ELSEVIER

Discrete Mathematics 192 (1998) 223–249

DISCRETE MATHEMATICS

On some relations between 2-trees and tree metrics

Bruno Leclerc[a,*], Vladimir Makarenkov[b]

[a] *Centre d'Analyse et de Mathématique Sociales, École des Hautes Études en Sciences Sociales, 54 bd Raspail, F-75270 Paris, Cedex 06, France*

[b] *Institute of Control Sciences, 65 Profsoyuznaya, Moscow 117806, Russia and Centre d'Analyse et de Mathématique Sociales, École des Hautes Études en Sciences Sociales, 54 bd Raspail, F-75270 Paris, Cedex 06, France*

Received 17 January 1997; revised 11 July 1997; accepted 22 July 1997

Abstract

A *tree function* (TF) t on a finite set X is a real function on the set of the pairs of elements of X satisfying the four-point condition: for all *distinct* $x,y,z,w \in X, t(xy)+t(zw) \leqslant \max\{t(xz)+t(yw), t(xw)+t(yz)\}$. Equivalently, t is representable by the lengths of the paths between the leaves of a valued tree T_ℓ. TFs are a straightforward generalization of the tree dissimilarities and tree metrics of the literature. A graph Θ is a 2-*tree* if it belongs to the following class $\mathscr{Q}$: an edge-graph belongs to $\mathscr{Q}$: if $\Theta' \in \mathscr{Q}$ and yz is an edge of Θ', then the graph obtained by the addition to Θ' of a new vertex x adjacent to y and z belongs to $\mathscr{Q}$. These graphs, and the more general k-trees, have been studied in the literature as generalizations of trees. It is first explicited here how to make a TF $t_{\Theta,d}$ correspond to any positively valued 2-tree Θ_d on X. Then, given a tree dissimilarity t, the set $\mathscr{Q}(t)$ of the 2-trees Θ such that $t=t_{\Theta,t}$ is studied. Any element of $\mathscr{Q}(t)$ gives a way of summarizing t by its restriction to a minimal subset of entries. Several characterizations and properties of the elements of $\mathscr{Q}(t)$ are given. We describe five classes of such elements, including two new ones. Associated with a dissimilarity of the general type, these classes of 2-trees lead to methods for the recognition and fitting of tree dissimilarities.

Résumé

Une fonction d'arbre (TF) t sur un ensemble fini X est une fonction réelle sur l'ensemble des parties à deux éléments de X vérifiant la condition des quatre points: pour tous $x,y,z,w \in X$, *distincts*, $t(xy)+t(zw) \leqslant \max\{t(xz)+t(yw), t(xw)+t(yz)\}$. De façon équivalente, t est représentable par les longueurs des chemins entre les feuilles d'un arbre valué T_ℓ. Ces fonctions constituent une généralisation immédiate des dissimilarités et distances d'arbres de la littérature. Un graphe Θ est un 2-*arbre* s'il appartient à la classe $\mathscr{Q}$ suivante: un graphe réduit à deux sommets adjacents appartient à $\mathscr{Q}$; si $\Theta' \in \mathscr{Q}$ et si yz est une arête de Θ', le graphe obtenu en ajoutant à Θ' un nouveau sommet x adjacent à y et z appartient à $\mathscr{Q}$. Ces graphes, et plus généralement les k-arbres, ont été étudiés dans la littérature comme généralisations des arbres. On montre d'abord ici comment une TF $t_{\Theta,d}$ correspond à tout 2-arbre positivement valué Θ_d sur X. Puis, étant donnée une dissimilarité d'arbre t, on étudie l'ensemble $\mathscr{Q}(t)$ des 2-arbres Θ tels que $t=t_{\Theta,t}$. Plusieurs caractérisations et propriétés de ses éléments sont obtenues. Tout élément de $\mathscr{Q}(t)$

* Corresponding author. E-mail: leclerc@ehess.fr.

PII S0012-365X(98)00073-9

fournit un résumé de t par sa restriction à un sous-ensemble d'entrées minimal; nous décrivons cinq classes de tels éléments, dont deux entièrement nouvelles. Associés aux dissimilarités quelconques, ces classes de 2-arbres conduisent à des méthodes de reconnaissance et d'ajustement des distances d'arbres.

1. Introduction

Ultrametrics and tree metrics are two well-known classes of metrics, both defined by a particular inequality. The ultrametric triangular inequality $u(x,z) \leqslant \max(u(x,y), u(y,z))$ is a special case of the four-point condition $t(x,y)+t(z,w) \leqslant \max\{t(x,z)+t(y,w), t(x,w)+t(y,z)\}$ satisfied by tree metrics (precise definitions are given in Section 2.2). Tree metrics are representable as path lengths between the leaves of a positively valued tree; the ultrametric case corresponds to the existence of a point equidistant from all the leaves in this tree representation.

The combinatorial study of the ultrametrics on a given finite set X leads to many properties related with circuits, cocycles and spanning trees of the complete graph K_X on X, and, in fact, with the cycle matroid of this graph [14,15,19]. Especially, every minimum spanning tree provides an abridgement (a summarization) of an ultrametric into $n-1$ entries: an ultrametric u on X is entirely defined by its restriction to the edges of a minimum spanning tree of K_X valued according to u.

Chaiken et al. [6] and Yushmanov [25] have independently associated to any tree metric t some sets of $2n-3$ pairs of elements of X such that t is defined by its restriction to these sets. Two facts have been pointed out about these works: on the one hand, it is shown in Makarenkov and Leclerc [20] that the abridgements of Chaiken et al. and Yushmanov are in fact the same (though defined in entirely different ways); but, on the other hand, two other abridgements, into $2n-3$ entries again, are obtained in Leclerc [16,18]; two new ones are also given in this paper. The existence of such constructions suggest that the properties of minimum spanning trees in relation with ultrametrics may extend, at least partly, as properties of some class of valued graphs in relation with tree metrics. The first purpose of this paper is to confirm this hypothesis, with the graphs called 2-trees instead of trees. The valued 2-trees related with tree metrics are those characterized in Theorems 4.2 and 4.7 of Section 4.1. Some combinatorial properties of tree metrics are also obtained, for instance a type of equality satisfied by tree dissimilarities on the 2-paths of their 2-trees abridgements (2-paths have in 2-trees a role similar to the paths of a tree). The uses of 2-trees abridgements in classical problems of data analysis, recognition and adjustment, are also discussed, and the works of Leclerc [18] and Makarenkov and Leclerc [20] about some particular abridgements are generalized.

The paper is organized as follows: definitions about graphs, trees and 2-trees are recalled in Section 2.1. An important feature of 2-trees is that they admit *elimination* (linear) *orders* $(x_1, x_2, \ldots, x_n)$ on X such that, for all $i=3,4,\ldots,n-1$, the subgraph induced by $x_1, \ldots, x_i$ is still a 2-tree. Definitions about dissimilarities are given in

Section 2.2, where several extensions of tree metrics, allowing negative values, are also considered; they are called tree dissimilarities and tree functions. In Section 3, a tree function is associated with any positively valued 2-tree Θ, endowed with an elimination order. It is then shown that this function does not depend in fact on the choice of a particular elimination order (Theorem 3.2) and is a tree metric with a simple condition (Proposition 3.3). The elements of the set $\mathscr{Q}(t)$ of the 2-tree abridgements of a given tree dissimilarity t are studied in Section 4. They are characterized by a property of their triangles in Theorem 4.2 and by two minimality properties in Theorem 4.7. It is also observed that the set $\mathscr{Q}(t)$ has in fact various elements and that the constructions of Yushmanov and others are far from being exceptions; two new classes of such graphs are described. In Section 5, the set $\mathscr{Q}(d)$ is generalized to any dissimilarity d and two ways of using an element of this set for the adjustment of a tree dissimilarity, or a tree metric, to d are proposed.

2. Definitions

2.1. Graphs

2.1.1. Graphs and trees

A *graph* is a pair $G=(V(G),E(G))$ where $V(G)$ is a finite *vertex set* and $E(G)$ is a set of unordered pairs of distinct elements of $V(G)$, the *edge set* of G. For sake of brevity, an edge is denoted vv' instead of $\{v,v'\}$; the vertices v and v' are said to be *adjacent*. The number of vertices of G is denoted as $\mu(G)$. When there is no risk of confusion, we write $V,E,\mu,\dots$ instead of $V(G),E(G),\mu(G),\dots$ The vertex v is a *leaf* if its degree $\partial(v)$ is equal to 1 and a 2-*leaf* if $\partial(v)=2$. In a *path* $(vv_1,v_1v_2,\dots,v_{k-1}v_k,v_kv')$ of G between two vertices v and v', all the vertices other than v and v' will be distinct and a path P will be identified with the set of its edges. If, moreover, $v=v'$, the path P is a *circuit* of G. The graph G is a *tree* if it is connected and has no circuits. A tree T has exactly $\mu(T)-1$ edges; its unique path between two distinct vertices v and v' is denoted as $T(vv')$. It has a number of leaves $\mu_1(T)$ comprised between 2 and $\mu(T)-1$. The tree T is a *path-tree* if $\mu_1(T)=2$, and a $(\mu-1)$-*star* if $\mu_1(T)=\mu-1$. The graph G is a *k-clique* if $\mu(G)=k$ and $uv\in E$ for all $u,v\in V(G)$. A *triangle of* G is a subset of $V(G)$ inducing a 3-clique; such a subset is denoted xyz instead of $\{x,y,z\}$.

A *valued graph* is an ordered pair $G_\ell=(G,\ell)$, where G is a graph and ℓ is a real *length* function on the edge set $E(T)$ of T. Then, $\ell(G)=\sum_{e\in E(G)}\ell(e)$. When this quantity is defined (that is when the graph is connected and has no circuit of negative length), we set, for two distinct vertices v and v' of G:

$$\delta(vv')=\mathrm{Min}_{P\text{ path of }G\text{ between }v\text{ and }v'}\sum_{e\in T(vv')}\ell(e).$$

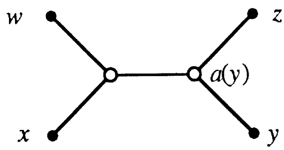

Fig. 1.

When T is a tree, $\delta(vv')$ is always defined as

$$\delta(vv') = \sum_{e \in T(vv')} \ell(e).$$

Let X be a finite set with n elements. An XLL-tree (*leaf labelled according to X tree*, frequently called an X-tree in the literature) is a tree satisfying two properties: (i) the leaf set of an XLL-tree T is X; and (ii) for any $v \in V(T) - X$, $\partial(v) \geqslant 3$. In an XLL-tree, the role of the vertices in $V(T) - X$ is just to determine the shape of the tree; they are called *latent* vertices and often indicated without labels in figures. An XLL-tree has at most $2n-2$ vertices and thus $2n-3$ edges, these numbers corresponding to the non-degenerate case where all the latent vertices have degree 3. The *articulation point* $a(x)$ of $x \in X$ is the unique latent vertex v adjacent to x (Fig. 1). In the case of a valued XLL-tree, the length $\ell(xa(x))$ is also denoted as $\rho(x)$. For more definitions and properties of such trees, see [3].

We make a valued XLL-tree T_ℓ correspond to any valued tree $T'_{\ell'}$ with X as set of leaves by repeating the following operation: choose a vertex u of degree two in $V(T') - X$, delete it and replace the edges vu and uv' incident to u with a unique edge vv'; set $\ell(vv') = \ell'(vu) + \ell'(uv')$. When no such vertex remains, an XLL-tree T is obtained; complete the end length function on T by setting $\ell(vv') = \ell'(vv')$ for any edge $vv' \in V(T) \cap V(T')$. The trees T and T' have the same leaves and define the same function δ on $X^{(2)}$. We say that T_ℓ is the *reduced XLL-tree* corresponding with T'_ℓ. For $Y \subseteq X$, the *restricted* YLL-tree $T_{\ell|Y}$ is obtained by deleting all the leaves not in Y and reducing the obtained tree.

2.1.2. k-trees and 2-trees

A graph Θ is a k-tree if it belongs to the class $\mathscr{KT}$ of graphs defined as follows:

- every complete graph on k vertices belongs to $\mathscr{KT}$;
- if G' belongs to $\mathscr{KT}$ and K is a k-clique of G', the graph obtained by the addition to G' of a new vertex x with the edges xc, $c \in K$, belongs to $\mathscr{KT}$.

A k-tree on n vertices has $k(2n-k-1)/2$ edges and $n-k$ $(k+1)$-cliques. A k-tree Θ with exactly 2 vertices, say x and y, of degree k, is called a *k-path*. The k-trees have been studied in the literature as generalizations of trees and for their algorithmic properties (see, for instance, [12, p. 100]; for the number of labelled k-trees on a set with k elements, see [21] or [4]). An induced subgraph of a k-tree Θ which is a k-path (with x and y as vertices of degree k) is a *k-path of* Θ (between x and y). Among the many properties of k-trees, we shall mainly use the following:

- *For $|X|>k$, a k-tree $\Theta=(X,E)$ has at least 2 vertices of degree k. The subgraph induced by the subset $X-\{x\}$ is a k-tree if and only if x is a vertex of degree k of Θ.*
- *If Θ is a k-tree on X, then, for all $x,y\in X$ such that $xy\notin E(\Theta)$, there exists a unique subgraph of Θ which is a k-path between x and y.*

The 1-trees are nothing but trees. We are concerned here with them and also with 2-trees (see [13,23] for this case). 2-trees have $2n-3$ edges and $n-2$ triangles. The number of triangles containing a given vertex x and the number of triangles containing a given edge xy are, respectively, denoted as $\partial_1(x)$ and $\partial_2(xy)$. A vertex x is a 2-leaf if and only if $\partial_1(x)=1$. The unique 2-path of Θ between x and y (not adjacent) is denoted as $\Theta[xy]$. The 2-tree Θ is a 2-*path* tree if it has exactly two 2-leaves. Among the 2-path-trees, we distinguish the *broken wheels*, with a vertex of degree $n-1$ (see Fig. 15 of Section 4.2.4 for an example) and the T-*paths* where no vertex has a degree exceeding 4 (see Fig. 13 of Section 4.1). For a 2-tree Θ on X and a fixed indexing $(x_1,x_2,\dots,x_n)$ of X, we set $X^i=\{x_1,\dots,x_i\}$ and denote as Θ^i the subgraph of Θ induced by X^i, for $i=2,3,\dots,n$. An *elimination order* (*of* Θ) is a linear order $(x_1,x_2,\dots,x_n)$ on X such that the following two equivalent conditions hold:

- for $i=3,4,\dots,n-1$, the subgraph Θ^i is a 2-tree on X^i;
- for $i=4,\dots,n$, the vertex x^i is a 2-leaf of Θ^i.

For instance, the 2-tree of Fig. 5 admits 216 elimination orders among the 5040 linear orders on seven elements. Then, for $i=3,4,\dots,n$, the unique triangle of Θ^i containing x^i is denoted as A^{i-2}.

2.2. Dissimilarities and tree functions

2.2.1. Types of tree functions

A *dissimilarity* on X is a real function d on $X\times X$ satisfying the following conditions for all $x,y\in X$: (D1) $d(x,y)=d(y,x)$; (D2) $d(x,y)\geqslant d(x,x)=0$. The notation $d(xy)$ will be used instead of $d(x,y)$ whenever the elements x and y are known to be distinct. With this notation, d is just a non-negative real function defined on the set $X^{(2)}$ of unordered pairs of distinct elements of X. Real functions on $X^{(2)}$ without this constraint will also be considered.

A dissimilarity d is a *metric* if it satisfies the classical metric triangular inequality: for all $x,y,z\in X, d(x,z)\leqslant d(x,y)+d(y,z)$. This property is satisfied by the minimum path length function δ of any positively valued connected graph. A dissimilarity d is an *ultrametric* when it satisfies the stronger inequality $d(x,z)\leqslant\max(d(x,y),d(y,z))$. It is a *tree metric* if it is representable by the lengths of the paths between the leaves of a non-negatively valued XLL-tree (called its *tree representation*); a tree metric is a metric. A dissimilarity d on X satisfies the *four-point condition* (F), if, for all $x,y,z,w\in X$,

$$d(x,y)+d(z,w)\leqslant\max\{d(x,z)+d(y,w),d(x,w)+d(y,z)\}. \tag{F}$$

Theorem 2.1 (Buneman [5], Dobson [10], Patrinos [22] and Zaretskii [26]). *For a dissimilarity d on X, the following two conditions are equivalent*:
(i) *d is a tree metric*;
(ii) *d satisfies the four-point condition.*
Moreover, a tree metric admits a unique tree representation.

Here the possibility of edges (adjacent to a leaf) with null length makes it not necessary to consider trees where X is not exactly the set of leaves. The previous result admits several extensions. A first one is straightforward: a dissimilarity d on X satisfies the *weak four-point condition* (W), if, for all *distinct* $x, y, z, w \in X$,

$$d(xy) + d(zw) \leqslant \max\{d(xz) + d(yw), d(xw) + d(yz)\}. \tag{W}$$

As recalled in Leclerc [18], a dissimilarity satisfying Condition (W) is not necessarily a metric, but has still a unique tree representation, possibly with negative lengths on the edges adjacent to the leaves. Such a dissimilarity will be called a *tree dissimilarity* (often abbreviated as TD in the sequel).

A further extension has been recently defined by Bandelt and Steel [2]. A dissimilarity d on X satisfies the *unsigned four-point condition* (U) if, for all $x, y, z, w \in X$, at least two among the three sums in Condition (W) are equal:

$$|\{d(x,y) + d(z,w), d(x,z) + d(y,w), d(x,w) + d(y,z)\}| \leqslant 2. \tag{U}$$

This condition characterizes dissimilarities, called *unsigned tree dissimilarities*, which are representable (uniquely again) by a valued tree with positive or negative lengths.

A real function d on $X^{(2)}$ satisfying Condition (W) (resp. Condition (U)) is called a *tree function*, abbreviated as TF (resp. an *unsigned tree function*). It is straightforward to make a tree dissimilarity correspond to such a function, with no change in the shape of the tree representation: add a convenient positive constant $2K$ to each value of d: it corresponds to the addition of K to the length of the edge $xa(x)$, for any leaf x (see Tables 1 and 2 and Figs. 2 and 3).

2.2.2. *Well-formed triples*

Given three distinct vertices u, v, w of a tree T, there is a unique vertex $m(u, v, w)$ which is common to the paths $T(uv), T(vw)$ and $T(uw)$. This vertex is called the *median* of the triple (u, v, w). If x, y, z are three distinct leaves of T, then $m(x, y, z) \neq x, y, z$. If t is the path length function of a valued tree T_ℓ, the distance between the vertex x

Table 1
A tree dissimilarity

b	1		
c	1	4	
d	4	7	1
	a	*b*	*c*

Table 2
An unsigned tree dissimilarity

b	5		
c	5	4	
d	3	2	6
	a	*b*	*c*

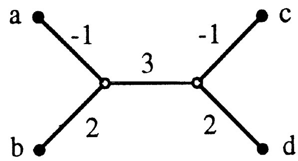

Fig. 2. The tree representation of the dissimilarity of Table 1.

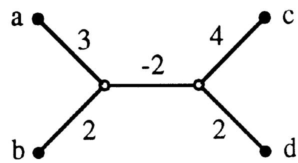

Fig. 3. The tree representation of the dissimilarity of Table 2.

and the path $T(yz)$ is equal to $t(x,m(x,y,z)) = \frac{1}{2}(t(xy)+t(xz)-t(yz))$; this quantity is denoted as $t(x,yz)$.

Consider a TF t and its valued X-tree representation T_ℓ. A triple (x,y,z) of leaves of T is said *well-formed* if the latent vertex $a(y)$ belongs to the path $T(xz)$. For instance, in the tree of Fig. 1, the triple (x,y,z) is well-formed, but not (y,x,z). The expression "well-formed triple" will be frequently abbreviated as WFT in the sequel. Several characterizations of WFTs in relation with the dissimilarity t are given in the next result:

Proposition 2.2. *For a tree function t on X, its tree representation T, and a triple (x,y,z) of leaves of T, the following four conditions are equivalent*:
(i) *the triple (x,y,z) is well-formed*;
(ii) $t(y,xz) = \min_{a,b\in X-\{y\}} t(y,ab)$;
(iii) $t(yz)-t(xz) = \min_{w\in X-\{x,y\}} t(yw)-t(xw)$ *holds*;
(iv) $t(xy)-t(xz) = \min_{w\in X-\{y,z\}} t(yw)-t(zw)$ *holds.*

Proof. (i)$\Rightarrow$(ii): If (x,y,z) is a WFT, then, $t(y,xz) = \ell(ya(y)) = \rho(y)$. Let $a,b\in X-\{y\}$ and consider the path $T(yv)$ from y to the first latent vertex v on the path $T(ab)$. The path $T(yv)$ includes the edge $ya(y)$, and so, has a length at least equal to $\rho(y)$. Then, $t(y,xz) = \min_{a,b\in X-\{y\}} t(y,ab)$.

(ii)$\Rightarrow$(iii): From (ii), we have $t(y,xz) = \min_{w\in X-\{x,y\}} t(y,xw)$ and, so, $t(xy)+t(yz)-t(xz) = \min_{w\in X-\{x,y\}} t(xy)+t(yw)-t(xw)$.

(iii)$\Rightarrow$(i): The quantity $t(yw)-t(xw)$ is minimized by those leaves w such that $a(y)$ belongs to the path $T(xw)$.

The implications (ii)$\Rightarrow$(iv)$\Rightarrow$(i) are obtained by replacing (iii) with (iv) above. $\square$

With these characterizations, the notion of a well-formed triple applies to a TF t as well as to an XLL-tree T (but Proposition 2.2 is no longer valid in the case of an unsigned tree function satisfying Condition (U)). They also imply the following property, which will be (generally implicitly) used in the proofs of Section 3: if a triple (x,y,z) is well-formed for t, then, for any subset Y containing x,y and z, it remains well-formed for the restriction tree and dissimilarity $T_{|Y}$ and $t_{|y}$.

Proposition 2.3. *Assume $n>3$. Let t be a TF on the set X and (x,y,z) a WFT of t. Then, the following equality holds for all $w \in X-\{x,y,z\}$:*

$$t(yw) = \max\{t(xw)+t(yz), t(xy)+t(zw)\} - t(xz).$$

Proof. The inequality $t(xw)+t(yz) \leqslant t(yw)+t(xz)$ comes from the hypothesis and Proposition 2.2; similarly, $t(zw)+t(xy) \leqslant t(yw)+t(xz)$. With Condition (W), these inequalities imply the result. □

Proposition 2.3 has the following consequence: for $n>3$, if the restriction $t_{|X-\{y\}}$ is known, only two values, $t(yz)$ and $t(xy)$ are necessary to determine the whole table of t, provided (x,y,z) is a WFT. So, it may be expected that, by consideration of successive WFTs, a TF is determined by $3+2(n-3)=2n-3$ conveniently chosen values. This is confirmed and precised in the next two sections.

3. From valued 2-trees to tree functions

Consider a positively valued 2-tree Θ_d on X, together with an elimination order $L=(x_1,\dots,x_n)$ of Θ. Let $\mathscr{A}=\{A^1,A^2,\dots,A^{n-2}\}$ be the corresponding sequence of triangles of Θ. Then, we determine a TF $t=t_{\Theta,d,L}$ to Θ_d such that $t(xy)=d(xy)$ for any edge xy of Θ as follows:

- For $i=3$, we have $A^1=x_1x_2x_3$. Any real function d on $(X^3)^{(2)}$ satisfies Condition (U). The unique $\{x_1,x_2,x_3\}$LL-tree representation of d consists of a 3-star T_ℓ with x_1,x_2,x_3 as leaves and a latent vertex $u=a(x_i)$, $i=1,2,3$ (Fig. 4). The length $\ell(x_1u)=\rho(x_1)$ of its edge x_1u is given by $2\rho(x_1)=d(x_1x_2)+d(x_1x_3)-d(x_2x_3)$; the lengths $\rho(x_2)$ and $\rho(x_3)$ are similarly determined.

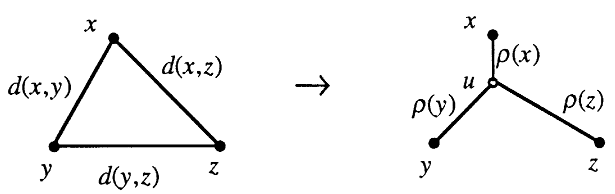

Fig. 4.

- Assume that, for $i-1$ ($3 \leqslant i \leqslant n$), a TF $t_{|X'}$ with its X^{i-1} LL-tree representation T_ℓ^{i-1} has been associated with Θ_d^{i-1}, on such a way that $t(aa') = d(aa')$ for any edge aa' of Θ^{i-1}. We have to add the leaf x_i to T^{i-1} to obtain the tree T^i. Set $x_i = z$ and $A^{i-2} = xyz$.

Set $2\sigma(x) = 2t(x, yz) = d(xy) + d(xz) - d(yz)$. The problem is to place the latent vertex $a(z)$ on the path $T^{i-1}(xy)$ in such a way that $t(xa(z)) = \sigma(x)$; then, we shall set $\ell(a(z)z) = \rho(z)$ in every case. Let $u_1 = a(x), u_2, \ldots, u_k = a(y)$ be the successive latent vertices of the path $T^{i-1}(xy)$. With the condition that t satisfies Condition (W), that is no edge between two latent vertices has a negative length, six cases may occur. The first two concern the situation where $t(xu_1) \leqslant \sigma(x) \leqslant t(xu_k)$:

(i) there are two vertices u_j and u_{j+1} such that $t(xu_j) < \sigma(x) < t(xu_{j+1})$. Set $V(T) = V(T') \cup \{a(z), z\}$ and $E(T^i) = (E(T^{i-1}) - \{u_j u_{j+1}\}) \cup \{a(z)u_j, a(z)u_{j+1}, a(z)z\}$, with the lengths: $\ell(a(z)u_j) = \sigma(x) - t(xu_j), \ell(a(z)u_{j+1}) = t(xu_{j+1}) - \sigma(x)$;

(ii) there exists a vertex u_j such that $\sigma(x) = t(x, u_j)$. Then, the vertex $a(z)$ is identical to u_j. Set $V(T^i) = V(T^{i-1}) \cup \{z\}$ and $E(T^i) = E(T^{i-1}) \cup \{a(z)z\}$;

When $\sigma(x) < t(xu_1)$, two cases must be considered:

(iii) if $\sigma(x) \neq 0$, set $V(T^i) = V(T^{i-1}) \cup \{a(z), z\}$ and $E(T^i) = (E(T^{i-1}) - \{xu_1\}) \cup \{xa(z), a(z)u_1, a(z)z\}$, with the lengths: $\ell(a(z)x) = \sigma(x)$ and $\ell(a(z)u_1) = t(xu_1) - \sigma(x)$;

(iv) if $\sigma(x) = 0$, then the vertex $a(z)$ is the same as $a(x)$ in the tree T^i. Precisely, $V(T^i) = V(T^{i-1}) \cup \{a(z), z\}$ and $E(T^i) = (E(T^{i-1}) - \{u_1 x\}) \cup \{u_1 a(z), a(z)x, a(z)z\}$, with the previous length of $u_1 x$ as length of $u_1 a(z)$ and $\ell(a(z)x) = 0$.

The last two cases concern the case where $\sigma(x) > t(xu_k)$:

(v) if $\sigma(x) \neq d(xy)$, set $V(T^i) = V(T^{i-1}) \cup \{a(z), z\}$ and $E(T^i) = (E(T^{i-1}) - \{u_k y\}) \cup \{ya(z), a(z)u_k, a(z)z\}$, with the lengths: $\ell(a(z)u_k) = \sigma(x) - t(x, u_k)$ and $\ell(a(z)y) = d(x, y) - \sigma(x)$;

(vi) if $\sigma(x) = d(xy)$, then $a(z) = a(y)$ in the tree T^i; set $V(T^i) = V(T^{i-1}) \cup \{a(z), z\}$ and $E(T^i) = (E(T^{i-1}) - \{u_k y\}) \cup \{u_k a(z), a(z)y, a(z)z\}$, with the previous length of $u_k y$ as length of $u_k a(z)$ and $\ell(a(z)y) = 0$.

For $i = n$, a valued XLL-tree T_ℓ^n is obtained, with its associated TF $t = t_{\Theta,d,L}$. Several properties of $t_{\Theta,d,L}$ are derived from this construction. In the proofs of the next results, we set, as above, $x_i = z$ and $A^{i-2} = xyz$. The basic point is that (x, z, y) is a WFT for T^i and, so, Proposition 2.3 applies.

Proposition 3.1. *Let uv be an edge of the tree representation T_ℓ of $t = t_{\Theta,d,L}$; there are at least two edges ab of Θ such that uv belongs to the paths $T(ab)$.*

Proof. We show this result for all the subgraphs Θ^i. It obviously holds for $i = 3$. Assume that it holds for Θ^{i-1}. The XLL-tree T_ℓ^i is obtained by adding the vertex $a(z)$ on the path $T^{i-1}(xy)$ and, generally, replacing an edge $u_p u_{p+1}$ of this path by the edges $u_p a(z)$ and $a(z)u_{p+1}$; these edges belong to the path $T^i(ab)$ for each edge ab of Θ such that $u_p u_{p+1}$ belongs to the path $T^{i-1}(ab)$. The edge $za(z)$ is also added.

It belongs to the paths $T^i(xz)$ and $T^i(yz)$ of Θ^i; by the induction hypothesis, every other edge of Θ^i belongs to the path $T^i(ab)$ for at least two edges ab of Θ^i. □

Theorem 3.2. *Let Θ_d be a positively valued 2-tree on X, and $t = t_{\Theta,d,L}$ the TF associated with an elimination order L of Θ. Let a, b, w and z be four distinct elements of X such that $wz \notin E(\Theta)$ and ab is an edge of the 2-path $\Theta[wz]$. The following equality holds*:

$$t(zw) + t(ab) = \max\{t(aw) + t(bz),\ t(az) + t(bw)\}.$$

Proof. It may be assumed without loss of generality that w is before z in the order L; let Y be the set of the vertices of the 2-path $\Theta[wz]$; by construction, the triple (x, z, y) is a WFT for the TF $t_{|Y}$. We prove the result by induction on the number k of triangles of this 2-path. If $k = 2$ (the minimum possible value with the hypotheses), it is a direct consequence of Proposition 2.3.

Assume $k > 2$. If $xy = ab$, then the result follows again from the fact that (x, z, y) is a WFT together with Proposition 2.3.

If, say, $y = b$, we have, by the induction hypothesis,

$$t(aw) + t(bx) \leqslant t(xw) + t(ab).$$

and, since the triple (x, z, b) is a WFT,

$$t(xw) + t(bz) \leqslant t(zw) + t(bx).$$

Combining these inequalities, we have

$$t(aw) + t(bz) \leqslant t(zw) + t(ab),$$

which, with Condition (W), gives the result.

If the four elements a, b, x, y are distinct, the induction hypothesis leads to
(a) $t(ax) + t(bw) \leqslant t(wx) + t(ab)$, and
(b) $t(ay) + t(bw) \leqslant t(wy) + t(ab)$;
Since the triple (x, z, y) is a WFT, we have also
(c) $t(wx) + t(yz) \leqslant t(wz) + t(xy)$, and
(d) $t(wy) + t(xz) \leqslant t(wz) + t(xy)$, and also
(e) $t(ay) + t(xz) = t(az) + t(xy)$, or
(f) $t(ax) + t(yz) = t(az) + t(xy)$.
From (a) and (c), we obtain
(g) $t(bw) + t(ax) + t(yz) - t(xy) \leqslant t(wz) + t(ab)$,
and, from (b) and (d),
(h) $t(bw) + t(ay) + t(xz) - t(xy) \leqslant t(wz) + t(ab)$.
We just have to report (e) in (h), or (f) in (g), to obtain

$$t(bw) + t(az) \leqslant t(wz) + t(ab),$$

which, with Condition (W), gives the result. □

As a consequence of this result, the TF $t_{\Theta,d,L}$ does not depend in fact on the particular elimination order L of Θ considered. So, it will be denoted by $t_{\Theta,d}$ in the sequel. Indeed, the value of $t(wz)$ depends only on the values of the edges of the 2-path $\Theta[wz]$. Another observation is that the quantity

$$\max\{t(aw)+t(bz), t(az)+t(bw)\} - t(ab)$$

does not depend on the edge ab of the 2-path $\Theta[wz]$, provided this edge is not incident to w or z. This accounts for the constraints that a tree dissimilarity t satisfies. Theorem 3.2 gives first solutions to two problems:

- the computation of $t(xy)$ by induction on $|\Theta[xy]|$ (already possible with Proposition 2.3);
- the verification of the equivalence of two valued 2-trees Θ_d and $\Theta'_{d'}$, in the sense that $t_{\Theta,d} = t'_{\Theta,d}$. This question is studied below in Section 4.

Proposition 3.3. *The TF $t_{\Theta,d}$ is a tree metric if and only if the values of d on the edges of each triangle of Θ satisfy the metric triangular inequality.*

Proof. If $t = t_{\Theta,d}$ is a metric, then, since d and t are equal on the edges of Θ, d satisfies the metric inequality on every triangle of Θ. The converse is obvious for $n=3$. Assume that $t_{|X^{i-1}}$ is a tree metric and that the triangle xyz satisfies the metric inequality. The quantity $\sigma(x)$ defined by $2\sigma(x) = d(xy) + d(xz) - d(yz)$ is then nonnegative, as for the quantities $\sigma(y) = d(xy) - \sigma(x)$ and $\rho(z)$ which are derived from similar expressions. So, the result holds for all integers i and for n. □

The observation that, if all the triangles of the 2-tree satisfy the metric inequality, then the TF $t_{\Theta,d}$ is a metric, has been already done in Leclerc [16] in the special case corresponding to Section 4.2.2 below. Figs. 5 and 6 show, respectively, a positively valued 2-tree and its triangles. The construction of the XLL-tree associated with the corresponding TF is suggested in Figs. 7 and 8: a 3-star is associated with each triangle, and these 3-stars are pairwise glued on their paths with common extremities.

Recall that $\ell(T)$ is the total length of the tree representation T_ℓ of $t = t_{\Theta,d}$ and that $\partial_2(xy)$ is the number of triangles of Θ including the pair xy. Consider the quantity $\lambda(\Theta_d) = \sum_{ab\in E(\Theta)}(2 - \partial_2(ab))d(ab)$. Its interest is due to the following result. Let $\rho_i(x_i)$ be the length of the edge $x_i a(x_i)$ in the tree T^i.

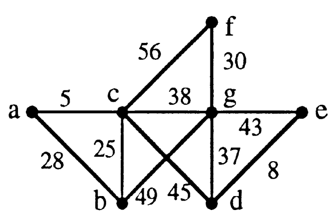

Fig. 5. 2-tree.

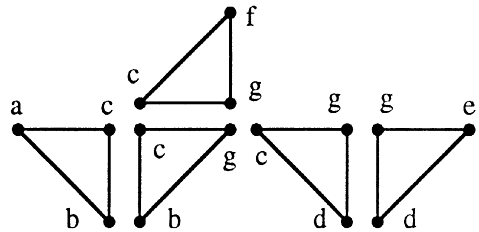

Fig. 6. Triangles.

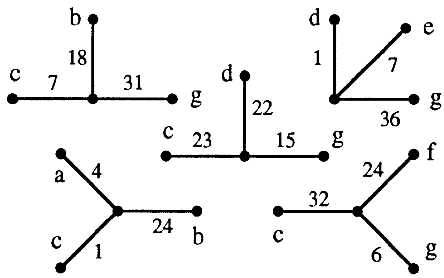

Fig. 7. 3-stars.

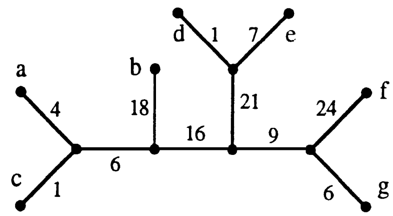

Fig. 8. XLL-tree.

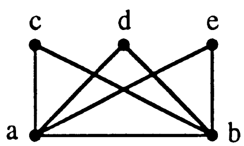

Fig. 9.

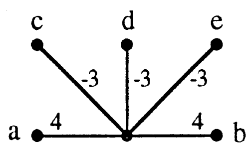

Fig. 10.

Proposition 3.4. *If* T_ℓ *is the tree representation of the TF* $t=t_{\Theta,d}$, *then* $\lambda(\Theta_d)=2\ell(T)$.

Proof. When adding the leaf x_i to T^{i-1}, the difference of lengths between the valued trees T_ℓ^i and T_ℓ^{i-1} is given by $\ell(T^i)-\ell(T^{i-1})=\rho_i(x_i)$; the degree ∂_2 of the new edges xx_i and yx_i is 1, and the number $2-\partial_2(xy)$ decreases by 1. □

So, if $(x_1,x_2,\ldots,x_n)$ is an elimination order of Θ, then, $\lambda(\Theta_t)=2\ell(T)=t(x_1x_2)+t(x_1x_3)+t(x_2x_3)+2\sum_{4\leqslant i\leqslant n}\rho_i(x_i)$. As a consequence, if Θ is a 2-path, then $\partial_2(ab)\leqslant 2$ for any edge ab of Θ, and the total length $\ell(T)$ is positive. This property is not general: the positively valued 2-tree of Fig. 9, valued by $d(xy)=1$ for any pair $xy\neq ab$ and $d(ab)=8$, leads to the valued XLL-tree of Fig. 10, of negative total length.

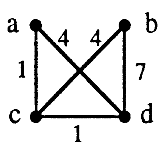

Fig. 11. Θ'.

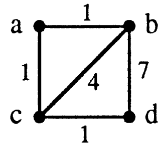

Fig. 12. Θ.

4. The 2-tree abridgements of a tree dissimilarity

4.1. The 2-trees associated with a tree dissimilarity t

In this section, it is assumed that a TD t on X is given, with its unique valued tree representation T_ℓ. A 2-tree Θ on X is denoted as Θ_t when it is valued according to the restriction of t to the edges of Θ. The set of all the 2-trees Θ such that $t = t_{\Theta,t}$ is denoted as $\mathcal{Q}(t)$. There exist generally 2-trees which do not belong to $\mathcal{Q}(t)$. For instance, both 2-trees Θ' of Fig. 11 and Θ of Fig. 12 are valued according to the TD t of Table 1. By Theorem 3.2, $t_{\Theta',t}(ab) = \max\{t(ac)+t(bd), t(ad)+t(bc)\} - t(cd) = 7 \neq t(ab)$, and $t_{\Theta,t}(ad) = \max\{t(ac)+t(bd), t(ab)+t(cd)\} - t(bc) = 4 = t(ad)$. So, Θ belongs to $\mathcal{Q}(t)$ while Θ' does not. A reason is that the 2-leaf b of Θ' belongs to the triangle bcd whereas the triple (c, b, d) is not well-formed. The results of the previous section allow us to state the following:

Theorem 4.1. *Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite set and a valued 2-tree Θ_d on X. Then, the TF $t = t_{\Theta,d}$ is the unique one satisfying*: (i) $\Theta \in \mathcal{Q}(t)$; *and* (ii) $t(xy) = d(xy)$ *for any edge xy of* Θ.

In this section, we complete this observation, firstly with a characterization of the elements of $\mathcal{Q}(t)$ based on Theorem 3.2.

Theorem 4.2. *For a TD t and a 2-tree Θ on X, the following three conditions are equivalent*:

(i) $\Theta \in \mathcal{Q}(t)$.

(ii) *There exists an elimination order $L = (x_1, x_2, \ldots, x_n)$ of Θ such that, for $i = 3, \ldots, n$, the triple (y_i, x_i, z_i) corresponding to the triangle $A^{i-2} = x_i y_i z_i$ of Θ^i is well-formed for $t_{|X^i}$*;

(iii) *For any elimination order $L = (x_1, x_2, \ldots, x_n)$ of Θ, the triple (y_i, x_i, z_i) corresponding to the triangle $A^{i-2} = x_i y_i z_i$ of Θ^i is well-formed for $t_{|X^i}$, for all $i = 3, \ldots, n$.*

Proof. Trivially, (iii) implies (ii), and, if (ii) is satisfied, then the construction of Section 3 may be performed to obtain the TD $t = t_{\Theta,d,L} = t_{\Theta,d}$; so, (ii) implies (i). It remains to show that (i) implies (iii). Let x be a 2-leaf of Θ^i, belonging to the unique triangle xyz of Θ^i, and let $w \in X^i - \{x, y, z\}$; the pair xw is not an edge of

Θ^i and, so, not an edge of Θ. Applying Theorem 3.2, we obtain the equality $t(wx)+t(yz)=\max\{t(wy)+t(xz),t(wz)+t(xy)\}$. Then, $t(wy)+t(xz)\leqslant t(wx)+t(yz)$ and, finally, $t(xz)-t(yz)=\min_{w\in X^i-\{x,y\}} t(xw)-t(yw)$, which, according to Proposition 2.2, proves that the triple (y,x,z) is well-formed for $t_{|X^i}$. □

To prove that, for any TD t, the set $\mathcal{Q}(t)$ is not empty, it suffices to exhibit a 2-tree satisfying Condition (ii) of Theorem 4.2. Such an element Λ of $\mathcal{Q}(t)$ may be constructed in a simple way. Let us determine a linear order $L=(x_1,x_2,\dots,x_n)$ in X as follows:

- Choose arbitrary x_{n-1} and x_n in X.
- For $k=2,\dots,n-1$, find $x_{n-k}\in X^{n-k}=X-\{x_n,\dots,x_{n-k+1}\}$ such that the triple $(x_{n-k},x_{n-k+2},x_{n-k+1})$ is well-formed for the restricted TD $t_{|X^{n-k+2}}$ (and for the reduced tree $T_{|X^{n-k+2}}$). In other terms, $t(x_{n-k}x_{n-k+2})-t(x_{n-k}x_{n-k+1})=\min_{x\in X^{n-k}} t(xx_{n-k+2})-t(xx_{n-k+1})$. Set $\{x_{n-k}x_{n-k+2},x_{n-k+1}x_{n-k+2}\}\subseteq E(\Lambda)$. Finally, set $x_1x_2\in E(\Lambda)$.

The obtained 2-tree Λ is a T-path. Observe that the total length of the tree representation T_ℓ of t, given by $\lambda(\Lambda_t)=t(x_1x_2)+t(x_{n-1}x_n)+\sum_{1\leqslant i\leqslant n-2} t(x_ix_{i+2})$, cannot be negative. Since $\lambda(\Theta_t)=2\ell(T)$ for any element Θ of $\mathcal{Q}(t)$ (by Proposition 3.4), we have:

Proposition 4.3. *If t is a TD on X and T_ℓ its tree representation, then $\ell(T)\geqslant 0$.*

Example. Let us start with $x_7=a$ and $x_6=b$ in the tree metric of Table 3, corresponding with the valued XLL-tree of Fig. 8. We obtain, among several possibilities, the order (g,e,f,d,c,b,a). The corresponding 2-tree Λ is given in Fig. 13.

In fact, several ways are given in the literature for summarizing a tree metric by $2n-3$ entries. They are surveyed in Section 4.2, where it is emphasized that they all consist of the recognition of specific classes of elements of $\mathcal{Q}(t)$. Indeed, this set appears to have many elements. In fact, there exists a 2-tree $\Theta\in\mathcal{Q}(t)$ admitting L as an elimination order for any arbitrary linear order $L=(x_1,x_2,\dots,x_n)$ and TF t on X. To obtain such a 2-tree, a sequence Θ^i of 2-trees on the sets X^i is constructed as follows:

Step 1. Θ^3 is the 3-clique on X^3. Each vertex of Θ^3, especially x^3, has degree 2.

Table 3

b	28					
c	5	25				
d	48	56	45			
e	54	62	51	8		
f	59	67	56	55	61	
g	41	49	38	37	43	30
	a	*b*	*c*	*d*	*e*	*f*

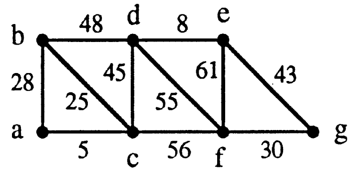

Fig. 13.

Step $i-1$ ($3\leqslant i\leqslant n-1$). Assume that a valued 2-tree Θ^i on X^i, has been obtained in such a way that $(x_1,\dots,x_i)$ is an elimination order of Θ^i. Choose arbitrarily $w\in X^i$ and consider the path $T(wx_{i+1})$ of the tree representation T of t. Let u be the first latent vertex on this path (starting from x_{i+1}) that belongs to a path $T(xy)$ for some edge xy of Θ^i. The 2-tree Θ^{i+1} is obtained by the addition to Θ^i of the vertex x_{i+1} and the edges xx_{i+1} and yx_{i+1}.

Proposition 4.4. *The previous procedure determines at the step* $n-2$ *a valued* 2*-tree* $\Theta=\Theta^n$ *which belongs to* $\mathcal{Q}(t)$ *and admits L as an elimination order.*

Proof. Trivially, the 2-tree Θ^3 is an element of $\mathcal{Q}(t_{|X^3})$. Assume that $\Theta^i\in\mathcal{Q}(t_{|X^i})$ and consider the step i. By Proposition 3.1, the latent vertex u is also the first one that belongs to some path of T between elements of X^i. So, the triple (x,x_{i+1},y) is well-formed for $t_{|X^{i+1}}$. Then, the 2-tree Θ^n satisfies the condition (ii) of Theorem 4.2 and belongs to $\mathcal{Q}(t)$. By construction, for $i=3,\dots,n$, the vertex x_i is a 2-leaf of Θ^i, which means that L is an elimination order of $\Theta=\Theta^n$. □

Thus, since the vertices of Θ^3 constitute a triangle of Θ, the elements of $\mathcal{Q}(t)$ cover all the triangles of X, and, so, all the pairs. This observation presents some analogy with the fact that the minimum spanning trees of an ultrametric on X cover all the pairs of elements of X [15,19].

Corollary 4.5. *Let t be a TD on X. For all distinct* $x,y,z\in X$, *there exists a* 2*-tree* $\Theta\in\mathcal{Q}(t)$ *such that* $\{x,y,z\}$ *is a triangle of* Θ.

A characterization of WFTs as triangles of 2-trees may be obtained as a complement of Proposition 2.2.

Corollary 4.6. *Let t be a TD on X. A triple* (x,z,y) *of elements of X is a WFT of t if and only if there exists a* 2*-tree* $\Theta\in\mathcal{Q}(t)$ *such that the vertex z is a* 2*-leaf and xyz is a triangle of* Θ.

Proof. The ‘if’ part is a direct consequence of the construction of $t_{\Theta,t}$ in Section 3. Conversely, assume that (x,z,y) is a WFT of t and consider a 2-tree $\Theta'\in\mathcal{Q}(t_{|X-\{z\}})$ such that xy is an edge of Θ'. By the hypothesis, the latent vertex $a(z)$ of the tree representation T_ℓ of t belongs to the path $T(xy)$. Then, the 2-tree Θ obtained by the addition of the vertex z and the edges xz and yz to Θ' belongs to $\mathscr{Q}(t)$. □

Finally, we give two characterizations of the elements of $\mathscr{Q}(t)$ by minimality properties. In the next statement, the inequality $t' \leqslant t$ corresponds to the pointwise order on mappings; it means $t(xy) \leqslant t'(xy)$ for all distinct $x, y \in X$. The definition of the pointwise minimum follows. Let $\mathscr{Q}_X$ be the set of all the 2-trees on X.

Theorem 4.7. *For a TD t and a 2-tree Θ on X, the following three conditions are equivalent*:
(i) $\Theta \in \mathscr{Q}(t)$;
(ii) $t = t_{\Theta,t} = \min\{t_{\Theta',t}\colon \Theta' \in \mathscr{Q}_X\}$;
(iii) $\lambda(\Theta_t) = \min\{\lambda(\Theta'_t)\colon \Theta' \in \mathscr{Q}_X\}$.

Proof. Let Θ' be an element of $\mathscr{Q}_X$. Set $t' = t_{\Theta',t}$. If xy is an edge of Θ', then, $t(xy) = t'(xy)$. Otherwise, we prove the inequality $t \leqslant t'$ by induction on the number k of triangles of the 2-path $\Theta'[xy]$. Let aby be the triangle containing y in this path; so, ab is an edge of Θ'. For $k = 2$, ax, ay, bx and by are all edges of Θ'. Then, $t(xy) \leqslant \max\{t(ax)+t(by), t(ay)+t(bx)\} - t(ab) = \max\{t'(ax)+t'(by), t'(ay)+t'(bx)\} - t'(ab) = t'(xy)$, this last equality due to Theorem 3.2.

For $k > 2, ab, ay$ and by are edges of Θ' and, by the induction hypothesis, $t(ax) \leqslant t'(ax)$ and $t(bx) \leqslant t'(bx)$. Then, $t(xy) \leqslant \max\{t(ax) + t(by), t(ay) + t(bx)\} - t(ab) \leqslant \max\{t'(ax) + t(by), t(ay) + t'(bx)\} - t(ab) = t'(xy)$.

The equality $t = t_{\Theta',t}$ is, by definition, obtained when $\Theta' \in \mathscr{Q}(t)$. The equivalence of properties (i) and (ii) follows.

In order to prove the equivalence of (i) and (iii), consider an arbitrary order $L = (x_1, x_2, \ldots, x_n)$ on X, an element Θ' of $\mathscr{Q}_X$ and two 2-trees Π and Π' belonging to $\mathscr{Q}(t)$ and $\mathscr{Q}(t_{\Theta',t})$, respectively. Assume that both Π and Π' admit L as an elimination order; this is possible, according to Corollary 4.5. By Proposition 3.4, $\lambda(\Theta_t) = \lambda(\Pi_t)$ and $\lambda(\Theta'_t) = \lambda(\Pi'_t)$. Let us show that $\lambda(\Pi_t) \leqslant \lambda(\Pi'_t)$, the equality being obtained only if $\Pi' \in \mathscr{Q}(t)$. This is true for $n = 3$ where, in fact, there is a unique 2-tree on X. For $n > 3$, consider, for $4 \leqslant i \leqslant n$, the triangle xyx_i of Π^i and the triangle $x'y'x_i$ of Π'^i. Since the first one corresponds with a WFT of t, the inequality $2\rho(x_i) = t(xx_i) + t(yx_i) - t(xy) \leqslant t(x'x_i) + t(y'x_i) - t(x'y') = 2\rho'(x_i)$ follows from Condition (ii) of Proposition 2.2 and is an equality only if (x', x_i, y') is also a WFT of t. Since $\lambda(\Pi_t) = t(x_1x_2) + t(x_1x_3) + t(x_2x_3) + 2\sum_{4 \leqslant i \leqslant n} \rho(x_i)$ and $\lambda(\Pi'_t) = t(x_1x_2) + t(x_1x_3) + t(x_2x_3) + 2\sum_{4 \leqslant i \leqslant n} \rho'(x_i)$, the result follows. □

4.2. Examples

A first example of elements of $\mathscr{Q}(t)$, the 2-paths Λ, depending, at least, on the initial choices of x_{n-1} and x_n, has been given in Section 4.1. Here we present four other classes of 2-trees which satisfy Condition (ii) of Theorem 4.1 and, so, belong to $\mathscr{Q}(t)$. The tree metric of Table 3 is again used for the illustrations of these classes.

4.2.1. Abridgement by decomposition

Let c be an element of X and consider the real function $\upsilon_{t,c}$ on $(X\backslash\{c\})^2$ defined by $\upsilon_{t,c}(xy)=t(xy)-t(cx)-t(cy)$. A classical fact is that, although it is generally not a dissimilarity (it has negative entries), $\upsilon_{t,c}$ satisfies the ultrametric inequality ([11]; see also [1]). Then, it is defined by the $n-2$ entries of a minimum spanning tree (MST), and the TD t by the $2n-3$ entries choosen as follows [18]:

- choose arbitrarily an element c of X; set $X'=X\backslash\{c\}$;
- compute the function $\upsilon_{t,c}$ on X'^2;
- determine an MST $M'=(X',A')$ on X' for $\upsilon_{t,c}$.

Set $R_c=\{cx\colon x\in X'\}\cup A'$, with $|R_c|=(n-1)+(n-2)=2n-3$. For a pair xy of elements of X', the value of $t(xy)$ is recovered from the entries of t on R_c by

$$t(xy)=t(cx)+t(cy)+\max\{t(zw)-t(cz)-t(cw)\colon zw\in M'(xy)\}.$$

The graph $\Delta=(X,R_c)$ depends on the choice of the element c and of the MST M'. Each edge xy of A' determines, with the edges cx and cy, a triangle of Δ, which is a 2-tree. Any elimination order of Δ is an elimination order of the 1-tree M'. A 2-leaf x of Δ is a leaf of M', adjacent to a unique vertex y. By the properties of MSTs, $\upsilon_{t,c}(xy)=\min_{z\in X\backslash\{c,x\}}\upsilon_{t,c}(xz)$, that is $t(xy)-t(cy)=\min_{z\in X\backslash\{c,x\}}t(xz)-t(cz)$; then, by Proposition 2.2, the triple (c,x,y) is well-formed for t. If x is deleted, the subgraph $M'_{X'-\{x\}}$ is still an MST for $(\upsilon_{t,c})_{|X'-\{x\}}$ and any of its leaves determines again a WFT. Finally, Δ is an element of $\mathscr{Q}(t)$.

Example. In the tree metric of Table 3, choose a as the element c mentioned above; the values of the function $\upsilon_{t,a}$ are given in Table 4. Choose $A'=\{be,cf,de,ef,fg\}$ (among 80 possible MSTs for $\upsilon_{t,a}$). The valued 2-tree Δ_t is given in Fig. 14.

4.2.2. Abridgement by minimum spanning trees

Another way of summarizing a TD by $2n-3$ entries is proposed in [16,18]. It leads to a 2-tree $\Sigma=(X,A\cup A'\}$ obtained as follows: let A be the edge set of an MST M of t; in case of ties on the values of t, this MST must be free of subsets of edges $\{xy,yy',y'z\}$ such that $t(xy)=t(xy'),t(yz)=t(y'z)$ and $t(yy')\leqslant\max(t(xy),t(yz))$;

Table 4

c	−8				
d	−20	−8			
e	−20	−8	−94		
f	−20	−8	−52	−52	
g	−20	−8	−52	−52	−70
	b	c	d	e	f

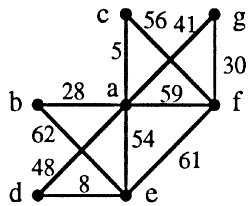

Fig. 14.

the construction of such an MST is straightforward, and it is shown in [18] (see also [8]) that, for any quadruple a,b,c,w of elements of X such that ab and bc are two edges of the path $M(aw)$, the following equality holds:
(i) $t(aw)=t(bw)+t(ac)-t(bc)$

Let $N(x)=\{y\in X\colon\ xy\in A\}$ for each vertex x of degree at least 2 in A. Compute the function $\upsilon_{t,x}$ on $(N(x))^2$ as in Section 4.2.2 and determine an MST A'_x on $N(x)$; take A' as the union of the A'_x. The following facts may be observed: every vertex of Σ has degree at least 2 in Σ; a 2-leaf z of Σ is a leaf of M; there are at least two leaves of M that are 2-leaves of Σ; the deletion of such a 2-leaf z gives a subgraph $\Sigma_{X-\{z\}}$ of the same type. Then, one shows by induction on n that Σ is a 2-tree on X. It remains to show that, if x is the unique neighbor of z in M, and y its other neighbor in Σ, then, the triple (y,x,z) is well formed. In this case, xy is an edge of M. Let $w\in X-\{x,y,z\}$, and xw_1 the first edge of the path $M(xw)$. Three cases may occur:

- $w_1=w$; the vertices w_1, y and z are all adjacent to x in M and one has locally the situation of Section 4.2.1: since z is a leaf and yz an edge of the MST A'_x, one has $t(yz)-t(xz)-t(xy)\leqslant t(zw)-t(xz)-t(xw)$, and $t(yz)-t(xy)\leqslant t(zw)-t(xw)$.
- $w_1=y\neq w$; from the equality (i) above, one has $t(zw)+t(xy)=t(yz)+t(xw)$, that is $t(yz)-t(xy)=t(zw)-t(xw)$.
- If w_1 differs from both y and w, one may combine the above cases to obtain the following inequality and equality: $t(yz)-t(xy)\leqslant t(zw_1)-t(xw_1)=t(zw)-t(xw)$.

Finally, $t(yz)-t(xy)=\min_{w\in X-\{x,y,z\}}(t(zw)-t(xw))$, which, by Proposition 2.2, implies that the triple (x,z,y) is well-formed.

The tree metric t of Table 3 has no ties and, so, a unique MST with $A=\{ac,bc,cg,dg,de,fg\}$ as edge set; with $A'_c=\{ab,bg\}, A'_d=\{eg\}$ and $A'_g=\{cd,cf\}$, the obtained valued 2-tree Σ_t is given in Fig. 5 above.

4.2.3. *A Dijkstra-type construction*

The following procedure leads to a 2-tree Ξ belonging to $\mathcal{Q}(t)$. As the previous one, it selects a set of edges with low values, but in a different way. The idea is to add at each step a vertex z minimizing the length of the new edge $za(z)$. It is close, in some way, to the Prim–Dijkstra [9,24] method for the construction of an MST.

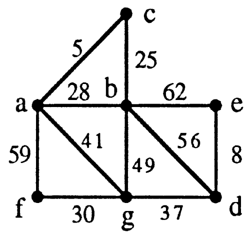

Fig. 15.

- Choose a pair x_1,x_2 of elements of X such that $t(x_1x_2)=\min_{x,y\in X,x\neq y} t(xy)$; set $X^2=\{x_1,x_2\}$ and $\{x_1x_2\}=E(\Xi^2)$;
- For $k=3,\dots,n$, select $x_k\in X-X^{k-1}$ and $xy\in E(\Sigma^{k-1})$ such that $t(x_kx)+t(x_ky)-t(xy)=\min_{z\in X-X^{k-1},x'y'\in E(\Sigma^{k-1})} t(zx')+t(zy')-t(x'y')$; the triple (x,x_k,y) is well-formed for $t_{|X^k}$. Set $X^k=X^{k-1}\cup\{x_k\}$ and $E(\Xi^k)=E(\Xi^{k-1})\cup\{xx_k,yx_k\}$.

The valued 2-tree Ξ_t obtained for the dissimilarity of Table 3 is given in Fig. 15.

4.2.4. Circular orders

A linear order $(x_1,x_2,\dots,x_n)$ on X is *diagonal* if the triple (x_{i-1},x_i,x_{i+1}) is well-formed for all integers $i=1,\dots,n$ (modulo n). It is shown in Chaiken et al. [6] that any order corresponding to a circular (say, clockwise) scanning of the leaves of T in a graphical planar representation of this tree is diagonal. Such orders are called *diagonal plane* by the previous authors, and *circular* in Makarenkov and Leclerc [20], where a counter-example of a diagonal order which is not circular is given.

Given a circular order C on X, the entire table of t may be recovered from its values on the set $R_C=\{x_1x_i\colon\ i=2,\dots n\}\cup\{x_{i-1}x_i\colon\ i=3,\dots,n\}$ [3], p. 60–62, 1991, p. 58–60, generalizing [6]. This set is the edge set of a 2-tree Ω on X of the broken wheel type, with x_2 and x_n as 2-leaves belonging, respectively, to the triangles $x_1x_2x_3$ and $x_1x_{n-1}x_n$. The triples (x_1,x_2,x_3) and (x_1,x_n,x_{n-1}) are well-formed. When deleting the leaf x_n in the tree representation T of t, the order $C'=x_1x_2\dots x_{n-1}$ is still circular for $t'=t_{|X-\{x_n\}}$ so, the vertex x_{n-1} is a 2-leaf of the 2-tree $\Omega_{X^{n-1}}$, belongs to the triangle $x_1x_{n-2}x_{n-1}$ and to the WFT (x_1,x_{n-1},x_{n-2}) of t', and so on. That Ω is an element of $\mathcal{Q}(t)$ follows from this reduction procedure. As a consequence of Proposition 3.4, the total length of the tree representation T_ℓ is given by $2\ell(T_\ell)=\lambda(\Omega_t)=t(x_1x_n)+\sum_{1\leqslant i\leqslant n-1} t(x_ix_{i+1})$.

Every $2n-3$-tuple $(d_{12},d_{13},d_{23},\dots,d_{1i},d_{i,i+1},\dots,d_{1n},d_{n-1,n})$ of positive numbers can be considered as the values of the edges of a 2-tree Ω on $X=\{x_1,\dots,x_n\}$ of the previous type, thus leading to a TF $t=t_{\Omega,d}$. Contrary to a claim encountered in the literature, t is not the only TF satisfying the properties $t(x_1x_i)=d_{1i}$ and $t(x_ix_{i+1})=d_{i,i+1}$, since there may exist a TF t' on X such that $t(xy)=t'(xy)$ for any edge xy of Ω, but with $\Omega\notin\mathcal{Q}(t')$. By Theorem 4.1, t is unique when adding that it admits C as a circular order.

In a 1984 paper, Yushmanov considers the orders $(x_1,x_2,\dots,x_n)$ on X such that, for $k=n,\ n-1,\dots,4$, the triple (x_1,x_k,x_{k-1}) is well formed in the tree T^k. He gives

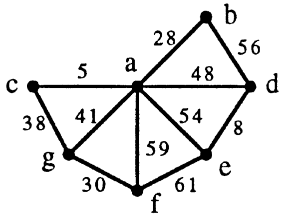

Fig. 16.

a first version of Proposition 2.2 and uses it for the determination of such an order directly from the table of a tree metric t. It is shown in Makarenkov and Leclerc [20] that Yushmanov and circular orders are in fact the same and an algorithm for finding a circular order with arbitrary x_1 and x_n is derived from Yushmanov's work.

In the tree of Fig. 8, the order (a,c,g,f,e,d,b) is circular. The corresponding valued 2-tree Ω_t is given in Fig. 16. It is worth noticing that the order (g,e,f,d,c,b,a) in the example of Section 4.1 (Fig. 13) is not circular. So, we have in fact two different ways of encoding a valued tree by a $(2n-3)$-tuple of distances between leaves. Although it has not been already mentioned in the literature, the method of this section seems more direct. On the other hand, circular orders are interesting from a geometrical point of view.

5. Fitting methods related with 2-trees

5.1. Valued 2-trees and dissimilarities

Abridgements of the types described in the previous section may be used for the storage of a tree metric in a reduced place (especially, types Λ and Ω correspond to encodings of trees by sequences of numbers). They are also a tool for questions relevant to data analysis:

- the adjustment, or fitting, problem: given a dissimilarity d on X, find a TD, or a tree metric, close (on some way) to d.
- the recognition problem: is a given dissimilarity d a TD? A way of obtaining an answer is to compute the tree metric $t=t_{\Theta,d}$ and to check the equalities $t(xy)=d(xy)$ for all pairs x,y of distinct elements.

For the use of 2-trees in these problems, we have first to extend some notions from tree dissimilarities to dissimilarities:

Definitions. Let d be a dissimilarity on X. A triple (x,y,z) of distinct elements of X is *well-formed* if, for any $w\in X-\{x,y,z\}$, the inequality $d(yz)-d(xz)\leqslant d(yw)-d(xw)$ holds.

A 2-tree Θ on X belongs to the set $\mathcal{Q}_1(d)$ if, for any elimination order $L=(x_1,x_2,\dots,x_n)$ of Θ, both triples (y_i,x_i,z_i) and (z_i,x_i,y_i) corresponding to the triangle $A^{i-2}=x_i y_i z_i$ of Θ^i are well-formed for $d_{|X^i}$, for all $i=3,\dots,n$.

A 2-tree Θ on X belongs to the set $\mathcal{Q}(d)$ if it admits an elimination order $L=(x_1,x_2,\dots,x_n)$ of Θ such that, for $i=3,\dots,n$, at least one of the triples (y_i,x_i,z_i) and (z_i,x_i,y_i) corresponding to the triangle $A^{i-2}=x_i y_i z_i$ of Θ^i is well-formed for $d_{|X^i}$.

Of course, $\mathcal{Q}_1(d)\subseteq\mathcal{Q}(d)$. Several remarks may be done about these definitions. Firstly, when d is not a TD, a triple (z,y,x) may not be well-formed whereas (x,y,z) is. Secondly, the equivalence of Conditions (ii) and (iii) of Theorem 4.2 is no longer valid; in other terms, the sets $\mathcal{Q}_1(d)$ and $\mathcal{Q}(d)$ are generally not identical. With these definitions, several classes of elements of $\mathcal{Q}(t)$ defined in Section 4 for a TD t extend to elements of $\mathcal{Q}_1(d)$ or $\mathcal{Q}(d)$:

- The MST-based 2-tree Δ of Section 4.2.1 may be constructed for any dissimilarity d, depending on the choice of an element c (among the n elements of X). It is straightforward to establish that $\Delta\in\mathcal{Q}_1(d)$.
- The construction of a linear order $L=(x_1,x_2,\dots,x_n)$ on X such that, for $k=2,\dots,n-1$, the triple $(x_{n-k},x_{n-k+2},x_{n-k+1})$ is well-formed for $d_{|X^{n-k+2}}$, leads to a T-path Λ. It mainly depends on the initial choice of x_n and x_{n-1} among $n(n-1)$ possible ones.
- Circular orders (called *Yushmanov orders* in Makarenkov and Leclerc [20], since no circular scanning remains) extend to dissimilarities. A difference is that, once the elements x_1 and x_n are choosen (among $n(n-1)$ possible choices), the obtained order C on X is unique in the general case. Such an order C allows to define a 2-tree Ω as in Section 4.2.4.

Moreover, several other classes of 2-trees are obtained in such a way that they belong to $\mathcal{Q}(t)$ for any TD t. They may be used in the problems addressed in Section 5.2 below.

- The MST-based 2-trees Σ of Section 4.2.2 may also be constructed for any dissimilarity d. They do not belong to $\mathcal{Q}_1(d)$ when d is not a TD. The 2-tree Σ is unique when no ties occur on the values of d or in the calculations.
- The Dijkstra-like construction in Section 4.2.3 leads to a 2-tree Ξ which, in the general case, is unique with the choice of the initial pair x_1x_2 minimizing the value of d.

All the constructions above may be obtained by an $O(n^2)$ algorithm. Of course, other 2-trees than those described here may be also considered. For instance, a 2-tree minimizing $\lambda(\Theta_d)$ in $\mathcal{Q}_X$ is an element of $\mathcal{Q}(t)$ for any TD t. At a first glance, to obtain such a 2-tree seems to be a difficult problem.

5.2. Fitting methods

The TF $t_{\Theta,d}$ corresponds to a 2-tree Θ_d, valued according to the restriction of d to the edges of Θ. Choosing a 2-tree of one type described in Section 5.1 as Θ ensures two interesting properties (the second by Proposition 3.3):

Table 5

b	3				
c	1	2			
d	2	5	2		
e	6	7	4	4	
f	7	5	4	6	7
	a	*b*	*c*	*d*	*e*

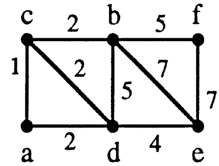

Fig. 17. Λ_d.

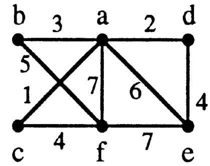

Fig. 18. Δ_d.

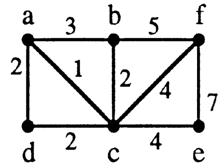

Fig. 19. Σ_d.

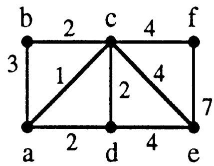

Fig. 20. Ξ_d.

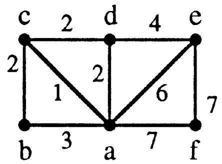

Fig. 21. Ω_d.

- if d is a TD, then $t_{\Theta,d} = d$;
- if d is a metric, then $t_{\Theta,d}$ is a tree metric.

Moreover, it is shown in Leclerc [18] that $t_{\Sigma,d}$ is always a TD. Otherwise, when d is not a metric, the 2-trees of the Λ, Δ, Ξ and Ω types may lead only to TFs; since the 2-trees of the Λ and Ω types are 2-paths, the tree representations of the TFs $t_{\Lambda,d}$ or $t_{\Omega,d}$ have positive total lengths.

Example. For the dissimilarity d (which is not a metric) of Table 5, Figs. 17–21 give the 2-trees Λ_d (with $x_5 = e$ and $x_6 = f$), Λ_d (with $c = a$), Σ_d (unique), Ξ_d (one of two solutions) and Ω_d (with $x_1 = a$ and $x_6 = f$); Figs. 22–26 give the corresponding valued XLL-trees and Tables 6–10 the TFs. In this example, $t_{\Sigma,d}$, $t_{\Xi,d}$ and $t_{\Omega,d}$ are tree metrics; $t_{\Lambda,d}$ is a tree dissimilarity and $t_{\Delta,d}$ is just a tree function.

A mapping $d \mapsto t_{\Theta,d}$ defines a fitting method that makes a tree function (a tree metric) correspond to any dissimilarity (metric), provided Θ is chosen as above. Clearly, such a purely combinatorial approach gives a straightforward solution to the recognition problem, but does not provide a good approximation technique with a metric criterion like the least-squares one. Here, we propose a fitting algorithm based on the least-squares criterion. Its output is a tree metric.

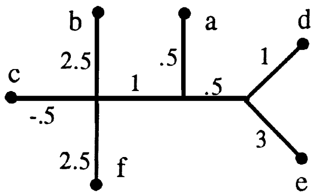

Fig. 22.

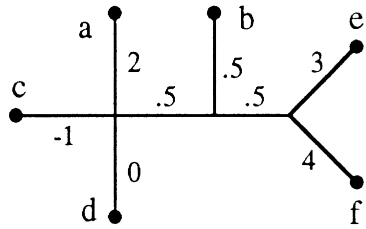

Fig. 23.

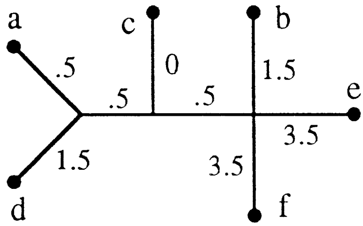

Fig. 24.

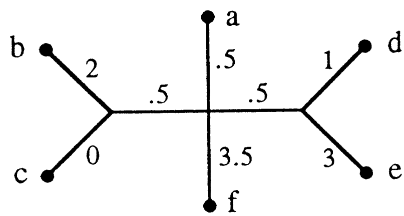

Fig. 25.

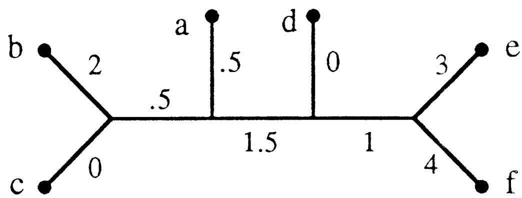

Fig. 26.

Table 6

$t_{\Lambda,d}$

b	4				
c	1	2			
d	2	5	2		
e	4	7	4	4	
f	4	5	2	5	7
	a	*b*	*c*	*d*	*e*

Table 7

$t_{\Delta,d}$

b	3				
c	1	0			
d	2	1	−1		
e	6	4	3	4	
f	7	5	4	5	7
	a	*b*	*c*	*d*	*e*

Algorithm 1. *Construction of a positively valued XLL-tree T_ℓ from a dissimilarity d, a 2-tree $\Theta \in \mathcal{Q}(d)$ and an elimination order $(x_1, x_2, \ldots, x_n)$ of Θ.*

Step 1. $V(T^2) := \{x_1, x_2\}$; $E(T^2) := x_1x_2$; $\ell(x_1x_2) := d(x_1, x_2)$

Step k $(2 \leqslant k \leqslant n-1)$. Let $xyx_{k+1} = A^{k-1}$ be the triangle of Θ^{k+1} containing x_{k+1}. The problem is to add the leaf x_{k+1} to the current valued tree T_ℓ^k with leaves $x_1, \ldots, x_k$, that is to find the place of the latent vertex $a_{k+1} = a(x_{k+1})$ on the path $T^k(xy)$.

Table 8
$t_{\Sigma,d}$

b	3				
c	1	2			
d	2	4	2		
e	5	5	4	6	
f	5	5	4	6	7
	a	b	c	d	e

Table 9
$t_{\Xi,d}$

b	3				
c	1	2			
d	2	4	2		
e	4	6	4	4	
f	4	6	4	5	7
	a	b	c	d	e

Table 10
$t_{\Omega,d}$

b	3				
c	1	2			
d	2	4	2		
e	6	8	6	4	
f	7	9	7	5	7
	a	b	c	d	e

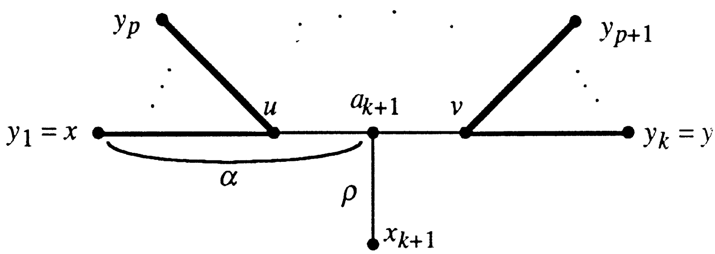

Fig. 27.

Let uv be an edge of this path, u being its extremity closest to x. Let $x=y_1,\dots,y_p$ be the leaves of T_ℓ^k which are on the same side of a_{k+1} as u, and $y_{p+1},\dots,y_k=y$ those on the side of v (Fig. 27). Taking into account the distances $t(y_i,xy)$ for all y_i, $2\leqslant i\leqslant k-1$, we obtain the following quantity to minimize for the best place for a_{k+1} on the edge uv, determined by $\alpha=t(xa_{k+1})$; this quantity is computed together with $\rho=\rho(x_{k+1})=\ell(a_{k+1}x_{k+1})$:

$$\begin{aligned}\text{minimize}\quad & (\alpha+\rho-d(y_1x_{k+1}))^2+(t(y_1y_k)-\alpha+\rho-d(y_kx_{k+1}))^2\\ & +\textstyle\sum_{2\leqslant i\leqslant p}(d(y_ix_{k+1})-(\alpha-t(y_1,y_iy_k)+t(y_i,y_1y_k)+\rho))^2\\ & +\textstyle\sum_{p+1\leqslant i\leqslant k-1}(d(y_ix_{k+1})-(t(y_1y_k)-\alpha\\ & -t(y_k,y_1y_i)+t(y_i,y_1y_k)+\rho))^2\\ \text{subject to}\quad & t(y_1u)\leqslant\alpha\leqslant t(y_1v);\quad \rho\geqslant 0,\end{aligned}$$

where $t(y_1u)$ and $t(y_1,y_iy_k)$, for instance, are the distances between vertices, or between vertex and path, in the tree T_ℓ^k.

This problem is solved with the use of Lagrange multipliers for each edge of the path $T^k_\ell(xy)$. Then, the edge realizing the minimum of the above function is chosen in relation with computed values of α and ρ; several cases may arise:

- If $\alpha=0$, set $V(T^{k+1})=V(T^k)\cup\{a_{k+1},x_{k+1}\}$. Replace the edge xu of T^k with three edges $a_{k+1}u$, xa_{k+1} and $a_{k+1}x_{k+1}$, with lengths $\ell(a_{k+1}u)=\ell(xu)$, $\ell(xa_{k+1})=0$ and $\ell(x_{k+1}a_{k+1})=\rho$.
- If $\alpha=t(xy)$, set $V(T^{k+1})=V(T^k)\cup\{a_{k+1},x_{k+1}\}$. Replace the edge yv of T^k with three edges $a_{k+1}v, ya_{k+1}$ and $a_{k+1}x_{k+1}$, with lengths $\ell(a_{k+1}v)=\ell(yv)$, $\ell(ya_{k+1})=0$ and $\ell(x_{k+1}a_{k+1})=\rho$.
- If $\alpha=t(xu)$ for some latent vertex u of the path $T^k(xy)$, set $V(T^{k+1})=V(T^k)\cup\{x_{k+1}\}$. Add the edge ux_{k+1} with length $\ell(x_{k+1}u)=\rho$ to T^k.
- Otherwise, there are two vertices u and v on the path $T^k(xy)$ such that $t(xu)<\alpha<t(xv)$; set $V(T^{k+1})=V(T^k)\cup\{a_{k+1},x_{k+1}\}$. Replace the edge uv of T^k with three edges $ua_{k+1}, a_{k+1}v$ and $a_{k+1}x_{k+1}$, with lengths $\ell(ua_{k+1})=\alpha-t(xu)$, $\ell(a_{k+1}v)=t(xv)-\alpha$ and $\ell(x_{k+1}a_{k+1})=\rho$.

When the given dissimilarity d is already a tree metric, the minimum sum of squares obtained at each step turns to be null and the final output is d; such an idempotence property is a fundamental requirement for a fitting method. A detailed presentation of this algorithm is made in Makarenkov and Leclerc [20] in the particular case where Θ is the 2-tree corresponding with a Yushmanov order C on X. The fact that such an order is circular for the obtained tree representation T is used to reduce the time complexity to $O(n^2)$. For the generalized version above, this complexity is at worse $O(n^3)$. Although Yushmanov orders already give good results, the possibility of using 2-trees of other types may lead to interesting developments.

6. Conclusion

Our initial purpose, to extend the properties of the MSTs of ultrametrics, is achieved in several ways by the 2-trees abridgements of tree metrics: the set $\mathcal{Q}(t)$ is a class of 2-trees, which constitute a classical extension of trees; the minimization of $\lambda(\Theta_t)$ by its elements corresponds with the minimum length property of any MST; the elements of $\mathcal{Q}(t)$ cover every pair and every triangle of elements of X; for all $\Theta\in\mathcal{Q}(t)$ and $xy\notin E(\Theta)$, $t(xy)$ is a function of the values of t on the edges of the 2-path of Θ between x and y.

Not surprisingly, there are important limitations to these analogies, and the properties observed here are more complicated than in the case of ultrametrics. An important point is that ultrametrics on X correspond with the residuated maps from the partition lattice Π_X to the set $\mathcal{R}^+$ of the non-negative real numbers (see for instance [17]) and, with the pointwise order, have an interesting lattice structure. This fact does not seen to have counterparts for tree dissimilarities (the only property of the pointwise order on these dissimilarities obtained here is Condition (ii) of Theorem 4.7); in the same direction, it may be verified that, contrary to trees, 2-trees are not the bases of a matroid on X.

As a final remark, it may be observed that the definition of the set $\mathcal{Q}(d)$, when d is a dissimilarity with no specified properties, is not satisfactory, since it excludes several classes of 2-trees that turn out to be interesting for the fitting problem. The obtainment of a good definition, and, if possible, the recognition of canonical 2-trees associated with any dissimilarity remains to be studied.

Acknowledgements

The stay in France of Vladimir Makarenkov was supported by the French Ministry of Foreign Affairs. The authors are indebted to Victor Chepoi for a fruitful remark at the conference "Discrete Metric Spaces" in Lyon. We also thank Alain Guénoche and an anonymous referee for advice and comments. This research was partially supported by Esprit LTR Project no. 20244-ALCOM-IT.

References

[1] H.J. Bandelt, Recognition of tree metrics, SIAM J. Discrete Math. 3 (1) (1990) 1–6.
[2] H.J. Bandelt, M. Steel, Symmetric matrices representable by weighted trees over a cancellative abelian monoid, SIAM J. Discrete Math. 8 (4) (1995) 517–525.
[3] J.P. Barthélemy, A. Guénoche, Les arbres et les représentations des proximités, Masson, 1988 (transl.) Paris. Trees and proximity representations, Wiley, New York, 1991.
[4] L.W. Beineke, R.E. Pippert, The number of labelled k-dimensional trees, J. Combin. Theory 6 (2) (1969) 200–205.
[5] P. Buneman, The recovery of trees from measures of dissimilarity, in: F.R. Hodson, D.G. Kendall, P. Tautu (Eds.), Mathematics in Archaeological and Historical Sciences, Edinburgh University Press, Edinburgh, 1971, pp. 387–395.
[6] S. Chaiken, A.K. Dewdney, P.J. Slater, An optimal diagonal tree-code, SIAM J. Algebraic Discrete Meth. 4 (1) (1983) 42–49.
[7] H. Colonius, H.H. Schulze, Tree structure for proximity data, Br. J. Math. Statist. Psychol. 34 (1981) 167–180.
[8] F. Critchley, On exchangeability-based equivalence relations induced by strongly Robinson and, in particular, by quadripolar Robinson dissimilarity matrices, in: B. Van Cutsem (Ed.), Classification and Dissimilarity Analysis, Lecture notes in Statistics, vol. 93, Springer, New York, 1994, pp. 173–199.
[9] E.W. Dijkstra, A note on two problems in connection with graphs, Numer. Math. 1 (1959) 269–271.
[10] A.J. Dobson, Unrooted trees for numerical taxonomy, J. Appl. Probab. 11 (1974) 32–42.
[11] J.S. Farris, A.G. Kluge, M.J. Eckart, A numerical approach to phylogenetic systematics, Syst. Zool. 19 (1970) 172–189.
[12] M.C. Golumbic, Algorithmic Graph Theory and Perfect Graphs, Academic Press, New York, 1980.
[13] F. Harary, E.M. Palmer, On acyclical simplicial complexes, Math. 15 (1968) 115–122.
[14] L. Hubert, Spanning trees and aspects of clustering, Br. J. Math. Statist. Psychol. 27 (1974) 14–28.
[15] B. Leclerc, Description combinatoire des ultramétriques, Math. Sci. Hum. 73 (1981) 5–37.
[16] B. Leclerc, Adjustments of tree metrics based on minimum spanning trees, in: E. Diday et al. (Eds.), New Approaches in Classification and Data Analysis, Springer, Berlin, 1994a, pp. 128–135.
[17] B. Leclerc, The residuation model for the ordinal construction of dissimilarities and other valued objects, in: B. Van Cutsem (Ed.), Classification and Dissimilarity Analysis, Lecture notes in Statistics, vol. 93, Springer, New York, 1994b, pp. 149–172.
[18] B. Leclerc, Minimum spanning trees for tree metrics: abridgements and adjustments, J. Classification 12 (1995) 207–241.
[19] B. Leclerc, Minimum spanning trees and types of dissimilarities, Eur. J. Combin. 17 (2/3) (1996) 255–264.

[20] V. Makarenkov, B. Leclerc, Tree metrics and their circular orders: some uses for the reconstruction and fitting of phylogenetic trees, in: B. Mirkin et al. (Eds.), Mathematical Hierarchies and Biology, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, American Mathematical Society, Providence, RI, 1997, pp. 183–208.

[21] J.W. Moon, The number of labelled k-trees, J. Combin. Theory 6 (2) (1969) 196–199.

[22] A.N. Patrinos, S.L. Hakimi, The distance matrix of a graph and its tree realization, Quart. Appl. Math. 30 (1972) 255–269.

[23] R.E. Pippert, L.W. Beineke, Characterizations of 2-dimensional trees, in: G. Chartrand, S.F. Kapoor (Eds.), The Many Facets of Graph Theory, Lecture Notes in Mathematics, vol. 110, Springer, Berlin, 1969, pp. 263–270.

[24] R.C. Prim, Shortest connexion network and some generalizations, Bell System Technol. J. 36 (1957) 1389–1401.

[25] S.V. Yushmanov, Construction of a tree with p leaves from $2p-3$ elements of its distance matrix (Russian), Mat. Zametki 35 (1984) 877–887.

[26] K. Zaretskii, Construction of a tree on the basis of a set of distances between its leaves (Russian), Uspekhi Mat. Nauk. 20 (1965) 90–92.