



UNIVERSITÀ
DEGLI STUDI
FIRENZE

FLORE

Repository istituzionale dell'Università degli Studi di Firenze

Families of multiweights and pseudostars

Questa è la Versione finale referata (Post print/Accepted manuscript) della seguente pubblicazione:

Original Citation:

Families of multiweights and pseudostars / Baldisserri, Agnese; Rubei, Elena. - In: ADVANCES IN APPLIED MATHEMATICS. - ISSN 0196-8858. - STAMPA. - 77(2016), pp. 86-100. [10.1016/j.aam.2016.03.001]

Availability:

This version is available at: 2158/1029150 since: 2021-03-23T19:06:05Z

Published version:

DOI: 10.1016/j.aam.2016.03.001

Terms of use:

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (<https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf>)

Publisher copyright claim:

(Article begins on next page)

Families of multiweights and pseudostars

Agnese Baldisserri Elena Rubei

Address of both authors: Dipartimento di Matematica e Informatica “U. Dini”,
viale Morgagni 67/A, 50134 Firenze, Italia

E-mail addresses: baldisser@math.unifi.it, rubei@math.unifi.it

Corresponding author: Elena Rubei

Abstract

Let $\mathcal{T} = (T, w)$ be a weighted finite tree with leaves $1, \dots, n$. For any $I := \{i_1, \dots, i_k\} \subset \{1, \dots, n\}$, let $D_I(\mathcal{T})$ be the weight of the minimal subtree of T connecting i_1, \dots, i_k ; the $D_I(\mathcal{T})$ are called k -weights of \mathcal{T} . Given a family of real numbers parametrized by the k -subsets of $\{1, \dots, n\}$, $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{k}}$, we say that a weighted tree $\mathcal{T} = (T, w)$ with leaves $1, \dots, n$ realizes the family if $D_I(\mathcal{T}) = D_I$ for any I .

In [14] Pachter and Speyer proved that, if $3 \leq k \leq (n+1)/2$ and $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{k}}$ is a family of positive real numbers, then there exists at most one positive-weighted essential tree \mathcal{T} with leaves $1, \dots, n$ that realizes the family (where “essential” means that there are no vertices of degree 2). We say that a tree P is a pseudostar of kind (n, k) if the cardinality of the leaf set is n and any edge of P divides the leaf set into two sets such that at least one of them has cardinality $\geq k$. Here we show that, if $3 \leq k \leq n-1$ and $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{k}}$ is a family of real numbers realized by some weighted tree, then there is exactly one weighted essential pseudostar $\mathcal{P} = (P, w)$ of kind (n, k) with leaves $1, \dots, n$ and without internal edges of weight 0, that realizes the family; moreover we describe how any other weighted tree realizing the family can be obtained from \mathcal{P} . Finally we examine the range of the total weight of the weighted trees realizing a fixed family.

1 Introduction

For any graph G , let $E(G)$, $V(G)$ and $L(G)$ be respectively the set of the edges, the set of the vertices and the set of the leaves of G . A **weighted graph** $\mathcal{G} = (G, w)$ is a graph G endowed with a function $w : E(G) \rightarrow \mathbb{R}$. For any edge e , the real number $w(e)$ is called the weight of the edge. If all the weights are nonnegative (respectively positive), we say that the graph is **nonnegative-weighted** (respectively **positive-weighted**); if the weights of the internal edges are nonzero, we say that the graph is **internal-nonzero-weighted** and, if all the weights are nonnegative and the ones of the

2010 Mathematical Subject Classification: 05C05, 05C12, 05C22

Key words: weighted trees, dissimilarity families

internal edges are positive, we say that the graph is **internal-positive-weighted**. For any finite subgraph G' of G , we define $w(G')$ to be the sum of the weights of the edges of G' . In this paper we will deal only with weighted finite trees.

Definition 1. Let $\mathcal{T} = (T, w)$ be a weighted tree. For any distinct $i_1, \dots, i_k \in V(T)$, we define $D_{\{i_1, \dots, i_k\}}(\mathcal{T})$ to be the weight of the minimal subtree containing i_1, \dots, i_k . We call this subtree “the subtree realizing $D_{\{i_1, \dots, i_k\}}(\mathcal{T})$ ”. More simply, we denote $D_{\{i_1, \dots, i_k\}}(\mathcal{T})$ by $D_{i_1, \dots, i_k}(\mathcal{T})$ for any order of i_1, \dots, i_k . We call the $D_{i_1, \dots, i_k}(\mathcal{T})$ the **k -weights** of \mathcal{T} and we call a k -weight of \mathcal{T} for some k a **multiweight** of \mathcal{T} .

If S is a subset of $V(T)$, the k -weights give a vector in $\mathbb{R}^{\binom{S}{k}}$. This vector is called **k -dissimilarity vector** of (\mathcal{T}, S) . Equivalently, we can speak of the **family of the k -weights** of (\mathcal{T}, S) .

If S is a finite set, $k \in \mathbb{N}$ and $k < \#S$, we say that a family of real numbers $\{D_I\}_{I \in \binom{S}{k}}$ is **treelike** (respectively p-treelike, nn-treelike, inz-treelike, ip-treelike) if there exist a weighted (respectively positive-weighted, nonnegative-weighted, internal-nonzero-weighted, internal-positive-weighted) tree $\mathcal{T} = (T, w)$ and a subset S of the set of its vertices such that $D_I(\mathcal{T}) = D_I$ for any k -subset I of S . If in addition $S \subset L(T)$, we say that the family is **l-treelike** (respectively p-l-treelike, nn-l-treelike, inz-l-treelike, ip-l-treelike).

Graphs and in particular weighted graphs may have applications in several disciplines, such as biology, psychology, archeology, engineering. Phylogenetic trees are positive-weighted trees whose vertices represent species and the weight of an edge is given by how much the DNA sequences of the species represented by the vertices of the edge differ. There is a wide literature concerning graphlike dissimilarity families and treelike dissimilarity families, in particular concerning methods to reconstruct positive-weighted trees from their dissimilarity families; these methods are used by biologists to reconstruct phylogenetic trees (see for example [13], [19] and [7], [17] for overviews); also archeologists represent evolutions of manuscripts by positive-weighted trees. See the introduction in [6] for some applications to psychology. Also the case of general weighted graphs can be interesting. For example, consider a web of pipes such that, when a particle or a material goes from a vertex of an edge to the other vertex, it gets or loses some quantity of a particular substance; call this quantity weight of the edge, so that the web becomes a weighted graph \mathcal{G} ; making a material run from a vertex i of the graph to another j , we get a value for $D_{i,j}(\mathcal{G})$; analogously, the numbers $D_{i_1, \dots, i_k}(\mathcal{G})$ can represent how much a material, by going from the leaf i_s to the leaves $i_1, \dots, i_s, \dots, i_k$, gets or loses of a certain substance. So weighted graphs can represent hydraulic webs or webs in the human body; moreover they can represent also railway webs where for some lines the difference between the earnings and the cost of the line is positive and for the other lines is zero or negative. If we know “how much we get or lose” by going from a leaf of a weighted tree to other leaves (the values D_{i_1, \dots, i_k}), we can try to reconstruct the weighted tree. It can be interesting, given a family of real numbers, $\{D_{i_1, \dots, i_k}\}_{i_1, \dots, i_k}$, to wonder if there exists a weighted tree with it as family of k -weights, if it is unique and, if it is not unique, which is the tree realizing the given family and with maximum/minimum total weight (if there exists); for instance, for “railway trees” where for some lines the difference between the earnings and the cost of the line is positive and for the other lines is zero or negative, we can be interested in searching the trees with maximum total weight. Moreover, in research fields where only positive-weighted trees are studied, if from experimental data we get a

family of k -weights, for some k , which is treelike but not positive-treelike, this can suggest that there is a systematic underestimate of the weights.

One of the first results on weighted trees is a criterion for a metric on a finite set to be nn - l -treelike, see [5], [18], [20]:

Theorem 2. *Let $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ be a set of positive real numbers satisfying the triangle inequalities. It is p -treelike (or nn - l -treelike) if and only if, for all distinct $a, b, c, d \in \{1, \dots, n\}$, the maximum of*

$$\{D_{a,b} + D_{c,d}, D_{a,c} + D_{b,d}, D_{a,d} + D_{b,c}\}$$

is attained at least twice.

As we have already said, also the study of general weighted trees can be interesting and, obviously, it cannot be deduced from the study of positive-weighted trees; in [3], Bandelt and Steel proved a result, analogous to Theorem 2, for general weighted trees:

Theorem 3. (Bandelt-Steel) *For any set of real numbers $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$, there exists a weighted tree \mathcal{T} with leaves $1, \dots, n$ such that $D_I(\mathcal{T}) = D_I$ for any $I \in \binom{\{1, \dots, n\}}{2}$ if and only if, for any $a, b, c, d \in \{1, \dots, n\}$, at least two among $D_{a,b} + D_{c,d}$, $D_{a,c} + D_{b,d}$, $D_{a,d} + D_{b,c}$ are equal.*

In [8] some results on graphs with minimal total weight among the ones realizing a given metric space were established.

For higher k the literature is more recent, see [1], [4], [9], [10], [11], [12], [14], [15], [16]. Three of the most important results for higher k are the following:

Theorem 4. (Herrmann, Huber, Moulton, Spillner, [9]). *If $n \geq 2k$, a family of positive real numbers $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{k}}$ is ip - l -treelike if and only if the restriction to every $2k$ -subset of $\{1, \dots, n\}$ is ip - l -treelike.*

Theorem 5. (Levy-Yoshida-Pachter, [11]) *Let $\mathcal{T} = (T, w)$ be a positive-weighted tree with $L(T) = \{1, \dots, n\}$. For any $i, j \in \{1, \dots, n\}$, define*

$$S(i, j) = \sum_{Y \in \binom{\{1, \dots, n\} - \{i, j\}}{k-2}} D_{i,j,Y}(\mathcal{T}).$$

Then there exists a positive-weighted tree $\mathcal{T}' = (T', w')$ such that $D_{i,j}(\mathcal{T}') = S(i, j)$ for all $i, j \in \{1, \dots, n\}$, the quartet system of T' is contained in the quartet system of T and, defined $T_{\leq s}$ the subforest of T whose edge set consists of edges whose removal results in one of the components having size at most s , we have $T_{\leq n-k} \cong T'_{\leq n-k}$.

Theorem 6. (Pachter-Speyer, [14]). *Let $k, n \in \mathbb{N}$ with $3 \leq k \leq (n+1)/2$. A positive-weighted tree \mathcal{T} with leaves $1, \dots, n$ and no vertices of degree 2 is determined by the values $D_I(\mathcal{T})$, where I varies in $\binom{\{1, \dots, n\}}{k}$.*

Pachter-Speyer Theorem raises naturally some questions. It is natural to ask what happens if $k > (n + 1)/2$: Pachter and Speyer showed that in this case the statement does not hold, but is there a class of trees such that at most one tree in this class realizes a given family $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{k}}$?

Furthermore we can wonder what happens if we consider general weighted trees instead of positive-weighted ones.

Finally, if a family $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{k}}$ is realized by more than one (positive-)weighted tree, can we say something about the ones, among the weighted trees realizing the family, that have maximum/minimum total weight (if they exist)?

To state our results we need the following definition.

Definition 7. *Let $k \in \mathbb{N} - \{0\}$. We say that a tree P is a **pseudostar** of kind (n, k) if $L(P) = n$ and any edge of P divides $L(P)$ into two sets such that at least one of them has cardinality greater than or equal to k .*

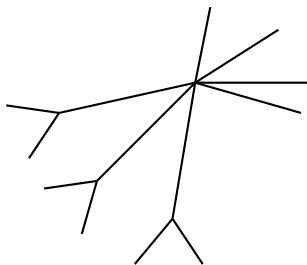


Figure 1: A pseudostar of kind $(10, 8)$

We prove that, if $3 \leq k \leq n - 1$, given a l-treelike family of real numbers, $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{k}}$, there exists exactly one internal-nonzero-weighted pseudostar \mathcal{P} of kind (n, k) with leaves $1, \dots, n$ and no vertices of degree 2 such that $D_I(\mathcal{P}) = D_I$ for any I ; it is positive-weighted if the family is p-l-treelike. Moreover, any other tree realizing the family $\{D_I\}_I$ and without vertices of degree 2 is obtained from the pseudostar by a certain kind of operations we call “OI operations” and by inserting some internal edges of weight 0 (see Definition 11 and Theorem 16). In particular we get that the statement of Pachter-Speyer Theorem holds also for general weighted trees.

Finally, in §4, given a p-l-treelike family $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{k}}$ in the set of positive real numbers, we examine the range of the total weight of the trees realizing it and we show that the pseudostar of kind (n, k) realizing it has maximum total weight (see Theorem 18); then we study the analogous problem for l-treelike families in \mathbb{R} (see Theorem 19).

2 Notation and some remarks

Notation 8. • Let $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x > 0\}$.

• For any $n \in \mathbb{N}$ with $n \geq 1$, let $[n] = \{1, \dots, n\}$.

• For any set S and $k \in \mathbb{N}$, let $\binom{S}{k}$ be the set of the k -subsets of S .

- Throughout the paper, the word “tree” will denote a finite tree.
- We say that a vertex of a tree is a **node** if its degree is greater than 2.
- Let F be a leaf of a tree T . Let N be the node such that the path p between N and F does not contain any node apart from N . We say that p is the **twig** associated to F . We say that an edge is **internal** if it is not an edge of a twig.
- We say that a tree is **essential** if it has no vertices of degree 2.
- Let T be a tree and let $\{i, j\} \in E(T)$. We say that a tree T' is obtained from T by **contracting** $\{i, j\}$ if there exists a map $\varphi : V(T) \rightarrow V(T')$ such that:

$$\varphi(i) = \varphi(j),$$

$$\varphi^{-1}(y) \text{ is a set with only one element for any } y \neq \varphi(i),$$

$$E(T') = \{\{\varphi(a), \varphi(b)\} \mid \{a, b\} \in E(T) \text{ with } \varphi(a) \neq \varphi(b)\}.$$

We say also that T is obtained from T' by **inserting** an edge.

- Let T be a tree and let S be a subset of $L(T)$. We denote by $T|_S$ the minimal subtree of T whose set of vertices contains S . If $\mathcal{T} = (T, w)$ is a weighted tree, we denote by $\mathcal{T}|_S$ the tree $T|_S$ with the weight induced by w .
- Let $\mathcal{T} = (T, w)$ be a weighted tree. We denote $w(T)$ by $D_{\text{tot}}(\mathcal{T})$ and we call it **total weight** of \mathcal{T} .
- Let $n, k \in \mathbb{N}$, $n \geq 3$ and $1 < k < n$. Given a family of real numbers $\{D_I\}_{I \in \binom{[n]}{k}}$, we say that a **weighted tree** $\mathcal{T} = (T, w)$ with $L(T) = [n]$ **realizes the family** $\{D_I\}_I$ if $D_I(\mathcal{T}) = D_I$ for any $I \in \binom{[n]}{k}$.

Definition 9. Let T be a tree.

We say that two leaves i and j of T are **neighbours** if in the path from i to j there is only one node; furthermore, we say that $C \subset L(T)$ is a **cherry** if any $i, j \in C$ are neighbours.

We say that a cherry is **complete** if it is not strictly contained in another cherry.

The **stalk** of a cherry is the unique node in the path with endpoints any two elements of the cherry. Let $i, j, l, m \in L(T)$. We say that $\langle i, j|l, m \rangle$ holds if in $T|_{\{i, j, l, m\}}$ we have that i and j are neighbours, l and m are neighbours, and i and l are not neighbours; in this case we denote by $\gamma_{i, j, l, m}$ the path between the stalk of $\{i, j\}$ and the stalk of $\{l, m\}$ in $T|_{\{i, j, l, m\}}$. The symbol $\langle i, j|l, m \rangle$ is called **Buneman’s index** of i, j, l, m .

Remark 10. (i) A pseudostar of kind $(n, n - 1)$ is a star, that is, a tree with only one node.

(ii) Let $k, n \in \mathbb{N} - \{0\}$. If $\frac{n}{2} \geq k$, then every tree with n leaves is a pseudostar of kind (n, k) , in fact if we divide a set with n elements into two parts, at least one of them has cardinality greater than or equal to $\frac{n}{2}$, which is greater than or equal to k .

Definition 11. Let $k, n \in \mathbb{N} - \{0\}$. Let $\mathcal{T} = (T, w)$ be a weighted tree with $L(T) = [n]$. Let e be an edge of T with weight y and dividing $[n]$ into two sets such that each of them has strictly less than k elements. Contract e and add y/k to the weight of every twig of the tree. We call this operation a k -IO operation on \mathcal{T} and we call the inverse operation a k -OI operation.

Remark 12. It is easy to check that, if $\mathcal{T} = (T, w)$ and $\mathcal{T}' = (T', w')$ are weighted trees with $L(T) = L(T') = [n]$ and \mathcal{T}' is obtained from \mathcal{T} by a k -IO operation on an edge e of weight y , we

have that \mathcal{T} and \mathcal{T}' have the same k -dissimilarity vector. Furthermore, if \mathcal{T} is positive-weighted we have that $D_{tot}(\mathcal{T}') > D_{tot}(\mathcal{T})$, precisely

$$D_{tot}(\mathcal{T}') = D_{tot}(\mathcal{T}) + \frac{n-k}{k}y.$$

Example. Let $k = 5$, $n = 8$. Consider the weighted trees in Figure 2, where the labelled vertices are the numbers in bold and the other numbers are the weights. The tree on the left is not a pseudostar of kind $(8, 5)$ because of the edge e ; the tree on the right is obtained from the one on the left by a 5-IO operation on e . The 5-weights of the two trees are the same.

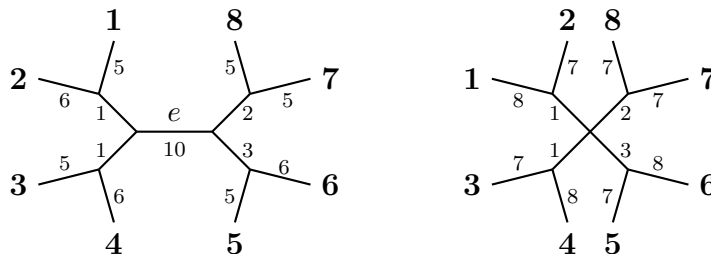


Figure 2: a 5-IO operation

3 Existence and uniqueness of a pseudostar realizing a tree-like family

The following proposition is useful for the proof of Theorem 16 but it can be interesting on its own; it shows that Buneman indices of weighted pseudostars can be recovered from their k -weights. The result was known only for positive-weighted trees.

Proposition 13. *Let $k, n \in \mathbb{N}$ with $2 \leq k \leq n-2$. Let $\mathcal{P} = (P, w)$ be a weighted tree with $L(P) = [n]$.*

1) *Let $i, l \in [n]$.*

(1.1) *If i, l are neighbours, then $D_{i,X}(\mathcal{P}) - D_{l,X}(\mathcal{P})$ does not depend on $X \in \binom{[n]-\{i,l\}}{k-1}$.*

(1.2) *If \mathcal{P} is an internal-nonzero-weighted essential pseudostar of kind (n, k) , then also the converse is true.*

2) *Let $i, j, l, m \in [n]$.*

(2.1) *If $\langle i, j | l, m \rangle$ holds or $P|_{i,j,l,m}$ is a star, then*

$$D_{i,m,R}(\mathcal{P}) + D_{j,l,R}(\mathcal{P}) = D_{i,l,R}(\mathcal{P}) + D_{j,m,R}(\mathcal{P})$$

for any $R \in \binom{[n]-\{i,j,l,m\}}{k-2}$.

(2.2) *Let $k \geq 4$ and \mathcal{P} be an internal-nonzero-weighted essential pseudostar of kind (n, k) . We have that $\langle i, j | l, m \rangle$ holds if and only if at least one of the following conditions holds:*

(a) *$\{i, j\}$ and $\{l, m\}$ are complete cherries in P ,*

(b) there exist $S, R \in \binom{[n] - \{i, j, l, m\}}{k-2}$ such that

$$D_{i,j,S}(\mathcal{P}) + D_{l,m,S}(\mathcal{P}) \neq D_{i,l,S}(\mathcal{P}) + D_{j,m,S}(\mathcal{P}), \quad (1)$$

$$D_{i,j,R}(\mathcal{P}) + D_{l,m,R}(\mathcal{P}) \neq D_{i,m,R}(\mathcal{P}) + D_{j,l,R}(\mathcal{P}). \quad (2)$$

Proof. (1.1) Obvious.

(1.2) Let \mathcal{P} be as in the assumptions and suppose that $D_{i,X}(\mathcal{P}) - D_{l,X}(\mathcal{P})$ does not depend on $X \in \binom{[n] - \{i, l\}}{k-1}$. For every $\delta \in [n]$, let $\bar{\delta}$ be the node on the path from i to l such that

$$\text{path}(i, l) \cap \text{path}(i, \delta) = \text{path}(i, \bar{\delta}).$$

Suppose, contrary to our claim, that i and l are not neighbours. Therefore, on the path between i and l there are at least two nodes. For any a, b nodes in the path between i and l , we say that $a \leq b$ if and only if $\text{path}(i, a) \subset \text{path}(i, b)$. Let x, y be two nodes on the path between i and l such that there is no node in the path between x and y apart from x and y ; thus in the path between x and y there is only one edge since P is essential. Suppose $x < y$, see Figure 3. We can divide $[n]$ into two disjoint subsets:

$$X = \{\delta \in [n] \mid \bar{\delta} \leq x\},$$

$$Y = \{\delta \in [n] \mid \bar{\delta} \geq y\}.$$

Since P is a pseudostar of kind (n, k) , then either $\#X \geq k$ or $\#Y \geq k$; suppose $\#X \geq k$ (we argue analogously in the other case); let $\gamma_1, \dots, \gamma_{k-1}$ be distinct elements of $X - \{i\}$ with $\overline{\gamma_{k-1}} = x$. Up to interchanging the names of $\gamma_1, \dots, \gamma_{k-2}$ (and correspondingly the names of $\overline{\gamma_1}, \dots, \overline{\gamma_{k-2}}$), we can suppose $\overline{\gamma_1} \leq \overline{\gamma_2} \leq \dots \leq \overline{\gamma_{k-1}}$. Let $\eta \in Y - \{l\}$ such that $\bar{\eta} = y$.

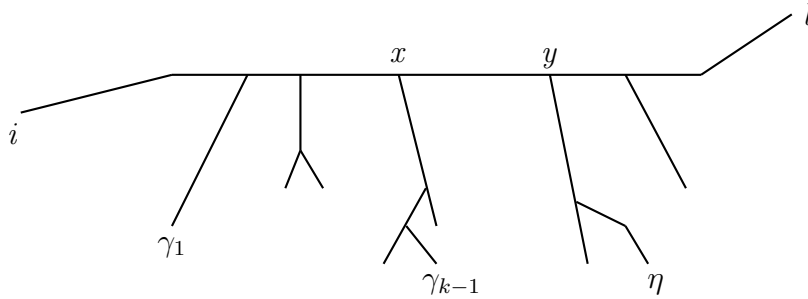


Figure 3: Neighbours in pseudostars

If $k \geq 3$, we have:

$$D_{i, \gamma_1, \dots, \gamma_{k-1}} - D_{l, \gamma_1, \dots, \gamma_{k-1}} = w(\text{path}(i, \overline{\gamma_1})) - w(\text{path}(l, \overline{\gamma_{k-1}}))$$

$$D_{i, \gamma_1, \dots, \gamma_{k-2}, \eta} - D_{l, \gamma_1, \dots, \gamma_{k-2}, \eta} = w(\text{path}(i, \overline{\gamma_1})) - w(\text{path}(l, \bar{\eta})).$$

Since the first members of the equalities above are equal by assumption, we have that

$$w(\text{path}(l, \overline{\gamma_{k-1}})) = w(\text{path}(l, \bar{\eta})),$$

that is

$$w(\text{path}(l, x)) = w(\text{path}(l, y)),$$

thus the weight of the edge $\{x, y\}$ must be 0, which contradicts the assumption.

If $k = 2$, we have:

$$D_{i, \gamma_1} - D_{l, \gamma_1} = w(\text{path}(i, \bar{\gamma}_1)) - w(\text{path}(l, \bar{\gamma}_1)) = w(\text{path}(i, x)) - w(\text{path}(l, x))$$

$$D_{i, \eta} - D_{l, \eta} = w(\text{path}(i, \bar{\eta})) - w(\text{path}(l, \bar{\eta})) = w(\text{path}(i, y)) - w(\text{path}(l, y)).$$

Since the first members of the equalities above are equal, we must have that the weight of the edge $\{x, y\}$ must be 0, which contradicts the assumption.

(2.1) Let $R \in \binom{[n] - \{i, j, l, m\}}{k-2}$, $A = P|_{i, j, l, m, R}$ and $A' = P|_{i, j, l, m}$. Suppose $\langle i, j | l, m \rangle$ holds. Call x the stalk of the cherry $\{i, j\}$ and y the stalk of the cherry $\{l, m\}$ in A' . Let us denote the set of the connected components of $A - A'$ by $C_{A-A'}$. For any $H \in C_{A-A'}$, let v_H be the vertex that is both a vertex of H and a vertex of A' . Then $D_{i, m, R}(\mathcal{P})$ is equal to

$$D_{i, m}(\mathcal{P}) + \sum_{H \in C_{A-A'}} w(H) + w \left(\bigcup_{\substack{H \in C_{A-A'} \text{ s.t.} \\ v_H \in V(\text{path}(j, x))}} \text{path}(v_H, x) \right) + w \left(\bigcup_{\substack{H \in C_{A-A'} \text{ s.t.} \\ v_H \in V(\text{path}(l, y))}} \text{path}(v_H, y) \right).$$

Analogous formulas hold for $D_{i, l, R}(\mathcal{P})$, $D_{j, m, R}(\mathcal{P})$, $D_{j, l, R}(\mathcal{P})$ and we can easily prove our statement. If A' is a star, we argue analogously.

(2.2) \Leftarrow Obviously (a) implies $\langle i, j | l, m \rangle$. Suppose (b) holds; then, if $P|_{i, j, l, m}$ were a star or $\langle i, l | j, m \rangle$ or $\langle i, m | j, l \rangle$ held, from (2.1) we would get a contradiction of the assumptions. Thus $\langle i, j | l, m \rangle$ holds. \Rightarrow Let us consider the path between i and l . We use the same notation as in (1.2). By assumption $\bar{j} < \bar{m}$. Let $m' \in [n]$ be such that \bar{m}' is the maximum node strictly less than \bar{m} and let $j' \in [n]$ be such that \bar{j}' is the minimum node strictly greater than \bar{j} . We could possibly have $j' = m$ and $m' = j$ or $j' = m'$. In Figure 4 we sketch the situation in case $j' < m'$.

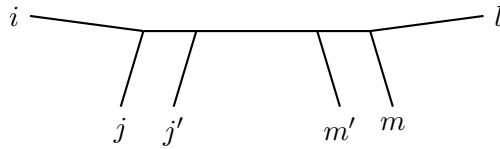


Figure 4: j' and m'

Since P is a pseudostar of kind (n, k) , we have:

$$\#\{x \in [n] \mid \bar{x} \leq \bar{m}'\} \geq k \quad \text{or} \quad \#\{x \in [n] \mid \bar{x} \geq \bar{m}\} \geq k \quad (3)$$

and

$$\#\{x \in [n] \mid \bar{x} \leq \bar{j}'\} \geq k \quad \text{or} \quad \#\{x \in [n] \mid \bar{x} \geq \bar{j}'\} \geq k. \quad (4)$$

• First suppose that there exists $s \in [n] - \{i, j\}$ such that $\bar{s} \leq \bar{j}$ and there exists $t \in [n] - \{l, m\}$ such that $\bar{t} \geq \bar{m}$. From (3) and (4) we get

$$\#\{x \in [n] \mid \bar{x} \leq \bar{m}'\} \geq k \quad \text{or} \quad \#\{x \in [n] \mid \bar{x} \geq \bar{j}'\} \geq k.$$

Suppose $\#\{x \in [n] \mid \bar{x} \leq \bar{m}'\} \geq k$ (the other case is analogous). Let R be a $(k-2)$ -subset of $\{x \in [n] - \{i, j\} \mid \bar{x} \leq \bar{m}'\}$ such that, if $m' \neq j$, then R contains s and m' . Then

$$\begin{aligned} D_{i,j,R}(\mathcal{P}) - D_{l,j,R}(\mathcal{P}) &= w(\text{path}(i, \min(\bar{j} \cup \bar{R})) - w(\text{path}(l, \max(\bar{j} \cup \bar{R}))) \\ &= w(\text{path}(i, \min(\bar{R})) - w(\text{path}(l, \bar{m}'))), \end{aligned}$$

$$\begin{aligned} D_{i,m,R}(\mathcal{P}) - D_{l,m,R}(\mathcal{P}) &= w(\text{path}(i, \min(\bar{m} \cup \bar{R})) - w(\text{path}(l, \max(\bar{m} \cup \bar{R}))) \\ &= w(\text{path}(i, \min(\bar{R})) - w(\text{path}(l, \bar{m}))) \end{aligned}$$

So we get that $D_{i,j,R}(\mathcal{P}) - D_{l,j,R}(\mathcal{P}) - D_{i,m,R}(\mathcal{P}) + D_{l,m,R}(\mathcal{P}) = -w(\{\bar{m}', \bar{m}\})$, which is nonzero by assumption. Thus $D_{i,j,R}(\mathcal{P}) + D_{l,m,R}(\mathcal{P}) \neq D_{l,j,R}(\mathcal{P}) + D_{i,m,R}(\mathcal{P})$; hence (2) holds.

• Now suppose that there exists $s \in [n] - \{i, j\}$ such that $\bar{s} \leq \bar{j}$ and there does not exist $t \in [n] - \{l, m\}$ such that $\bar{t} \geq \bar{m}$ (analogously if the converse holds). Then $\#\{x \in [n] \mid \bar{x} \leq \bar{m}'\} \geq k$. By taking R to be a $(k-2)$ -subset of $\{x \in [n] - \{i, j\} \mid \bar{x} \leq \bar{m}'\}$ such that, if $m' \neq j$, then R contains s and m' , we conclude as above that (2) holds.

• Finally, if there does not exist $s \in [n] - \{i, j\}$ such that $\bar{s} \leq \bar{j}$ and there does not exist $t \in [n] - \{l, m\}$ such that $\bar{t} \geq \bar{m}$, then (a) holds.

By considering the path between i and m , we get analogously that either (1) holds or (a) holds. \square

The following proposition characterizes Buneman indices in terms of k -weights in the case $k = 3$.

Proposition 14. *Let $n \geq 5$. Let $\mathcal{P} = (P, w)$ be an essential and internal-nonzero-weighted tree with $L(\mathcal{P}) = [n]$ (so it is a pseudostar of kind $(n, 3)$). Let $i, j, l, m \in [n]$.*

We have that $\langle i, j \mid l, m \rangle$ holds if and only if at least one of the following conditions holds:

(a) *there exists $r \in [n] - \{i, j, l, m\}$ such that the inequality*

$$D_{i,j,l}(\mathcal{P}) + D_{m,r,l}(\mathcal{P}) \neq D_{i,r,l}(\mathcal{P}) + D_{m,j,l}(\mathcal{P})$$

holds and the inequalities obtained from this by swapping i with j and/or l with m hold.

(b) *for any $r \in [n] - \{i, j, l, m\}$, the following inequalities hold:*

$$D_{i,j,r}(\mathcal{P}) + D_{m,l,r}(\mathcal{P}) \neq D_{i,m,r}(\mathcal{P}) + D_{j,l,r}(\mathcal{P}),$$

$$D_{i,j,r}(\mathcal{P}) + D_{m,l,r}(\mathcal{P}) \neq D_{i,l,r}(\mathcal{P}) + D_{j,m,r}(\mathcal{P}).$$

Proof. \Rightarrow Let x be the stalk of the cherry $\{i, j\}$ in $P|_{i,j,l,m}$ and let y be the stalk of the cherry $\{l, m\}$ in $P|_{i,j,l,m}$. Suppose first that in $V(\gamma_{i,j,l,m})$ there are some nodes of P different from x and y ; call c the node of P in $V(\gamma_{i,j,l,m}) - \{x, y\}$ such that $\text{path}(x, c) \subset \text{path}(x, c')$ for any c' node of P in $V(\gamma_{i,j,l,m}) - \{x, y\}$ (that is c is the node in $\gamma_{i,j,l,m}$ “nearest” to x). Let $r \in [n]$ be such that $\text{path}(x, y) \cap \text{path}(x, r) = \text{path}(x, c)$. For such an r , we have the inequalities in (a), in fact the edge $\{x, c\}$ has nonzero weight by assumption. Thus, if (a) does not hold, then there are no nodes of P in $V(\gamma_{i,j,l,m}) - \{x, y\}$, hence $\gamma_{i,j,l,m}$ is an edge; by assumption $w(\gamma_{i,j,l,m}) \neq 0$ and we can prove easily that (b) holds.

\Leftarrow We can easily prove that, if (a) holds or (b) holds, then $P|_{i,j,l,m}$ is not a star and $\langle i, m \mid j, l \rangle$ and $\langle i, l \mid j, m \rangle$ do not hold. \square

Corollary 15. *Let $n, k \in \mathbb{N}$ with $3 \leq k \leq n - 2$. Let $\{D_I\}_{I \in \binom{[n]}{k}}$ be a family of real numbers. The D_I for $I \in \binom{[n]}{k}$ determine the Buneman's indices of an internal-nonzero-weighted essential pseudostar $\mathcal{P} = (P, w)$ of kind (n, k) with $L(P) = [n]$ and realizing the family $\{D_I\}_I$.*

In fact, by part 1 of Proposition 13, the D_I for $I \in \binom{[n]}{k}$ determine the complete cherries of an internal-nonzero-weighted essential pseudostar $\mathcal{P} = (P, w)$ of kind (n, k) with $L(P) = [n]$ and realizing the family $\{D_I\}_I$; so, by part 2 for $k \geq 4$ they determine its Buneman's indices. For $k = 3$ we can use Proposition 14.

Theorem 16. *Let $n, k \in \mathbb{N}$ with $3 \leq k \leq n - 1$. Let $\{D_I\}_{I \in \binom{[n]}{k}}$ be a family of real numbers. If it is l -treelike, then there exists exactly one internal-nonzero-weighted essential pseudostar \mathcal{P} of kind (n, k) realizing the family. Any other weighted essential tree realizing the family $\{D_I\}_I$ can be obtained from \mathcal{P} by k -OI operations and by inserting internal edges of weight 0.*

If the family $\{D_I\}_{I \in \binom{[n]}{k}}$ is p - l -treelike, then \mathcal{P} is positive-weighted and any other positive-weighted essential tree realizing the family $\{D_I\}_I$ can be obtained from \mathcal{P} by k -OI operations.

Proof. Let $\mathcal{T} = (T, w)$ be a weighted tree with $L(T) = [n]$ and realizing the family $\{D_I\}_{I \in \binom{[n]}{k}}$. Obviously we can suppose that it is essential. By k -IO operations and contracting the internal edges of weight 0 we can change \mathcal{T} into an internal-nonzero-weighted essential pseudostar \mathcal{P} of kind (n, k) ; it realizes the family $\{D_I\}_I$ by Remark 12. If \mathcal{T} is positive-weighted, obviously also \mathcal{P} is positive-weighted.

If $k = n - 1$, it is easy to see that there exists at most a weighted essential star with leaves $1, \dots, n$ realizing the family $\{D_I\}_I$.

Suppose $k \leq n - 2$. By Corollary 15, the D_I for $I \in \binom{[n]}{k}$ determine the Buneman's indices, and then they determine P , in fact it is well known that the Buneman's indices of a tree determine the tree (see for instance [7]). We have to show that the D_I determine also the weights of the edges of \mathcal{P} . The argument is completely analogous to the proof of the theorem in [14]; we sketch it for the convenience of the reader. Let e be an edge of P which is not a twig. Then there exist $i, j, l, m \in [n]$ such that $e = \gamma_{i,j,l,m}$ (see Definition 9 for the meaning of $\gamma_{i,j,l,m}$); since P is a pseudostar of kind (n, k) , there exists $R \in \binom{[n] - \{i,j,l,m\}}{k-2}$ such that e is not an edge of $P|_R$. Then

$$2w(e) = D_{i,m,R}(\mathcal{P}) + D_{j,l,R}(\mathcal{P}) - D_{i,j,R}(\mathcal{P}) - D_{l,m,R}(\mathcal{P}),$$

so $w(e)$ is determined by the D_I . For any $I \in \binom{[n]}{k}$ we have that

$$D_I(\mathcal{P}) = \sum_{\substack{e \in E(T|_I) \\ e \text{ not twig}}} w(e) + \sum_{i \in I} w(e_i), \quad (5)$$

where e_i denotes the twig associated to i .

So, for any $i, j \in [n]$ and for any $S \in \binom{[n] - \{i,j\}}{k-1}$, we have:

$$w(e_i) - w(e_j) = \sum_{l \in (iS)} w(e_l) - \sum_{l \in (jS)} w(e_l) = \left(D_{iS} - \sum_{\substack{e \in E(T|_{iS}) \\ e \text{ not twig}}} w(e) \right) - \left(D_{jS} - \sum_{\substack{e \in E(T|_{jS}) \\ e \text{ not twig}}} w(e) \right).$$

Hence the difference of the weights of the twigs is determined by the D_I . From the formula (5) we get the weight of every twig. \square

Corollary 17. *The statement of Pachter-Speyer Theorem holds also for general weighted trees.*

We point out that the unicity statement of Theorem 16 in the case of positive-weighted trees can be deduced also from Theorem 5; anyway it was not stated there (the definition of pseudostar is new). Observe that the complexity of an IO -operation on a tree with n leaves is $O(n)$ and the internal edges of an essential tree with n leaves are at most n , so changing an essential tree into a pseudostar of kind (n, k) requires not more than $O(n^2)$ elementary operations.

Finally we want to mention that in [2], by using also the idea of pseudostar introduced in this paper, we give a characterization of treelike families of real numbers parametrized by k -subsets of a finite set and that the website <http://web.math.unifi.it/users/baldisser/Downloads.html> contains a program (which uses the ideas of [2]) whose aim is to establish if a family of real numbers is treelike and, if so, to compute the unique internal-nonzero-weighted essential pseudostar realizing the family.

4 The range of the total weight

Let $\{D_I\}_{I \in \binom{[n]}{k}}$ be a p -l-treelike family in \mathbb{R}_+ . If $2 \leq k \leq (n+1)/2$ we know that there exists a unique positive-weighted essential tree $\mathcal{T} = (T, w)$ with $L(\mathcal{T}) = [n]$ and realizing the family (see Theorem 6). On the other hand, for $k > (n+1)/2$ this statement no longer holds and, if we call U the set of all positive-weighted trees realizing the family $\{D_I\}_I$, we can wonder which is the range of the total weight of the weighted trees in U .

Theorem 18. *Let $k, n \in \mathbb{N}$ with $3 \leq k \leq n-1$. Let $\{D_I\}_{I \in \binom{[n]}{k}}$ be a p -l-treelike family of positive real numbers and let U be the set of the positive-weighted trees with $[n]$ as set of leaves and realizing the family $\{D_I\}_I$. Call \mathcal{P} the unique essential pseudostar of kind (n, k) in U (see Theorem 16). The following statements hold:*

(i)

$$\sup_{\mathcal{T} \in U} \{D_{tot}(\mathcal{T})\} = D_{tot}(\mathcal{P})$$

and the supremum is attained only by \mathcal{P} ;

(ii) if $\#U > 1$, then

$$\inf_{\mathcal{T} \in U} \{D_{tot}(\mathcal{T})\} = D_{tot}(\mathcal{P}) - (n-k) \cdot m$$

where m is the minimum among the weights of the twigs of \mathcal{P} ; the infimum is not attained.

Proof. (i) Let $\mathcal{T} = (T, w)$ be a weighted tree in U . Without changing the dissimilarity family and the total weight we can suppose that it is essential. By using several k -IO operations, we can transform it into a pseudostar of kind (n, k) . By Remark 12 the dissimilarity family does not change, so, by Theorem 16, the pseudostar of kind (n, k) we have obtained must be the unique essential pseudostar of kind (n, k) in U , that is \mathcal{P} . By Remark 12 we have that $D_{tot}(\mathcal{T}) \leq D_{tot}(\mathcal{P})$; furthermore, if \mathcal{T} is different from \mathcal{P} , then $D_{tot}(\mathcal{T}) < D_{tot}(\mathcal{P})$.

(ii) Suppose $\#U > 1$. Then we can make a k -OI operation on \mathcal{P} : we add an edge of weight kx , where $x < m$, in such a way that the edge divides the tree in two trees each with less than k leaves, and we subtract x from the weight of every twig of \mathcal{P} . Let \mathcal{T} be the tree we have obtained. We have

$$D_{tot}(\mathcal{T}) = D_{tot}(\mathcal{P}) + k \cdot x - n \cdot x = D_{tot}(\mathcal{P}) - (n - k) \cdot x$$

Obviously, the limit of $D_{tot}(\mathcal{T})$, as x approaches m , is $D_{tot}(\mathcal{P}) - (n - k) \cdot m$.

Finally, let $\mathcal{A} \in U$. Without changing the dissimilarity family and the total weight we can suppose that it is essential. We can transform \mathcal{A} into \mathcal{P} by several k -IO operations, contracting edges with weights y_1, y_2, \dots, y_s and adding $\frac{y_1 + y_2 + \dots + y_s}{k}$ to the weight of every twig. Then we get:

$$D_{tot}(\mathcal{P}) = D_{tot}(\mathcal{A}) + \frac{n - k}{k}(y_1 + y_2 + \dots + y_s). \quad (6)$$

Furthermore, since, to obtain \mathcal{P} from \mathcal{A} , we have added $\frac{y_1 + y_2 + \dots + y_s}{k}$ to the weight of every twig, we have that

$$m > \frac{y_1 + y_2 + \dots + y_s}{k}. \quad (7)$$

Thus, from (6) and (7) we get:

$$D_{tot}(\mathcal{A}) = D_{tot}(\mathcal{P}) - \frac{n - k}{k}(y_1 + y_2 + \dots + y_s) > D_{tot}(\mathcal{P}) - (n - k) \cdot m.$$

□

The following theorem answers the analogous problem for general weighted trees.

Theorem 19. *Let $k, n \in \mathbb{N}$ with $3 \leq k \leq n - 1$. Let $\{D_I\}_{I \in \binom{[n]}{k}}$ be a l -treelike family of real numbers and let U be the set of weighted trees with $[n]$ as set of leaves and realizing the family $\{D_I\}_I$.*

(i) *If in U there are only weighted pseudostars of kind (n, k) (for instance if $k \leq \frac{n}{2}$), then $D_{tot}(\mathcal{P}) = D_{tot}(\mathcal{P}')$ for any $\mathcal{P}, \mathcal{P}' \in U$; in particular*

$$\inf_{\mathcal{P} \in U} \{D_{tot}(\mathcal{P})\} = \sup_{\mathcal{P} \in U} \{D_{tot}(\mathcal{P})\}.$$

(ii) *If in U there are weighted trees that are not pseudostars of kind (n, k) , then*

$$\inf_{\mathcal{T} \in U} \{D_{tot}(\mathcal{T})\} = -\infty, \quad \sup_{\mathcal{T} \in U} \{D_{tot}(\mathcal{T})\} = +\infty.$$

Proof. (i) Let \mathcal{P} and \mathcal{P}' be in U . Denote by $\hat{\mathcal{P}}$ and $\hat{\mathcal{P}}'$ the weighted trees obtained respectively from \mathcal{P} and \mathcal{P}' by contracting the internal edges of weight 0. Let $\bar{\mathcal{P}}$ and $\bar{\mathcal{P}}'$ be the weighted essential trees equivalent respectively to $\hat{\mathcal{P}}$ and $\hat{\mathcal{P}}'$. Obviously both $\bar{\mathcal{P}}$ and $\bar{\mathcal{P}}'$ pseudostars of kind (n, k) and realize the family $\{D_I\}_I$; so, by Theorem 16, they are equal. Thus $D_{tot}(\bar{\mathcal{P}}) = D_{tot}(\bar{\mathcal{P}}')$, therefore $D_{tot}(\mathcal{P}) = D_{tot}(\mathcal{P}')$.

(ii) Let $\mathcal{T} = (T, w)$ be an element of U that is not a pseudostar of kind (n, k) ; then there is an edge \bar{e} dividing the tree into two trees such that each of them has less than k leaves. Let $z \in \mathbb{R}$. We define on T a new weight w' as follows:

$$w'(\bar{e}) := w(\bar{e}) + z;$$

for every twig t ,

$$w'(t) := w(t) - \frac{1}{k}z;$$

for any edge e different from \bar{e} and not contained in a twig, we define $w'(e) = w(e)$.

Let $\mathcal{T}' = (T, w')$. We have that $D_I(\mathcal{T}') = D_I$ for any $I \in \binom{[n]}{k}$, so $\mathcal{T}' \in U$. Furthermore

$$D_{tot}(\mathcal{T}') = D_{tot}(\mathcal{T}) + z - \frac{n}{k}z = D_{tot}(\mathcal{T}) - \frac{n-k}{k}z.$$

Hence $\lim_{z \rightarrow -\infty} D_{tot}(\mathcal{T}') = +\infty$, while $\lim_{z \rightarrow +\infty} D_{tot}(\mathcal{T}') = -\infty$. □

Acknowledgements. This work was supported by the National Group for Algebraic and Geometric Structures, and their Applications (GNSAGA-INdAM). The first author was supported by Ente Cassa di Risparmio di Firenze.

References

- [1] A. Baldisserri, E. Rubei, On graphlike k -dissimilarity vectors, *Ann. Comb.* 18 (2014) 356-381
- [2] A. Baldisserri, E. Rubei, Treelike families of multiweights, arXiv:1404.6799, v4 December 2015
- [3] H-J Bandelt, M.A. Steel, Symmetric matrices representable by weighted trees over a cancellative abelian monoid, *SIAM J. Discrete Math.* 8 (1995) 517-525
- [4] C. Bocci, F. Cools, A tropical interpretation of m -dissimilarity maps, *Appl. Math. Comput.* 212 (2009) 349-356
- [5] P. Buneman, A note on the metric properties of trees, *J. Combinatorial Theory Ser. B* 17 (1974) 48-50
- [6] H. Colonius, H.H. Schultze, Tree structure from proximity data, *British J. Math. Statist. Psych.* 34 (1981) 167-180
- [7] A. Dress, K. T. Huber, J. Koolen, V. Moulton, A. Spillner, *Basic phylogenetic combinatorics*, Cambridge University Press, Cambridge, 2012
- [8] A. Dress, Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces, *Adv. Math.* 53 (1984) 321-402
- [9] S.Herrmann, K.Huber, V.Moulton, A.Spillner, Recognizing treelike k -dissimilarities, *J. Classification* 29 (2012) 321-340
- [10] B. Iriarte Giraldo, Dissimilarity vectors of trees are contained in the tropical Grassmannian, *Electron. J. Combin.* 17 (2010)
- [11] D. Levy, R. Yoshida, L. Pachter, Beyond pairwise distances: neighbor-joining with phylogenetic diversity estimates, *Mol. Biol. Evol.* 23 (2006) 491-498

- [12] C. Manon, Dissimilarity maps on trees and the representation theory of $SL_m(\mathbb{C})$, J. Algebraic Combin. 33 (2011) 199-213
- [13] M. Nei, N.Saitou, The neighbor joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406-425
- [14] L. Pachter, D. Speyer, Reconstructing trees from subtree weights, Appl. Math. Lett. 17 (2004) 615-621
- [15] E. Rubei, Sets of double and triple weights of trees, Ann. Comb. 15 (2011) 723-734
- [16] E. Rubei, On dissimilarity vectors of general weighted trees, Discrete Math. 312 (2012) 2872-2880
- [17] C. Semple, M. Steel, Phylogenetics, Oxford University Press, Oxford, 2003
- [18] J.M.S. Simoes Pereira, A Note on the Tree Realizability of a distance matrix, J. Combinatorial Theory 6 (1969) 303-310
- [19] J.A. Studier, K.J. Keppler, A note on the neighbor-joining algorithm of Saitou and Nei, Mol. Biol. Evol. 5 (1988) 729-731
- [20] K.A. Zaretskii, Constructing trees from the set of distances between pendant vertices, Uspehi Mat. Nauk 20 (1965) 90-92