theorem 7.2.5). Although this is not the case for $k \geq 3$, in Section 5 we shall prove that under certain circumstances it may still be possible to recover a tree from a $k$-dissimilarity $D$ on $X$ in case it is equidistant on every $(2k-1)$-element subset of $X$ (see Theorem 5).

We conclude the paper by considering 3-dissimilarities in more detail. It is well-known (see e.g. Gordon 1987, and Semple and Steel 2003) that treelike and equidistant 2-dissimilarities can be characterized in terms the 4-point and ultrametric condition, respectively (for more details see Section 6.). Thus, for $k \geq 3$, we can ask for similar “$m$-point” conditions that characterize treelike/equidistant $k$-dissimilarities. This question has been studied in Rubei (2011) for the case $k = 3$, where a recursive characterization is provided, and related problems are considered in Bocci and Cools.
Recognizing Treelike $k$-Dissimilarities (2009) in the context of tropical geometry. In addition, a necessary (but not sufficient) $(k + 2)$-point condition is given for the general case in Pachter and Speyer (2004, p. 618).

In the last section, we provide explicit 6-point characterizations for 3-dissimilarities that are treelike/equidistant, which can be regarded as generalizations of the 4-point/ultrametric conditions (Theorem 7). We conclude with a short discussion as to why finding similar conditions for $k \geq 4$ appears to be somewhat more challenging.

2. Preliminaries on Phylogenetic Trees

For the remainder of this paper, $X$ will always denote a non-empty, finite set. Also, for a $k$-dissimilarity $D : \binom{X}{k} \to \mathbb{R}$ and $\{x_1, x_2, \ldots, x_k\} \in \binom{X}{k}$, we will write $D(x_1, x_2, \ldots, x_k)$ instead of $D(\{x_1, x_2, \ldots, x_k\})$.

We now recall some further definitions concerning phylogenetic trees (for more details see Semple and Steel 2003). Let $T = (V, E)$ be a phylogenetic tree on $X$. A vertex $v \in V \setminus X$ is called an interior vertex of $T$. The set of leaves of $T$, that is, the set $X$, is also denoted by $L(T)$. Recall that it is assumed that all interior vertices have degree at least three. An edge $e \in E$ is called pendant if it is incident to a leaf of $T$. All other edges are called interior edges.

Now, two phylogenetic trees $T_1 = (V_1, E_1)$ and $T_2 = (V_2, E_2)$ on the same set $X$ are isomorphic if there exists a bijective map $\iota : V_1 \to V_2$ such that $\iota(x) = x$ holds for all $x \in X$ and $\{u, v\} \in E_1$ if and only if $\{\iota(u), \iota(v)\} \in E_2$ for any two distinct $u, v \in V_1$. In case we also have edge-weightings $\omega_i : E_i \to \mathbb{R}$, $i \in \{1, 2\}$, the weighted phylogenetic trees $(T_1, \omega_1)$ and $(T_2, \omega_2)$ are isomorphic if, in addition, $\omega_1(\{u, v\}) = \omega_2(\{\iota(u), \iota(v)\})$ holds for every edge $\{u, v\} \in E_1$. Note that interior edges with weight 0 can give rise to non-isomorphic weighted phylogenetic trees that induce the same $k$-dissimilarity. Therefore, in the following we will always implicitly assume that in any weighted phylogenetic tree interior edges are assigned positive weights. We call such edge-weightings interior-positive, for short.

We also apply the above terminology to (weighted) rooted phylogenetic trees with the following minor adaptations. For two rooted phylogenetic trees $T_1$ and $T_2$ with roots $\rho_1$ and $\rho_2$ to be isomorphic we require, in addition, that $\iota(\rho_1) = \rho_2$ holds. Note that in a weighted, rooted phylogenetic tree $(T, \rho, \omega)$ with $\omega$ equidistant, every interior edge has a non-negative weight while pendant edges might have negative weights (cf. Figure 1(c)). Again, to avoid non-isomorphic weighted, rooted phylogenetic trees giving rise to the same $k$-dissimilarity, we always assume that the edge-weightings are interior-positive. A rooted phylogenetic tree $T = (V, E)$ on $X$ with root
Figure 2. The weighted phylogenetic tree \((T(M), \omega(M))\) arising from the weighted, rooted phylogenetic tree in Figure 1(c) for \(M = 2\).

\(\rho\) is binary if every vertex in \(V \setminus (X \cup \{\rho\})\) has degree precisely three and \(\rho\) has degree two.

Next note that for every weighted, rooted phylogenetic tree \((T, \rho, \omega)\) with, not necessarily non-negative, equidistant edge-weighting \(\omega\) there exists a constant \(M \geq 0\) such that the edge-weighting \(\omega_M\), that assigns weight \(\omega(e)\) to every interior edge \(e\) and weight \(\omega(e) + M\) to every pendant edge \(e\), is also equidistant and non-negative. Thus, given a weighted, rooted phylogenetic tree \((T, \rho, \omega)\) on \(X\) with \(\omega\) equidistant, we can construct, for any sufficiently large constant \(M \geq 0\), a weighted phylogenetic tree \((T(M), \omega(M))\) on \(X\) with \(\omega(M)\) non-negative and interior-positive as follows: If \(\rho\) has degree at least three, then put \(T(M) = T\) and \(\omega(M) = \omega\). Otherwise, delete \(\rho\) and connect the two vertices \(u\) and \(v\) adjacent to \(\rho\) by a new edge with weight \(\omega_M(\{\rho, u\}) + \omega_M(\{\rho, v\})\) (cf. Figure 2). Note that if \(M\) is known, we can completely recover \((T, \rho, \omega)\) from \((T(M), \omega(M))\).

For every rooted phylogenetic tree \(T = (V, E)\) on \(X\) with root \(\rho\) there is a natural partial ordering \(\leq_T\) on \(V\) with unique minimal element \(\rho\) defined by \(v \leq_T w\) if and only if \(v\) is a vertex of the unique path from \(w\) to \(\rho\) in \(T\). The rooted subtree \(T_v\) of \(T\) induced by \(v \in V\) has vertex set \(\{u \in V : v \leq_T u\}\) and root \(v\). In addition, for any equidistant edge-weighting \(\omega\) of \(T\), we define the height \(h_{(T, \omega)}(v)\) of \(v\), also referred to as the height of \(T_v\), as the value \(d_{(T, \omega)}(v, x)\) for any leaf \(x\) of \(T_v\). Note that this height is well-defined in view of the fact that \(\omega\) is equidistant.

Finally, for rooted, as well as unrooted, phylogenetic trees \(T\) on \(X\) we denote, for any subset \(Y \subseteq X\), the smallest subtree of \(T\) containing the vertices in \(Y\) by \(T|_Y\) and refer to it as the restriction of \(T\) to \(Y\). (To formally view \(T|_Y\) as a phylogenetic tree on \(Y\), we suppress any vertices of degree 2.) In case \(T\) is rooted, we also consider \(T|_Y\) as a rooted phylogenetic tree where we distinguish the minimal element in the vertex set of \(T|_Y\) with respect to the partial order \(\leq_T\) as the root of the restriction. And, in case \(T\) has an edge-weighting \(\omega\), we consider \(T|_Y\) as a weighted tree with edge-weighting \(\omega|_Y\) obtained by restricting \(\omega\) to the edge set of \(T|_Y\).
3. Determining Trees

We begin this section by stating a uniqueness theorem that will be useful later:

**Theorem 3.** For every integer \( k \geq 2 \) and every set \( X \) with at least \( 2k - 1 \) elements we have:

(i) Two weighted phylogenetic trees \((T_1, \omega_1)\) and \((T_2, \omega_2)\) on \( X \) with \( \omega_i \) non-negative and interior-positive, \( i \in \{1, 2\} \), are isomorphic if and only if \( D^k_{(T_1, \omega_1)} = D^k_{(T_2, \omega_2)} \) holds.

(ii) Two weighted, rooted phylogenetic trees \((T_1, \rho_1, \omega_1)\) and \((T_2, \rho_2, \omega_2)\) on \( X \) with \( \omega_i \) equidistant and interior-positive, \( i \in \{1, 2\} \), are isomorphic if and only if \( D^k_{(T_1, \omega_1)} = D^k_{(T_2, \omega_2)} \) holds.

Note that for \( k = 2 \) parts (i) and (ii) of this theorem are well-known (see e.g. Semple and Steel 2003, Theorem 7.1.8). Moreover, part (i) is just a restatement of Theorem 1 above due to Pachter and Speyer, and part (ii) immediately follows from part (i) by considering the weighted phylogenetic trees \((T_1(M), \omega_1(M))\) and \((T_2(M), \omega_2(M))\) for some sufficiently large constant \( M \geq 0 \).

We now prove the first part of Theorem 2:

**Theorem 4.** Let \( k \geq 2 \) and \( D \) be a \( k \)-dissimilarity map on a set \( X \), \(|X| \geq 2k\).

(i) \( D \) is treelike if and only if the restriction of \( D \) to every \( 2k \)-element subset of \( X \) is treelike.

(ii) \( D \) is equidistant if and only if the restriction of \( D \) to every \( 2k \)-element subset of \( X \) is equidistant.

**Proof:** (i) For \( k = 2 \) this is well-known (see e.g. Semple and Steel 2003). So we shall assume in the following that \( k \geq 3 \) holds. Clearly, if \( D \) is treelike, then also the restriction to every \( 2k \)-element subset of \( X \) is treelike.

Conversely, assume that the restriction of \( D \) to every \( 2k \)-element subset of \( X \) is treelike. Note that this implies that the restriction of \( D \) to every \( i \)-element subset \( Y \) of \( X \), \( k \leq i \leq 2k \), is treelike, that is, there exists a weighted phylogenetic tree \( (T_Y, \omega_Y) \) on \( Y \) with \( \omega_Y \) non-negative and interior-positive such that \( D|_Y = D^k_{(T_Y, \omega_Y)} \) holds.

Now consider an arbitrary pair of elements \( \{a, b\} \in \binom{X}{2} \). We claim that in any weighted phylogenetic tree \((T_Z, \omega_Z)\), \( Z \in \binom{X}{2k-1} \), \( \{a, b\} \subseteq Z \), the induced distance \( d_{(T_Z, \omega_Z)}(a, b) \) is the same. To show this equivalence, it suffices to consider such sets \( Z, Z' \) with \( Z' = (Z \setminus \{x\}) \cup \{y\} \) for two distinct elements \( x, y \in X \setminus \{a, b\} \). We now consider the weighted phylogenetic tree.
\((T_Y, \omega_Y)\) for the \(2k\)-element set \(Y := Z \cup \{y\}\). Since \(|Z| = |Z'| = 2k - 1\), it follows by Theorem 3(i) that \((T_Z, \omega_Z)\) is isomorphic to \((T_Y|Z, \omega_Y|Z)\) and \((T_Z', \omega_Z')\) is isomorphic to \((T_Y|Z', \omega_Y|Z')\). This isomorphism implies that the induced distance between \(a\) and \(b\) is the same for \((T_Z, \omega_Z)\) and \((T_Z', \omega_Z')\), as claimed.

As a consequence, for every pair \((a, b) \in \binom{X}{2}\), the restriction of \(D\) to any \((2k-1)\)-element subset of \(X\) containing \(a\) and \(b\) yields the same distance between \(a\) and \(b\), which we denote by \(\delta(a, b)\). Note that the restriction of the so-defined 2-dissimilarity \(\delta\) on \(X\) to every 4-element subset of \(X\) is treelike: For any four distinct elements \(a, b, c, d \in X\) we can select an arbitrary \(Z \in \binom{X}{2}\) with \(\{a, b, c, d\} \subseteq Z\) and in the weighted phylogenetic tree \((T_Z, \omega_Z)\) the induced distances between \(a, b, c, d\) will equal the corresponding values of \(\delta\). Hence, (since the theorem holds for \(k = 2\)) there exists a unique weighted phylogenetic tree \((T, \omega)\) on \(X\) with \(\omega\) non-negative and interior-positive such that \(D^2_{(T, \omega)} = \delta\) holds. Moreover, the restriction of \((T, \omega)\) to any \(2k\)-element subset \(Z \subseteq X\) is isomorphic to \((T_Z, \omega_Z)\). Hence, the \(k\)-dissimilarity \(D^k_{(T, \omega)}\) must be \(D\).

(ii) Again, if a \(k\)-dissimilarity on \(X\) is equidistant, so is its restriction to each \(2k\)-subset of \(X\). So, let \(D\) be a \(k\)-dissimilarity on \(X\) such that its restriction to each \(Y \in \binom{X}{2k}\) is represented by a weighted, rooted phylogenetic tree \((T_Y, \rho_Y, \omega_Y)\) on \(Y\) with \(\omega_Y\) equidistant and interior-positive. Then, for some sufficiently large \(M \geq 0\), all the weighted phylogenetic trees \((T_Y(M), \omega_Y(M))\) are such that \(\omega_Y(M)\) is non-negative and interior-positive. Therefore, by the first part of the theorem, there exists a unique weighted phylogenetic tree \((T, \omega)\) on \(X\) with \(\omega\) non-negative and interior-positive such that \(D^k_{(T, \omega)}(A) = D(A) + kM\) holds for all \(A \in \binom{X}{k}\). Moreover, \((T, \omega)\) must be isomorphic to \((T'(M), \omega'(M))\) for some weighted, rooted phylogenetic tree \((T', \rho', \omega')\) on \(X\) with \(\omega'\) equidistant and interior-positive, since otherwise there would exist some \(Y \in \binom{X}{2k}\) such that \(\omega_Y\) is not equidistant in view of the fact that, for all \(Y \in \binom{X}{2k}\), \((T|_Y, \omega|_Y)\) is isomorphic to \((T_Y(M), \omega_Y(M))\). Hence \(D\) must equal \(D^k_{(T', \omega')}\), as required.

Using this theorem we now show that, for fixed \(k \geq 3\), it is possible to efficiently check when a \(k\)-dissimilarity is treelike/equidistant. Note that any algorithm to check whether a given \(k\)-dissimilarity \(D\) is treelike/equidistant needs to read \(D\) first. Assuming that \(D\) is given as the list of values it takes on for each \(k\)-element subset of \(X\), this yields a lower bound of \(|X|^k\) on the run-time of any such algorithm.

**Corollary 1.** For any fixed \(k \geq 3\) and any \(k\)-dissimilarity \(D\) on \(X\), there is an algorithm with run-time in \(O(f(k) \cdot |X|^{2k})\) to decide whether \(D\) is
Recognizing Treelike $k$-Dissimilarities

**Proof:** Given a $k$-dissimilarity $D$ on $X$, it suffices to check for every $Z \in \binom{X}{2k}$ whether $D|_Z$ is treelike/equidistant. To do this, one can enumerate all (isomorphism classes of) unweighted phylogenetic trees (rooted or unrooted) with $2k$ leaves labeled by the elements in $Z$. Note that the number of these trees depends on $k$ but not on $|X|$. For each of those trees $T$, it remains to check if there exists an edge-weighting $\omega$ with certain properties so that $D^k_{(T,\omega)} = D|_Z$ holds. The latter can be phrased as a test whether a system of linear equations and inequalities has a solution, a problem for which a polynomial time algorithm is known (see e.g. Schrijver 1986). Therefore, one can check in $O(f(k))$ time whether $D|_Z$ is treelike/equidistant where $f$ is a function that does not depend on $|X|$. Since the number of $2k$-element subsets of $X$ is in $O(|X|^{2k})$, this establishes the required run-time bound.

4. Sharpness of the Bounds

In this section, we shall prove the second part of Theorem 2, that is, we shall prove that the bounds presented in the theorem are indeed sharp. More specifically, for each $k \geq 3$, we will present an example of a $k$-dissimilarity $D$ whose restriction to every $(2k-1)$-element subset is treelike/equidistant while $D$ is not treelike/equidistant. These examples will be presented in Examples 1 and 2 below.

We begin by presenting a useful lemma. Assume we have $k \geq 3$, $|X| \geq k$ and that $(T = (V, E), \rho, \omega)$ is a weighted, rooted phylogenetic tree on $X$ with $\omega$ equidistant and interior-positive such that $\rho$ is adjacent with precisely two vertices $u$ and $v$ (cf. Figure 3(a)). Put $a = \omega(\{\rho, u\})$ and $b = \omega(\{\rho, v\})$. Now define, for each $\alpha \in \mathbb{R}$ with $\alpha < 2 \min\{a, b\}$, a new equidistant edge-weighting $\omega(k, \alpha)$ for $T$ (cf. Figure 3(b)) by putting, for all $e \in E$,

$$
\omega(k, \alpha)(e) = \begin{cases} 
\omega(e) + \alpha/k, & \text{if } e \text{ is incident to a leaf}, \\
\omega(e) - \alpha/2, & \text{if } e \text{ is incident to } \rho, \\
\omega(e), & \text{else}.
\end{cases}
$$

**Lemma 1.** Suppose $k \geq 3$, $|X| \geq k$ and that $(T = (V, E), \rho, \omega)$ is a weighted, rooted phylogenetic tree on $X$ with $\omega$ equidistant and interior-positive such that $\rho$ is adjacent with precisely two vertices $u$ and $v$ and put $a = \omega(\{\rho, u\})$ and $b = \omega(\{\rho, v\})$. Then for each $\alpha \in \mathbb{R}$ with $\alpha < 2 \min\{a, b\}$ and $A \in \binom{X}{k}$ we have
Figure 3. (a) A weighted, rooted phylogenetic tree \((T, \rho, \omega)\) with an equidistant and interior-positive edge-weighting \(\omega\). (b) The same rooted phylogenetic tree but with the equidistant edge-weighting \(\omega(k, \alpha)\) for \(k = 5\) and \(\alpha = 10\). Note that \(D^5_{(T, \omega)} = D^5_{(T, \omega(5, 10))}\).

\[
D^k_{(T, \omega(k, \alpha))}(A) = \begin{cases} 
D^k_{(T, \omega)}(A) + \alpha, & \text{if } A \subseteq L(T_u) \text{ or } A \subseteq L(T_v), \\
D^k_{(T, \omega)}(A), & \text{else}.
\end{cases}
\]

In particular, \((T, \rho, \omega)\) and \((T, \rho, \omega(k, \alpha))\) induce the same \(k\)-dissimilarity if and only if we have \(\alpha = 0\) or \(|L(T_u)| \leq k - 1\) and \(|L(T_v)| \leq k - 1\) hold.

**Proof:** Let \(A \subseteq \binom{X}{k}\). The restriction \(T|_A\) contains \(k\) pendant edges. Moreover, \(T|_A\) contains the two edges incident in \(T\) with \(\rho\) if and only if we have \(A \cap L(T_u) \neq \emptyset\) and \(A \cap L(T_v) \neq \emptyset\). Hence we have \(D^k_{(T, \omega(k, \alpha))}(A) = D^k_{(T, \omega)}(A) + k \frac{\alpha}{k}\) in case \(A \subseteq L(T_u)\) or \(A \subseteq L(T_v)\) holds and we have \(D^k_{(T, \omega(k, \alpha))}(A) = D^k_{(T, \omega)}(A) + k \frac{\alpha}{k} - 2 \frac{\alpha}{2} = D^k_{(T, \omega)}(A)\) otherwise, as claimed.

The second assertion trivially holds if \(\alpha = 0\). If \(\alpha \neq 0\), then it holds since all \(A \in \binom{X}{k}\) contain leaves from both \(L(T_u)\) and \(L(T_v)\) if and only if \(|L(T_u)| \leq k - 1\) and \(|L(T_v)| \leq k - 1\) hold.

**Example 1 (Equidistant).** Let \(k \geq 3\) and \((T, \rho, \omega)\) be a weighted, rooted phylogenetic tree on \(X\) with \(\omega\) equidistant and interior-positive such that the root \(\rho\) of \(T\) is adjacent with precisely two vertices \(u\) and \(v\). Put \(a = \omega(\{\rho, u\})\) and \(b = \omega(\{\rho, v\})\). Assume that \(|L(T_u)|, |L(T_v)| \geq k\). Now, choose some non-zero \(\alpha < 2 \min\{a, b\}\) and define a \(k\)-dissimilarity \(D\) on \(X\) via

\[
D(A) := \begin{cases} 
D^k_{(T, \omega(k, \alpha))}(A), & \text{if } A \subseteq L(T_u), \\
D^k_{(T, \omega)}(A), & \text{else},
\end{cases}
\]

for all \(A \subseteq \binom{X}{k}\). We first show that the restriction of \(D\) to any \((2k - 1)\)-element subset of \(X\) is equidistant: For any \(Y \in \binom{X}{2k - 1}\) we define a weighted, rooted phylogenetic tree \((T_Y, \rho_Y, \omega_Y)\) on \(Y\) with \(\omega_Y\) equidistant and interior-positive by setting
\[(T_Y, \rho_Y, \omega_Y) := \begin{cases} (T, \rho, \omega(a))|_Y, & \text{if } |Y \cap L(T_u)| \geq k, \\ (T, \rho, \omega)|_Y, & \text{else.} \end{cases}\]

By the definition of \(D\) it follows that \(D|_Y = D^k_{(T_Y, \omega_Y)}\) holds.

We now show that \(D\) is not equidistant. It suffices to show that there exists some subset \(Z\) of \(X\) such that \(D|_Z\) is not equidistant. So, let \(Z \subseteq X\) be such that \(|Z \cap L(T_u)| = |Z \cap L(T_v)| = k\) (such a subset exists since \(|L(T_u)|, |L(T_v)| \geq k\), and suppose there exists some weighted, rooted phylogenetic tree \((T_Z, \rho_Z, \omega_Z)\) on \(Z\) with \(\omega_Z\) equidistant and interior-positive such that \(D|_Z = D^k_{(T_Z, \omega_Z)}\) holds. For every \(a \in Z \cap L(T_u)\) we define the weighted, rooted phylogenetic tree \((T_a, \rho_a, \omega_a) := (T_Z, \rho_Z, \omega_Z)|_{Z \setminus \{a\}}\) on \(Z \setminus \{a\}\). Choose distinct \(x, y \in Z \cap L(T_u)\) which are not adjacent to a common vertex of degree 3 (this is possible since \(|Z \cap L(T_u)| \geq k > 2\). By Theorem 3(ii), \((T_x, \rho_x, \omega_x)\) is isomorphic to \((T_Z \setminus \{x\}, \rho_Z \setminus \{x\}, \omega_Z \setminus \{x\})\) and \((T, \rho, \omega)|_{Z \setminus \{x\}}\), and \((T_y, \rho_y, \omega_y)\) is isomorphic to \((T_Z \setminus \{y\}, \rho_Z \setminus \{y\}, \omega_Z \setminus \{y\})\) and \((T, \rho, \omega)|_{Z \setminus \{y\}}\).

By our choice of \(x\) and \(y\), up to isomorphism, there exists only one possible weighted, rooted phylogenetic tree on \(Z\) whose restriction to \(Z \setminus \{x\}\) and \(Z \setminus \{y\}\) is \(T_x\) and \(T_y\), respectively. Hence \((T_Z, \rho_Z, \omega_Z)\) is isomorphic to \((T, \rho, \omega)|_Z\). However, this contradicts the fact that \((T_Z, \rho_Z, \omega_Z)\) induces \(D|_Z\) since \(D(Z \cap L(T_u)) = D^k_{(T_z, \omega_Z)}(Z \cap L(T_u)) \neq D^k_{(T, \omega)}(Z \cap L(T_u)) = D^k_{(T_z, \omega)}(Z \cap L(T_u))\), where the first equality is by definition and the second follows from the fact that \((T_Z, \rho_Z, \omega_Z)\) and \((T, \rho, \omega)|_Z\) are isomorphic.

**Example 2 (Treelike).** An example for the case \(k = 3\) was given by Chepoi and Fichet (2007). Here we give an example for general \(k\). Based on the weighted, rooted phylogenetic tree \((T, \rho, \omega)\) in Example 1, consider \((T(M), \omega(M))\) for some sufficiently large constant \(M \geq 0\) and choose \(\alpha > 0\). Then, using the same arguments as in Example 1, it is straightforward to show that the \(k\)-dissimilarity \(D\) constructed in the same way as in Example 1 is not treelike while its restriction to every \((2k - 1)\)-element subset of \(X\) is treelike.

5. The Case \(2k - 1\) for Equidistant \(k\)-Dissimilarities

It is well-known that if the restriction of a 2-dissimilarity \(D\) on \(X\) to every subset of \(X\) of size 3 is equidistant, then \(D\) is equidistant (Semple and Steel 2003, Theorem 7.2.5). In contrast, in Example 1, we have seen that for \(k \geq 3\) a \(k\)-dissimilarity \(D\) is not necessarily equidistant if its restriction to every \((2k - 1)\)-element subset is equidistant. However, we shall now prove that we can still recover a tree if the restriction of such a \(D\) to all \((2k - 1)\)-element subsets \(Y \subseteq X\) is induced by a weighted, rooted phylogenetic tree.
(T, \rho, \omega) on Y with \omega equidistant and interior-positive such that (T, \rho, \omega) is generic, that is, T is binary and no two distinct interior vertices have the same height.

**Theorem 5.** Let k \geq 3 and D be a k-dissimilarity map on X such that, for all Y \in \binom{X}{2k-1}, there exists a generic, weighted, and rooted phylogenetic tree (T_Y, \rho_Y, \omega_Y) on Y with \omega_Y equidistant and interior-positive such that D|_Y = D_{(T_Y, \omega_Y)} holds. Then there exists a binary rooted phylogenetic tree T on X such that, for all Y \in \binom{X}{2k-1}, the unweighted, rooted phylogenetic trees T|_Y and T_Y are isomorphic.

To prove this theorem, we shall use a well-known result about collections of rooted phylogenetic trees each having three leaves, that will allow us to “merge” trees, which we now recall. A triplet on X is a pair (\{a,b\}, c) with a, b, c \in X distinct, which we denote also by ab|c. The set of all triplets on X is denoted by \mathcal{R}(X), and a subset \mathcal{R} of \mathcal{R}(X) is called a triplet system on X. Given a rooted phylogenetic tree T on X, the triplet system \mathcal{R}_T of T is the set of all triplets ab|c on X such that the path from a to b in T is vertex-disjoint from the path from c to the root \rho in T. It is easily seen that for a rooted phylogenetic tree T on X and a rooted phylogenetic tree T' on Y \subseteq X, we have \mathcal{R}_T' \subseteq \mathcal{R}_T if T' is isomorphic to T|_Y. We now state the aforementioned result:

**Theorem 6** (Theorem 9.2 (ii) in Dress, Huber, Koolen, Moulton, and Spillner 2011). A rooted phylogenetic tree T on X is, up to isomorphism, uniquely determined by the triplet system \mathcal{R}_T. Moreover, given a triplet system \mathcal{R} \subseteq \mathcal{R}(X) there exists a rooted phylogenetic tree on X with \mathcal{R}_T = \mathcal{R} if and only if \mathcal{R} satisfies the following two conditions:

(R1) For any three elements a, b, c \in X at most one of the triplets ab|c, bc|a and ca|b is contained in \mathcal{R}.

(R2) For any four elements a, b, c, d \in X, ab|c \in \mathcal{R} implies ad|c \in \mathcal{R} or ab|d \in \mathcal{R}.

To prove Theorem 5 we will also use the following rather technical result:

**Lemma 2.** Let k \geq 3 be an integer, X a set with |X| = 2k - 2 and (T_1, \rho_1, \omega_1) and (T_2, \rho_2, \omega_2) be two generic weighted, rooted phylogenetic trees on X with \omega_i equidistant and interior-positive, i \in \{1, 2\}. If D^k_{(T_1, \omega_1)} = D^k_{(T_2, \omega_2)} then T_1 and T_2 are isomorphic as unweighted, rooted phylogenetic trees.

**Proof:** We distinguish two cases. First consider the case that at least one of T_1 and T_2, say T_1, contains a vertex v_1 such that the set B of leaves of
the rooted subtree \((T_1)_{v_2}\) has cardinality \(k - 1\). Define \(A := X \setminus B\). We claim that \(T_2\) must contain a vertex \(v_2\) such that the set of leaves of \((T_2)_{v_2}\) is \(B\). To establish this, first note that \(D^k_{(T_1, \omega_1)}(A \cup \{b\}) = D^k_{(T_1, \omega_1)}(A \cup \{b'\})\) and, therefore, in view of \(D^k_{(T_1, \omega_1)} = D^k_{(T_2, \omega_2)}\), also \(D^k_{(T_2, \omega_2)}(A \cup \{b\}) = D^k_{(T_2, \omega_2)}(A \cup \{b'\})\) must hold for all \(b, b' \in B\). This implies that, for every \(b \in B\), the height of the vertex \(w_b\) where the path from \(b\) to the root of \(T_2\) first meets the subtree \(T_2|A\) must be the same. Hence, since \(T_2\) is generic, \(w_b = w_{b'}\) holds for all \(b, b' \in B\). But then the tree \(T_2|B\) must equal the tree \((T_2)_{v_2}\) for some vertex \(v_2\) of \(T_2\), as claimed.

Next, we claim that \(T_1|A\) and \(T_2|A\) as well as \(T_1|B\) and \(T_2|B\) are isomorphic as unweighted, rooted phylogenetic trees. By Theorem 6 it suffices to show that \(R(T_1|A) = R(T_2|A)\) and \(R(T_1|B) = R(T_2|B)\) holds. In the following we will focus on the set \(A\). A completely analogous argument yields \(R(T_1|B) = R(T_2|B)\). So, consider three arbitrary distinct elements \(a, b\) and \(c\) in \(A\) and an arbitrary \((k - 2)\)-element subset \(C\) of \(B\). Up to relabeling, Figure 4 depicts the possible cases for the structure of the tree \(T_1|C \cup \{a, b, c\}\). Note that, by the assumption that \(T_1\) is generic, cases (b), (d), (g), (i), (j) and (k) are ruled out. In the remaining cases we have:

(a) \(D^k_{(T_1, \omega_1)}(C \cup \{a, b\}) = D^k_{(T_1, \omega_1)}(C \cup \{a, c\}) < D^k_{(T_1, \omega_1)}(C \cup \{b, c\})\)

(c) (and, similarly, (e) and (f)) \(D^k_{(T_1, \omega_1)}(C \cup \{b, c\}) < D^k_{(T_1, \omega_1)}(C \cup \{a, b\}) = D^k_{(T_1, \omega_1)}(C \cup \{a, c\})\)

(h) \(D^k_{(T_1, \omega_1)}(C \cup \{b, c\}) < D^k_{(T_1, \omega_1)}(C \cup \{a, c\}) < D^k_{(T_1, \omega_1)}(C \cup \{a, b\})\)

So, we have \(ab|c \in R(T_1|A)\) if and only if either

1. \(D^k_{(T_1, \omega_1)}(C \cup \{a, c\}) = D^k_{(T_1, \omega_1)}(C \cup \{b, c\}) \neq D^k_{(T_1, \omega_1)}(C \cup \{a, b\})\) holds, or

2. \(D^k_{(T_1, \omega_1)}(C \cup \{a, b\}), D^k_{(T_1, \omega_1)}(C \cup \{a, c\})\) and \(D^k_{(T_1, \omega_1)}(C \cup \{b, c\})\) are pairwise distinct and \(D^k_{(T_1, \omega_1)}(C \cup \{a, b\})\) is the smallest value among them.

It follows that, in view of the fact that \(D^k_{(T_1, \omega_1)} = D^k_{(T_2, \omega_2)}\) holds, we have \(R(T_1|A) = R(T_2|A)\), as required.

Now, to show that the trees \(T_1\) and \(T_2\) are isomorphic as unweighted, rooted phylogenetic trees, it remains to show that \(T_1|B\) and \(T_2|B\) are attached to \(T_1|A\) and \(T_2|A\), respectively, at the same position. To establish this result, consider the set \(A^*\) containing all \(a \in A\) such that \(D^k_{(T_1, \omega_1)}(B \cup \{a\})\) is minimal. Note that there must exist vertices \(w_1\) in \(T_1\) and \(w_2\) in \(T_2\) such that \(T_1|A^* = (T_1)_{w_1}\) and \(T_2|A^* = (T_2)_{w_2}\) hold. Moreover, note that \(v_1\) and \(w_1\) as
Figure 4. Schematic representations of the cases considered in the proof of Lemma 2 in the context of establishing that $R(T_1|_A) = R(T_2|_A)$ holds.

well as $v_2$ and $w_2$ must be adjacent to a common vertex, namely the vertex where $T_1|_B$ and $T_2|_B$ are attached to $T_1|_A$ and $T_2|_A$, respectively. It follows that $T_1$ and $T_2$ are isomorphic as unweighted, rooted phylogenetic trees.

Next consider the case that neither $T_1$ nor $T_2$ contains a vertex such that the subtree induced by that vertex has precisely $k - 1$ leaves. Choose some $M \in \mathbb{R}$ large enough and consider the weighted phylogenetic trees $(T_1(M), \omega_1(M))$ and $(T_2(M), \omega_2(M))$. Note that, for every edge $e$ of $T_j(M)$, $j \in \{1, 2\}$, removing $e$ from $T_j(M)$ yields two subtrees, one of which has at least $k$ leaves. That is the crucial property used in the proof of Theorem 1 presented in Pachter and Speyer (2004), and the lower bound $|X| \geq 2k - 1$ stated in this theorem is only needed to ensure that this property holds. Hence, even in case $|X| = 2k - 2$ the proof can be applied as long as we have this property. Therefore $T_1(M)$ and $T_2(M)$ are isomorphic (even as weighted phylogenetic trees!) and, hence, also $T_1$ and $T_2$, as required.

Proof of Theorem 5. Define $\mathcal{R} := \bigcup_{Y \in \binom{X}{2k-1}} \mathcal{R}_{T_Y}$. We have to show that Conditions (R1) and (R2) hold for $\mathcal{R}$.

To show that (R1) holds, it suffices to show that for all $Z \in \binom{X}{2k-2}$, and distinct $x, y \in X \setminus Z$, $a, b, c \in Z$ at most one of the triplets $ab|c$, $bc|a$ and $ca|b$ is contained in $\mathcal{R}_{T_{Z \cup \{x\}}} \cup \mathcal{R}_{T_{Z \cup \{y\}}}$. Lemma 2 implies that the phylogenetic trees $T^* := T_{\{Z \cup \{x\}\}}|_Z$ and $T_{Z \cup \{y\}}|_Z$ are isomorphic. By Theorem 6, Condition (R1) holds for $\mathcal{R}_{T^*}$, as required.
We now show that \((\mathcal{R}2)\) holds. Let \(a, b, c, d\) be distinct elements of \(X\) and suppose \(ab|c \in \mathcal{R}\). Then there exists some \(Y \in \binom{X}{2k-1}\) such that \(ab|c \in \mathcal{R}_{T_Y}\). If \(d \in Y\), then we have \(ad|c \in \mathcal{R}\) or \(ab|d \in \mathcal{R}\) since \((\mathcal{R}2)\) holds for \(\mathcal{R}_{T_Y}\). If \(d \notin Y\), take some \(x \in Y \setminus \{a, b, c\}\) and define \(Y' := (Y \setminus \{x\}) \cup \{d\}\). Again, by Lemma 2, the phylogenetic trees \(T_Y|_{Y\setminus\{x\}}\) and \(T_{Y'}|_{Y\setminus\{x\}}\) are isomorphic, hence \(ab|c\) is also an element of \(\mathcal{R}_{T_{Y'}}\), and hence \(ad|c \in \mathcal{R}\) or \(ab|d \in \mathcal{R}\) since \((\mathcal{R}2)\) holds for \(\mathcal{R}_{T_{Y'}}\).

\[\blacksquare\]

**Remark 1.** We suspect, but have not been able to prove, that, for \(k \geq 3\), if a \(k\)-dissimilarity on \(X\) is induced by an arbitrary weighted, rooted phylogenetic tree \((T_Y, \rho_Y, \omega_Y)\) with \(\omega_Y\) equidistant and interior-positive for all \(Y \in \binom{X}{2k-1}\), then it determines a rooted phylogenetic tree \(T\) on \(X\) such that, for all \(Y \in \binom{X}{2k-1}\), the tree \(T_Y\) is isomorphic to \(T|_Y\). Furthermore, depending on the topology of the unweighted phylogenetic tree \(T\) arising in this way, it might even still be possible to assign weights to the edges of \(T\) so that the edge-weighting is equidistant and the induced \(k\)-dissimilarity is \(D\). For example, if the number of leaves in one of the subtrees induced by the vertices adjacent to the root of \(T\) is smaller than \(k\), then one can extend the arguments in the proof of Theorem 4 to show that one can indeed construct a suitable edge-weighting for \(T\), and hence \(D\) is equidistant in this case.

### 6. 3-Dissimilarities

In this section, we prove that treelike and equidistant 3-dissimilarities can be characterized by certain 6-point conditions. We begin by recalling some conditions for characterizing treelike and equidistant 2-dissimilarities (see e.g. Smolenskii 1962; Zaretsky 1965; Buneman 1971; Gordon 1987; and Semple and Steel 2003).

It is well-known that a 2-dissimilarity \(D\) on \(X\) is treelike if and only if \(D\) is non-negative, it satisfies the *triangle inequality* (i.e., \(D(x_1, x_3) \leq D(x_1, x_2) + D(x_2, x_3)\)) holds for any three distinct elements \(x_1, x_2, x_3 \in X\) and

\[
D(x, x') + D(y, y') \leq \max\{D(x, y) + D(x', y'), D(x, y') + D(x', y)\}
\]

(1)

holds for any four distinct \(x, x', y, y' \in X\). Similarly, it is known that \(D\) is equidistant if and only if

\[
D(x, y) \leq \max\{D(x, z), D(z, y)\}
\]

(2)

holds for any three distinct \(x, y, z \in X\).

Inequalities (1) and (2) are commonly called the *4-point* and *ultrametric* conditions, respectively. Note that non-negativity of \(D\) and the triangle
inequality follow from the 4-point condition if one defines \( D(a, a) = 0 \) for all \( a \in X \) and then drops the requirement that the elements are pairwise distinct. However, as we view a \( k \)-dissimilarity as being a map from \( \binom{X}{k} \) into \( \mathbb{R} \), we need to explicitly require these additional properties.

We now present similar conditions that characterize treelike/equidistant 3-dissimilarities that are obtained by associating with every 3-dissimilarity a suitable 2-dissimilarity. The construction of this 2-dissimilarity is similar to the approach followed in the context of so-called perimeter models considered, for example, in Heiser and Bennani (1997) and Chepoi and Fichet (2007).

**Theorem 7.** Let \( D \) be a 3-dissimilarity on a set \( X \) with \( |X| \geq 5 \).

(i) \( D \) is treelike if and only if for all \( \{a, b, c, d, e\} \in \binom{X}{5} \)

\[
D(a, c, d) + D(a, c, e) + D(a, d, e) + D(b, c, d) + D(b, c, e) + D(b, d, e) \\
\leq 2 \left( D(a, b, c) + D(a, b, d) + D(a, b, e) + D(c, d, e) \right),
\]

\[
2 \left( D(a, c, d) + D(a, c, e) + D(b, d, e) \right) \\
\leq D(a, b, c) + D(a, b, d) + D(a, b, e) + D(a, d, e) + D(b, c, d) + D(b, c, e) + D(c, d, e),
\]

\[
D(a, c, d) + D(a, c, e) + D(b, d, e) \\
\leq \max \left\{ \frac{D(a, b, d) + D(a, b, e) + D(c, d, e)}{D(a, d, e) + D(b, c, d) + D(b, c, e)} \right\},
\]

and for all \( \{a, b, c, d, e, e'\} \in \binom{X}{6} \)

\[
2D(a, b, e) - D(a, c, e) - D(a, d, e) - D(b, c, e) - D(b, d, e) + 2D(c, d, e) \\
= 2D(a, b, e') - D(a, c, e') - D(a, d, e') - D(b, c, e') - D(b, d, e') + 2D(c, d, e').
\]

(ii) \( D \) is equidistant if and only if for all \( \{a, b, c, d, e\} \in \binom{X}{5} \)

\[
D(a, b, e) + D(c, d, e) \leq \max \left\{ \frac{D(a, c, e) + D(b, d, e)}{D(a, d, e) + D(b, c, e)} \right\}
\]

and for all \( \{a, b, c, d, e, e'\} \in \binom{X}{6} \) Equation (6) holds.
**Proof.** For any $Y = \{a, b, c, d, e\} \in \binom{X}{5}$ we define a map $\delta_Y : \binom{Y}{2} \to \mathbb{R}$ as follows. First define the vector

$$v_Y := (D(a, b, c), D(a, b, d), D(a, b, e), D(a, c, d), \ldots, D(c, d, e))^T$$

as well as the following matrix and its inverse (note that $A$ has full rank):

$$A = \begin{bmatrix}
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{bmatrix}$$

$$A^{-1} = \frac{1}{6} \cdot \begin{bmatrix}
2 & 2 & 2 & -1 & -1 & -1 & -1 & -1 & 2 & 2 \\
2 & -1 & -1 & 2 & 2 & -1 & -1 & -1 & 2 & -1 \\
-1 & 2 & -1 & 2 & -1 & 2 & -1 & 2 & -1 & -1 \\
-1 & -1 & 2 & -1 & 2 & 2 & -1 & -1 & 2 & -1 \\
-1 & -1 & 2 & 2 & -1 & -1 & 2 & -1 & 2 & -1 \\
-1 & -1 & 2 & 2 & -1 & -1 & 2 & -1 & -1 & 2 \\
-1 & 2 & -1 & -1 & 2 & -1 & -1 & 2 & -1 & 2 \\
2 & -1 & -1 & -1 & 2 & -1 & -1 & 2 & -1 & 2 \\
2 & -1 & -1 & -1 & 2 & -1 & -1 & 2 & -1 & 2
\end{bmatrix}$$

Then, using the notation

$$u_Y = (\delta_Y(a, b), \delta_Y(a, c), \delta_Y(a, d), \delta_Y(a, e), \delta_Y(b, c), \ldots, \delta_Y(d, e))^T,$$

the map $\delta_Y$ is defined by the unique solution of the system of linear equations

$$2v_Y = A \cdot u_Y. \quad (8)$$

In particular we have:

$$3\delta_Y(a, b) = 2D(a, b, c) + 2D(a, b, d) + 2D(a, b, e) - D(a, c, d)$$
$$- D(a, c, e) - D(a, d, e) - D(b, c, d) - D(b, c, e)$$
$$- D(b, d, e) + 2D(c, d, e),$$

$$3\delta_Y(a, c) = 2D(a, b, c) - D(a, b, d) - D(a, b, e) + 2D(a, c, d)$$
$$+ 2D(a, c, e) - D(a, d, e) - D(b, c, d) - D(b, c, e)$$
$$+ 2D(b, d, e) - D(c, d, e),$$
\[
3\delta_Y(a, d) = -D(a, b, c) + 2D(a, b, d) - D(a, b, e) + 2D(a, c, d) \\
- D(a, c, e) + 2D(a, d, e) - D(b, c, d) + 2D(b, c, e)
\]

(9)
\[
- D(b, d, e) - D(c, d, e),
\]

\[
3\delta_Y(b, c) = 2D(a, b, c) - D(a, b, d) - D(a, b, e) - D(a, c, d) \\
- D(a, c, e) + 2D(a, d, e) + 2D(b, c, d) + 2D(b, c, e) \\
- D(b, d, e) - D(c, d, e),
\]

\[
3\delta_Y(b, d) = -D(a, b, c) + 2D(a, b, d) - D(a, b, e) - D(a, c, d) \\
+ 2D(a, c, e) - D(a, d, e) + 2D(b, c, d) - D(b, c, e) \\
+ 2D(b, d, e) - D(c, d, e),
\]

\[
3\delta_Y(c, d) = -D(a, b, c) - D(a, b, d) + 2D(a, b, e) + 2D(a, c, d) \\
- D(a, c, e) - D(a, d, e) + 2D(b, c, d) - D(b, c, e) \\
- D(b, d, e) + 2D(c, d, e).
\]

It is not hard to see that \(D|_Y\) is treelike/equidistant if and only if \(\delta_Y\) is treelike/equidistant. This key observation will allow us to translate the 4-point/ultrametric condition characterizing when a 2-dissimilarity is treelike/equidistant into conditions for when a 3-dissimilarity is treelike/equidistant.

But first, we need some condition that ensures that, for any two distinct \(a, b \in X\) and any two distinct \(Z, Z' \in (\frac{X}{5})\) with \(\{a, b\} \subseteq Z \cap Z'\), we have \(\delta_Z(a, b) = \delta_{Z'}(a, b)\). Clearly, it suffices to consider such sets \(Z, Z' \in (\frac{X}{5})\) with \(|Z \cap Z'| = 4\). In particular, for \(Z = \{a, b, c, d, e\}\) and \(Z' = (Z \setminus \{e\}) \cup \{e'\}\) we obtain Equation (6) which ensures that the map \(\delta : (\frac{X}{2}) \to \mathbb{R}\) defined by putting \(\delta(a, b) := \delta_Z(a, b)\) for an arbitrary \(Z \in (\frac{X}{5})\) with \(\{a, b\} \subseteq Z\) is well-defined. In this case \(\delta_Z\) is treelike/equidistant for all \(Z \in (\frac{X}{5})\) if and only if \(\delta\) is treelike/equidistant if and only if \(D\) is treelike/equidistant.

We now prove the two assertions of the theorem: (i) Using the equations in (9), it is not hard to check that the conditions for \(\delta\) (to be non-negative, to satisfy the triangle inequality and the 4-point condition) translate into Inequalities (3), (4) and (5) respectively. (ii) Again it is not hard to check that, using the equations in (9), the ultrametric condition on \(\delta\) translates into Inequality (7).

Note that Bocci and Cools (2009, Theorem 3.2) showed that there exists a family of maps \(\phi_k\) from the set of all treelike 2-dissimilarities to the set of all treelike \(k\)-dissimilarities that maps a 2-dissimilarity \(D^2_{(T, \omega)}\) induced by a weighted phylogenetic tree \((T, \omega)\) on \(X\) with \(\omega\) non-negative to the treelike \(k\)-dissimilarity \(\phi_k(D^2_{(T, \omega)}) = D^k_{(T, \omega)}\). For \(k = 3\) this map can be thought of as multiplication with the matrix \(A\) as considered in the proof.
of Theorem 7, but for $k \geq 4$ it appears that no such simple representation is possible.

Indeed, the key observation used in proving the above result is that the restriction of any treelike/equidistant $3$-dissimilarity $D$ to every $5$-element subset $Y \subseteq X$ can be related to a unique treelike/equidistant $2$-dissimilarity on $Y$ by a system of linear equations (as this allowed the straight-forward translation of the $4$-point/ultrametric condition into a $5$-point condition). Unfortunately, it seems that there are problems when we try to apply this idea in case $k \geq 4$, even for $k = 4$.

More specifically, first note that, although the restriction of any treelike $4$-dissimilarity $D$ to every $6$-element subset $Y \subseteq X$ can be related to a treelike $2$-dissimilarity on $Y$ by a system of linear equations, to do this one has to select a suitable ordering of the elements in $Y$ (see also Bocci and Cools 2009, Theorem 2.2). In contrast, in the case $k = 3$ any ordering works. Moreover, the system of linear equations does not need to have a unique solution, that is, even after fixing a suitable ordering, there can be more than one $2$-dissimilarity on $Y$ associated to the $4$-dissimilarity $D$. This further complicates the translation of the $4$-point condition into some form of $8$-point condition for when $D$ is treelike.

References

BANDELT, H.-J. (1990), “Recognition of Tree Metrics,” SIAM Journal on Discrete Mathematics, 3, 1–6.

BANDELT, H.-J., and DRESS, A.W.M. (1994), “An Order-Theoretic Framework for Overlapping Clustering,” Discrete Mathematics, 136, 21–37.

BOCCI, C., and COOLS, F. (2009), “A Tropical Interpretation of $m$-Dissimilarity Maps,” Applied Mathematics and Computation, 212, 349–356.

BUNEMAN, P. (1971), “The Recovery of Trees from Measures of Dissimilarity,” in Mathematics in the Archaeological and Historical Sciences, eds. D.G. Kendall and P. Tautu, Edinburgh: Edinburgh University Press, pp. 387–395.

CHEPOI, V., and FICHET, B. (2007), “A Note on Three-Way Dissimilarities and Their Relationship with Two-Way Dissimilarities,” in Selected Contributions in Data Analysis and Classification, ed. P. Brito et al., Berlin: Springer, pp. 465–475.

CULBERSON, J., and RUDNICKI, P. (1989), “A Fast Algorithm for Constructing Trees from Distance Matrices,” Information Processing Letters, 30, 215–220.

DE SOETE, G. (1983), “A Least Squares Algorithm for Fitting Additive Trees to Proximity Data,” Psychometrika, 48, 621–626.

DEZA, M.-M., and ROSENBERG, I.G. (2000), “$n$-Semimetrics,” European Journal of Combinatorics, 21, 797–806.

DRESS, A.W.M., HUBER, K.T., KOOLEN, J., MOULTON, V., and SPILLNER, A. (2011), Basic Phylogenetic Combinatorics, Cambridge: Cambridge University Press.

DRESS, A.W.M., and STEEL, M. (2007), “Phylogenetic Diversity over an Abelian Group,” Annals of Combinatorics, 11, 143–160.

FAITH, D.P. (1992), “Conservation Evaluation and Phylogenetic Diversity,” Biological Conservation, 61, 1–10.
FELSENSTEIN, J. (2003), *Inferring Phylogenies*, Sunderland, Massachusetts: Sinauer Associates.

GORDON, A.D. (1987), “A Review of Hierarchical Classification,” *Journal of the Royal Statistical Society. Series A. General*, 150, 119–137.

GRISHIN, N. (1999), “A Novel Approach to Phylogeny Reconstruction from Protein Sequences,” *Journal of Molecular Evolution*, 48, 264–273.

HAYASHI, C. (1972), “Two Dimensional Quantification Based on the Measure of Dissimilarity Among Three Elements,” *Annals of the Institute of Statistical Mathematics*, 24, 251–257.

HEISER, W.J., and BENNANI, M. (1997), “Triadic Distance Models: Axiomatization and Least Squares Representation,” *Journal of Mathematical Psychology*, 41, 189–206.

JOLY, S., and LE CALVÉ, G. (1995), “Three-Way Distances,” *Journal of Classification*, 12, 191–205.

LEVY, D., YOSHIDA, R., and PACHTER, L. (2006), “Beyond Pairwise Distances: Neighbor-Joining with Phylogenetic Diversity Estimates,” *Molecular Biology and Evolution*, 23, 491–498.

PACHTER, L., and SPEYER, D. (2004), “Reconstructing Trees from Subtree Weights,” *Applied Mathematics Letters*, 17, 615–621.

RUBEI, E. (2011), “Sets of Double and Triple Weights of Trees,” *Annals of Combinatorics*, 15, 723–734.

SCHRIJVER, A. (1986), *Theory of Linear and Integer Programming*, Wiley-Interscience Series in Discrete Mathematics, Chichester: John Wiley & Sons Ltd.

SEMBLE, C., and STEEL, M. (2003), *Phylogenetics* (Vol. 24), *Oxford Lecture Series in Mathematics and Its Applications*, Oxford: Oxford University Press.

SMOLENSKII, Y.A. (1962), “A Method for the Linear Recording of Graphs,” *U.S.S.R. Computational Mathematics and Mathematical Physics*, 2, 396–397.

STEEL, M. (2005), “Phylogenetic Diversity and the Greedy Algorithm,” *Systematic Biology*, 54, 527–529.

WARRENS, M.J. (2010), “n-Way Metrics,” *Journal of Classification*, 27, 173–190.

ZARETSKY, K. (1965), “Reconstruction of a Tree from the Distances Between Its Pendant Vertices,” *Uspekhi Matematicheskikh Nauk (Russian Mathematical Surveys)*, 20, 90–92.