Discrete Applied Mathematics 220 (2017) 35–45

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

A characterization of dissimilarity families of trees

Agnese Baldisserri, Elena Rubei *

Dipartimento di Matematica e Informatica "U. Dini", viale Morgagni 67/A, 50134 Firenze, Italy

A R T I C L E I N F O

Article history:
Received 11 January 2016
Received in revised form 17 September 2016
Accepted 8 December 2016
Available online 10 January 2017

Keywords:
Weighted trees
Dissimilarity families

A B S T R A C T

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

In this paper we find some equalities and inequalities characterizing the families of real numbers parametrized by the k-subsets of $\{1, \ldots, n\}$ that are the families of k-weights of weighted trees whose leaf set is equal to $\{1, \ldots, n\}$ and whose weights of the internal edges are positive (where we say that an edge e is internal if there exists a path with endpoints of degree greater than 2 and containing e).

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 $w : E(G) \to \mathbb{R}$. 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*). We say that an edge e is internal if there exists a path with endpoints of degree greater than 2 and containing e. If the weights of the internal edges are nonzero, we say that the graph is *internal-nonzero-weighted* and, if the weights of the 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 $i_1, \ldots, i_k \in V(T)$, we define $D_{\{i_1,\ldots,i_k\}}(\mathcal{T})$ to be the weight of the minimal subtree containing $i_1, \ldots, i_k$. We call this subtree "the subtree realizing $D_{\{i_1,\ldots,i_k\}}(\mathcal{T})$". More simply, we denote $D_{\{i_1,\ldots,i_k\}}(\mathcal{T})$ by $D_{i_1,\ldots,i_k}(\mathcal{T})$ for any order of $i_1, \ldots, i_k$. We call the $D_{i_1,\ldots,i_k}(\mathcal{T})$ the *k-weights* of $\mathcal{T}$ and we call a k-weight of $\mathcal{T}$ for some k a *multiweight* of $\mathcal{T}$.

Throughout the paper we will use the following standard notation: for any set Z and $k \in \mathbb{N}$, we denote by $\binom{Z}{k}$ the set of the k-subsets of Z. If Z is a subset of $V(T)$, the k-weights $D_{i_1,\ldots,i_k}(\mathcal{T})$ with $i_1, \ldots, i_k \in Z$ give a vector in $\mathbb{R}^{\sharp\binom{Z}{k}}$. This vector is called *k-dissimilarity vector* of $(\mathcal{T}, Z)$. Equivalently, we can speak of the *family of the k-weights* of $(\mathcal{T}, Z)$ or of the *k-dissimilarity family of* $(\mathcal{T}, Z)$. The vertices in Z are said *labelled*.

* Corresponding author.
E-mail address: rubei@math.unifi.it (E. Rubei).

http://dx.doi.org/10.1016/j.dam.2016.12.007
0166-218X/

If Z is a finite set, $k \in \mathbb{N}$ and $k < \#Z$, we say that a family of real numbers $\{D_I\}_{I \in \binom{Z}{k}}$ 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 Z of the set of its vertices such that $D_I(\mathcal{T}) = D_I$ for any k-subset I of Z. In this case, we say also that $\mathcal{T}$ realizes the family $\{D_I\}_{I \in \binom{Z}{k}}$. If in addition $Z \subset L(T)$, we say that the family is *l-treelike* (respectively p-l-treelike, nn-l-treelike, inz-l-treelike, ip-l-treelike). In this paper we will consider only the problem of l-treelikeness, that is, we will consider labels only on the leaves.

Weighted graphs have applications in several disciplines, such as biology, psychology (see for instance [7]), archaeology, engineering. Phylogenetic trees are 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 weighted trees from their dissimilarity families; these methods, for instance the so-called neighbour-joining method, are used by biologists to reconstruct phylogenetic trees. See for example [16,20] and, for overviews, [8,17]. Weighted graphs can also represent hydraulic webs or railway webs where the weight of a line is the difference between the earnings and the cost of the line or the length of the line. It can be interesting, given a family of real numbers, $\{D_{i_1,\ldots,i_k}\}_{i_1,\ldots,i_k}$, to wonder if there exists a weighted tree with it as family of k-weights; moreover the study of the subset of the treelike vectors, and in particular of the equalities and inequalities characterizing it, can be useful if we search a tree whose k-weights have some characteristics, for instance satisfy some given equalities or inequalities. We recall also that the importance of the study of k-weights for $k \geq 3$ is due to the fact that they seem statistically more reliable than 2-weights (see [13,19]).

We recall the most important results concerning treelike dissimilarity families.

A criterion for a metric on a finite set to be nn-l-treelike was established in [6,18,21]:

Definition 2. Let $n \in \mathbb{N}$ with $n \geq 4$. Let $\{D_I\}_{I \in \binom{\{1,\ldots,n\}}{2}}$ be a family of positive real numbers. We say that the D_I satisfy the *4-point condition* if and only if, for all distinct $a, b, c, d \in \{1, \ldots, n\}$, 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.

Theorem 3. *Let $n \in \mathbb{N}$ with $n \geq 4$. Let $\{D_I\}_{I \in \binom{\{1,\ldots,n\}}{2}}$ be a family of positive real numbers satisfying the triangle inequalities. It is p-treelike (or nn-l-treelike) if and only if the 4-point condition holds.*

Also the study of general weighted trees can be interesting and, in 1995, Bandelt and Steel proved a result, analogous to Theorem 3, for general weighted trees:

Theorem 4 (*Bandelt–Steel, [3, Theorem 1]*)**.** *Let $n \in \mathbb{N}$ with $n \geq 4$. For any family of real numbers $\{D_I\}_{I \in \binom{\{1,\ldots,n\}}{2}}$, there exists a weighted tree $\mathcal{T}$ with leaves $1, \ldots, n$ such that $D_I(\mathcal{T}) = D_I$ for any $I \in \binom{\{1,\ldots,n\}}{2}$ if and only if the so-called relaxed 4-point condition holds, i.e., for any $a, b, c, d \in \{1, \ldots, n\}$, at least two among $D_{a,b} + D_{c,d}$, $D_{a,c} + D_{b,d}$, $D_{a,d} + D_{b,c}$ are equal.*

An easy variant of the theorems above is the following:

Theorem 5. *Let $n \in \mathbb{N}$ with $n \geq 4$. For any family of real numbers $\{D_I\}_{I \in \binom{\{1,\ldots,n\}}{2}}$, there exists an internal-positive weighted tree $\mathcal{T}$ with leaves $1, \ldots, n$ such that $D_I(\mathcal{T}) = D_I$ for any $I \in \binom{\{1,\ldots,n\}}{2}$ if and only if the 4-point condition holds.*

In fact, if the 4-point condition holds, in particular the relaxed 4-point condition holds; so by Theorem 4, there exists a weighted tree $\mathcal{T}$ with leaves $1, \ldots, n$ and with 2-weights equal to the D_I; it is easy to see that, since the 4-point condition holds, the weights of the internal edges of $\mathcal{T}$ are nonnegative; by contracting the internal edges of weight 0, we get an ip-weighted tree with leaves $1, \ldots, n$ and with 2-weights equal to the D_I.

For higher k the literature is more recent, see [1,4,9–15]. Three of the most important results for k-weights with $k \geq 3$ are the following.

Theorem 6 (*Herrmann, Huber, Moulton, Spillner, [9, Theorem 2]*)**.** *Let $n, k \in \mathbb{N} - \{0\}$. If $n \geq 2k$, a family of positive real numbers $\{D_I\}_{I \in \binom{\{1,\ldots,n\}}{k}}$ is ip-l-treelike if and only if its restriction to every $2k$-subset of $\{1, \ldots, n\}$ is ip-l-treelike.*

Theorem 7 (*Pachter–Speyer, [13, Theorem 1]*)**.** *Let $k, n \in \mathbb{N}$ with $3 \leq k \leq (n + 1)/2$. A positive-weighted tree $\mathcal{T}$ with leaves $1, \ldots, n$ and no vertices of degree 2 is determined by the values $D_I(\mathcal{T})$, where I varies in $\binom{\{1,\ldots,n\}}{k}$.*