The asymptotic enhanced negative type of finite ultrametric spaces

Ian Doust, Stephen Sánchez, and Anthony Weston

Abstract. The $p$-negative type gap $\Gamma_X(p)$ of a finite ultrametric space $(X, d)$ is positive for all $p \geq 0$. Let $\alpha$ denote the minimum non-zero distance in $(X, d)$. In this paper we compute the quantity

$$\Gamma_X(\infty) := \lim_{p \to \infty} \frac{\Gamma_X(p)}{\alpha^p}.$$ 

It turns out that this limit is positive and it may be computed explicitly by an elegant combinatorial formula. This leads to new, asymptotically sharp, families of enhanced $p$-negative type inequalities for $(X, d)$. Indeed, suppose that $G \in (0, \Gamma_X(\infty))$. Then, for all sufficiently large $p$, we have

$$\frac{G^2}{2} \left( \sum_{k=1}^n |\eta_k| \right)^2 + \sum_{j,i=1}^n d(z_j, z_i)^p \eta_j \eta_i \leq 0$$

for each finite subset $\{z_1, \ldots, z_n\} \subseteq X$ and each choice of real numbers $\eta_1, \ldots, \eta_n$ with $\eta_1 + \cdots + \eta_n = 0$.

1. Introduction and statement of the main result

Since the early 1990s there has been renascent interest in embedding properties of negative type metrics. One compelling reason for this is a fundamental link to the design of algorithms for cut problems [1, 7, 2]. The subject of this paper is the important and closely related class of strict negative type metrics. The first systematic treatment of strict negative type metrics appears in the elegant papers of Hjorth et al. [13, 14]. Informally, a metric space is of strict negative type when all of the non-trivial negative type inequalities for the space are strict. In more precise terms we have the following definition.

Definition 1.1. Let $p \geq 0$ and let $(X, d)$ be a metric space.

1. $(X, d)$ has $p$-negative type if and only if for all finite subsets $\{z_1, \ldots, z_n\} \subseteq X$ and all scalar $n$-tuples $\eta = (\eta_1, \ldots, \eta_n) \in \mathbb{R}^n$ that satisfy $\eta_1 + \cdots + \eta_n = 0$, we have:

$$\sum_{j,i=1}^n d(z_j, z_i)^p \eta_j \eta_i \leq 0.$$  \hfill (1.1)

2. $(X, d)$ has strict $p$-negative type if and only if it has $p$-negative type and the inequalities (1.1) are strict except in the trivial case $\eta = (0, 0, \ldots, 0)$.

The first hints of the negative type conditions may be traced back to an 1841 paper of Cayley [6]. Notably, a metric $d$ on a finite set $X$ is of 1-negative type if and only if $(X, \sqrt{d})$ may be isometrically embedded into some Euclidean space. This is a well-known consequence of Theorem 1 in Schoenberg [21]. Faver et al. [12] modified Schoenberg’s proof to show that a metric $d$ on a finite set $X$ is of strict 1-negative type if and only if $(X, \sqrt{d})$ may be isometrically embedded into some Euclidean space as an affinely independent set. In fact, for any $p \in [0, 2]$, a metric $d$ on a finite set $X$ is of strict $p$-negative type if and only if $(X, \sqrt{d^p})$ may be isometrically embedded into some Euclidean space as an affinely independent set ([12 Corollary 3.4]). As every finite metric space has strict $p$-negative type for some $p \in (0, 2)$, it follows that every finite simple connected graph endowed with the ordinary graph metric may be realized as an affinely independent set in some Euclidean space ([12 Corollary 3.8]). We recall that graph realization only requires the length of each
edge to be preserved by the embedding. Other distances within the graph may be distorted. In more general settings one may simply wish to embed a finite metric space of (strict) $p$-negative type into a normed space such as $L_1$ or $L_2$ with minimal distortion. Being able to do so in the case $p = 1$ has significant implications for the design of approximation algorithms \[11\] \[7\] \[2\]. Lately, ultrametric spaces (or, more specifically, $k$-hierarchically well-separated trees) have figured prominently in work on embeddings of finite metric spaces. Interesting papers along these lines include Bartal \[3\] \[4\]. In this paper we examine strict $p$-negative type properties of finite ultrametric spaces in the limit as $p \to \infty$. Importantly, a metric space ultrametric if and only if it has strict $p$-negative type for all $p \geq 0$ (\[12\] Corollary 5.3]). Doust and Weston \[9\] introduced a way to quantify the degree of strictness of the inequalities \[11\]. This notion underpins our work.

**Definition 1.2.** Let $(X, d)$ be a metric space that has $p$-negative type for some $p \geq 0$. Then the $p$-negative type gap of $(X, d)$ is defined to be the largest non-negative constant $\Gamma = \Gamma_X(p)$ that satisfies

$$\frac{\Gamma}{2} \left( \sum_{k=1}^{n} |\eta_k| \right)^2 + \sum_{j,i=1}^{n} d(z_j, z_i)^p \eta_j \eta_i \leq 0 \quad (1.2)$$

for all finite subsets \{ $z_1, \ldots, z_n$ \} $\subseteq X$ and all choices of real numbers $\eta_1, \ldots, \eta_n$ with $\eta_1 + \cdots + \eta_n = 0$.

In practice, it is a non-trivial exercise in combinatorial optimization to determine the precise value of $\Gamma_X(p)$. Most known results deal with what are seemingly the two most tractable cases: $p = 0$ or $p = 1$. A formula for the 1-negative type gap of a finite metric tree was derived in \[9\] using the method of Lagrange multipliers. The same technique was used by Weston \[23\] to compute $\Gamma_X(0)$ for each finite metric space $(X, d)$. Notably, the formula given for $\Gamma_X(0)$ in \[23\] only depends upon $|X|$. For a finite metric space $(X, d)$ of strict $p$-negative type, Wolf \[24\] has derived some general matrix-based formulas for computing $\Gamma_X(p)$.

An application in \[24\] computes the 1-negative type gap of each odd cycle $C_{2k+1}$, $k \geq 1$. The formulas in \[24\] also require non-trivial combinatorial optimization and depend upon being able to find the inverse of the $p$-distance matrix of the underlying finite metric space $(X, d)$. For any given $p \geq 0$, Li and Weston \[18\] have shown that a finite metric metric space $(X, d)$ has strict $p$-negative type if and only if $\Gamma_X(p) > 0$. In particular, if $(X, d)$ is ultrametric, then $\Gamma_X(p) > 0$ for all $p \geq 0$. Conversely, if $\Gamma_X(p) > 0$ for all $p \geq 0$, then $(X, d)$ is ultrametric. These statements follow from aforementioned results in \[12\] and \[18\]. At this point it is worth recalling the formal definition of an ultrametric.

**Definition 1.3.** A metric $d$ on a set $X$ is said to be ultrametric if for all $x, y, z \in X$, we have:

$$d(x, y) \leq \max\{d(x, z), d(y, z)\} \quad (1.3)$$

De Groot \[8\] characterized ultrametric spaces up to homeomorphism as the strongly zero-dimensional metric spaces. In fact, ultrametric spaces arise in very specific ways as the end spaces or leaves of certain tree-like structures, and it is for this reason that they are exceptionally important in fields as diverse as computational logic \[19\], data analysis \[5\], non-commutative geometry \[16\], and quantum mechanics \[17\]. Hughes \[15\] has given a categorical equivalence between the end spaces of infinite trees and complete ultrametric spaces. More precisely; \[15\] Theorem 6.9 states that there is an equivalence from the category of geodesically complete, rooted $\mathbb{R}$-trees and equivalence classes of isometries at infinity, to the category of complete ultrametric spaces of finite diameter and local similarity equivalences. On the other hand, certain Hierarchical Clustering methods take as their input a finite set $X$ and output a dendrogram. Loosely speaking, a dendrogram is a nested family of partitions that is usually represented graphically as a rooted tree. In fact, dendrograms come in two main flavors: proximity and threshold. Proximity dendrograms and finite ultrametric spaces are equivalent. Indeed, given a finite set $X$, there is a natural bijection between the collection of all proximity dendrograms over $X$ and the collection of all ultrametrics on $X$. A nice account of this well-known equivalence is given by Carlsson and Mémoli \[5\] Theorem 9.

Now suppose that $(X, d)$ is a finite ultrametric space with minimum non-zero distance $\alpha$. As noted, $\Gamma_X(p) > 0$ for all $p \geq 0$. The purpose of this paper is to examine the limiting behavior of the ratio $\Gamma_X(p)/\alpha^p$ as $p \to \infty$. It turns out that there is an intriguing pattern. Our main result shows that the limit

$$\Gamma_X(\infty) = \lim_{p \to \infty} \frac{\Gamma_X(p)}{\alpha^p}$$

exists and is necessarily finite. Indeed, we derive an explicit combinatorial formula for $\Gamma_X(\infty)$. In order to state this formula it is helpful to introduce some additional notation and terminology. The following function
\[ \vartheta : \mathbb{N} \setminus \{1\} \rightarrow \mathbb{Q} \] figures prominently:

\[ \vartheta(n) = \frac{1}{2} \left( \left\lfloor \frac{n}{2} \right\rfloor^{-1} + \left\lceil \frac{n}{2} \right\rceil^{-1} \right). \]

The next notion has a simple interpretation in terms of proximity dendrograms. (See Remark 2.4)

**Definition 1.4.** Let \((X, d)\) be a finite ultrametric space with minimum non-zero distance \(\alpha\). Let \(z \in X\) be given. The closed ball \(B_z(\alpha) = \{ x \in X : d(x, z) \leq \alpha \}\) will be called a coterie in \(X\) if \(|B_z(\alpha)| > 1\).

It is a simple matter to verify the following statements about coteries. (See Remark 2.4.)

**Theorem 1.5.** Let \((X, d)\) be a finite ultrametric space with at least two points and minimum non-zero distance \(\alpha\). If \(B_1, B_2, \ldots, B_l\) are the distinct coteries of \((X, d)\), then

\[ \lim_{p \rightarrow \infty} \frac{\Gamma_X(p)}{a^p} = \left( \sum_{k=1}^l \vartheta(|B_k|) \right)^{-1}. \] (1.4)

As noted, the left hand side of (1.4) will be denoted \(\Gamma_X(\infty)\). We call \(\Gamma_X(\infty)\) the asymptotic negative type constant of \((X, d)\). In proving Theorem 1.5, we will see that the ratio \(\Gamma_X(p)/a^p\) is either constant or strictly increasing on \((0, \infty)\). (It is constant if and only if \(d\) is a scalar multiple of the discrete metric on \(X\).) Hence, recalling the definition of \(\Gamma_X(p)\), we obtain the following immediate corollary of Theorem 1.5

**Corollary 1.6.** Let \((X, d)\) be a finite ultrametric space with at least two points and minimum non-zero distance \(\alpha\). Let \(G \in (0, \Gamma_X(\infty))\) be given. Then, for all sufficiently large \(p\), we have

\[ \frac{G}{2} \left( \sum_{k=1}^n |\eta_k| \right)^2 + \sum_{j,i=1}^n \frac{d(z_j, z_i)}{a^\rho} \eta_j \eta_i \leq 0 \] (1.5)

for each finite subset \(\{z_1, \ldots, z_n\} \subseteq X\) and each choice of real numbers \(\eta_1, \ldots, \eta_n\) with \(\eta_1 + \cdots + \eta_n = 0\).

The organization of the rest of the paper is as follows. Section 2 recalls the relationship between proximity dendrograms and finite ultrametric spaces. Section 2 also discusses a basic method for computing the \(p\)-negative type gap of a metric space using normalized simplices and \(p\)-simplex gaps. Section 3 develops the combinatorial framework that underpins our proof of Theorem 1.5. Section 4 then uses analytical techniques to complete the derivation of Theorem 1.5.

A number of basic conventions are used in this paper. \(\mathbb{N}\) denotes the set of positive integers and is referred to as the set of natural numbers. Given \(n \in \mathbb{N}\), we use \(\lfloor n \rfloor\) to denote the segment of the first \(n\) natural numbers: \{1, 2, \ldots, \(n\)\}. Sums indexed over the empty set are always taken to be 0. In relation to Definition 1.1 in the case \(p = 0\) (and elsewhere), we define \(0^0 = 0\).

2. Proximity dendrograms, finite ultrametric spaces and normalized simplices

In this section we briefly recall the bijection between proximity dendrograms and finite ultrametric spaces. We also discuss the use of normalized simplices to compute the \(p\)-negative type gap \(\Gamma_X(p)\) of a metric space \((X, d)\). This provides the necessary theoretical framework for the derivation of our main results in the subsequent sections. We assume throughout that \(|X| > 1\).

We begin with a review of partitions and proximity dendrograms. Our approach and notation is largely based on that of Carlsson and Mémoli.

**Definition 2.1.** Given a non-empty finite set \(X\), we let \(\mathcal{P}(X)\) denote the set of all partitions of \(X\). Given a partition \(\mathcal{V} \in \mathcal{P}(X)\) we call each \(v \in \mathcal{V}\) a block of \(\mathcal{V}\). Given partitions \(\mathcal{V}, \mathcal{V}' \in \mathcal{P}(X)\) we say that \(\mathcal{V}'\) is a refinement of \(\mathcal{V}\) if for every block \(v' \in \mathcal{V}'\) there exists a block \(v \in \mathcal{V}\) such that \(v' \subseteq v\). A refinement \(\mathcal{V}'\) of \(\mathcal{V}\) is said to be proper if \(\mathcal{V}' \neq \mathcal{V}\).

Proximity dendrograms are predicated in terms of nested collections of partitions in the following manner.

**Definition 2.2.** A proximity dendrogram is a triple \(\mathcal{D} = (X, \{\alpha_0, \alpha_1, \ldots, \alpha_l\}, \pi)\) that consists of a non-empty finite set \(X\), finitely many real numbers \(0 = \alpha_0 < \alpha_1 < \cdots < \alpha_L\) and a function \(\pi : \{\alpha_0, \alpha_1, \ldots, \alpha_L\} \rightarrow \mathcal{P}(X)\) with the following properties:
(1) \( \pi(\alpha_0) = \{x : x \in X\} \),
(2) \( \pi(\alpha_\ell) = \{X\} \), and
(3) if \( k \in [\ell] \), then \( \pi(\alpha_{k-1}) \) is a proper refinement of \( \pi(\alpha_k) \).

We call the set \( \{\alpha_0, \alpha_1, \ldots, \alpha_\ell\} \) the proximity part of \( D \).

Notice that conditions (1), (2) and (3) in Definition 2.2 imply that \( \ell \geq 1 \). Every proximity dendrogram generates a directed rooted tree \( T \) and this ultimately leads to the definition of an ultrametric on \( X \).

\[
\begin{align*}
\pi(\alpha_0) & : \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}, \{g\}\} \\
\pi(\alpha_1) & : \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}, \{g\}\} \\
\pi(\alpha_2) & : \{\{a, b, c\}, \{d\}, \{e, f, g\}\} \\
\pi(\alpha_3) & : \{\{a, b, c\}, \{d, e, f, g\}\} \\
\pi(\alpha_4) & : \{\{a, b, c, d, e, f, g\}\}
\end{align*}
\]

**Figure 1.** A proximity dendrogram with the sequence of partitions on the left and the generated directed tree on the right. The subtree \( T(\nu) \) generated by the node \( \nu \) is circled.

**Definition 2.3.** Let \( D = (X, \{\alpha_0, \alpha_1, \ldots, \alpha_\ell\}, \pi) \) be a given proximity dendrogram. The nodes of the tree \( T = T(D) \) generated by \( D \) consist of the blocks appearing in the partitions \( \pi(\alpha_0), \ldots, \pi(\alpha_\ell) \). The edge set of \( T \) is then determined by the Hasse diagram for this node set, partially ordered by set inclusion. If there is an edge between nodes \( \nu \) and \( w \), it is directed to go from \( \nu \) to \( w \) if \( w \subset \nu \). In this case we say that \( w \) is to the left of \( \nu \). Notice that the leaves of \( T \) are the blocks in \( \pi(\alpha_0) \) and the root is the block \( r = X \).

The level of a node \( \nu \in T \) is the smallest \( k \) such that \( \nu \in \pi(\alpha_k) \). For each \( k \geq 0 \), we let \( \Pi_k \) denote the set of nodes at level \( k \). Each level \( k \) node \( \nu \in T \) generates a subtree \( T(\nu) \) of \( T \). By definition, \( T(\nu) \) is the subtree of \( T \) that contains of \( \nu \) together with all nodes and edges that are to the left of \( \nu \). The subtree \( T(\nu) \) naturally inherits the direction of \( T \). We note that if \( k = 0 \), then \( T(\nu) \) degenerates to a singleton. We let \( b(\nu) \) denote the degree of the node \( \nu \) in the subtree \( T(\nu) \) and call it the left degree of \( \nu \) in \( T \). It signifies the number of edges in \( T \) emanating to the left from \( \nu \). Notice that if \( k \in [\ell] \) and \( \nu \in \Pi_k \), then \( b(\nu) \geq 2 \).

So far, the proximity part \( \{\alpha_0, \alpha_1, \ldots, \alpha_\ell\} \) of a given proximity dendrogram \( D = (X, \{\alpha_0, \alpha_1, \ldots, \alpha_\ell\}, \pi) \) has played a muted role in our definitions. However, given \( x, y \in X \), it is not a difficult matter to verify that

\[
d_D(x, y) = \min \{\alpha_k : x, y \text{ belong to the same block of } \pi(\alpha_k)\}
\]

defines an ultrametric on \( X \). Conversely, if \( d \) is an ultrametric on a finite set \( X \) with at least two elements, then there is a unique proximity dendrogram \( D = (X, \{\alpha_0, \alpha_1, \ldots, \alpha_\ell\}, \pi) \) such that \( d = d_D \). We refer to Carlsson and Mémoli [5, Theorem 9] for more specific details concerning this bijection. We call \( D \) the proximity dendrogram for \( (X, d) \). Note that \( \alpha_1 < \alpha_2 < \cdots < \alpha_\ell \) are the non-zero distances in \( (X, d) \).

**Remark 2.4.** According to Definition 1.1, a coterie in \( X \) is a closed ball \( B_x(\alpha_1) \subseteq X \) with \( |B_x(\alpha_1)| > 1 \). Looking at the formula (2.1) for \( d_D(x, y) \) we see that the coteries in \( X \) are precisely the blocks in \( \Pi_1 \).
Now let \((X, d)\) be a given finite ultrametric space with at least two points. We may assume that \(d = d_\Psi\) for some unique proximity dendrogram \(D = (X, \{\alpha_0, \alpha_1, \ldots, \alpha_t\}, \pi)\). It is a simple observation that \((X, d^p)\) is a finite ultrametric space for each given \(p > 0\). We note that the proximity dendrogram for \((X, d^p)\) is simply \(D^p = (X, \{\alpha_0^p, \alpha_1^p, \ldots, \alpha_t^p\}, \pi')\) where \(\pi'(\alpha_k^p) = \pi(\alpha_k)\) for each \(k \geq 0\). It is thus the case that the trees generated by \(D\) and \(D^p\) are identical. In other words, \(T(D) = T(D^p)\) for all \(p > 0\). The metric \(d' = d/\alpha_0\) on \(X\) is also an ultrametric. Moreover, it is obvious that \((X, d)\) and \((X, d')\) share the same proximity dendrogram tree \(T\). The minimum non-zero distance in \((X, d')\) is 1 and it is not difficult to see from Definition 1.2 that

\[
\Gamma_{(X, d)}(p) = \alpha_0^p \cdot \Gamma_{(X, d')}(p).
\]

Provided we take this scaling into account, we may always assume that the minimum distance in our finite ultrametric space is 1. It turns out that doing so is not purely cosmic and, in fact, comnotes a surprisingly versatile technical tool. The full extent of this statement will become clearer in Sections 3 and 4. Prior to that we need to describe a method for computing the \(p\)-negative type gap \(\Gamma_X(p)\) of a metric space \((X, d)\).

The technique we wish to describe is based on the notion of normalized simplices.

**Definition 2.5.** Let \(X\) be a set and suppose that \(s, t > 0\) are integers. A normalized \((s, t)\)-simplex in \(X\) is a collection of pairwise distinct points \(x_1, \ldots, x_s, y_1, \ldots, y_t \in X\) together with a corresponding collection of positive real numbers \(m_1, \ldots, m_s, n_1, \ldots, n_t\) that satisfy \(m_1 + \cdots + m_s = 1 = n_1 + \cdots + n_t\). Such a configuration of points and real numbers will be denoted by \(D(\omega) = [x_j(m_j); y_i(n_i)]\) and will simply be called a simplex when no confusion can arise. We call the vertices \(x_1, \ldots, x_s \in D(\omega)\) the \(M\)-team and the vertices \(y_1, \ldots, y_t \in D(\omega)\) the \(N\)-team.

Simplices with weights on the vertices were introduced by Weston [22] to study the generalized roundness of finite metric spaces. The basis for the following definition is derived from the original formulation of generalized roundness due to Enflo [11]. Enflo introduced generalized roundness in order to address a problem of Smirnov concerning the uniform structure of Hilbert spaces. There are very intimate ties between negative type and generalized roundness. See, for example, Prassidis and Weston [20].

**Definition 2.6.** Suppose \(p \geq 0\) and let \((X, d)\) be a metric space. For each normalized \((s, t)\)-simplex \(D(\omega) = [x_j(m_j); y_i(n_i)]\) in \(X\) we define

\[
\gamma^p_D(\omega) = \sum_{j=1}^{s,t} m_j n_i d(x_j, y_i)^p - \sum_{1 \leq j_1 < j_2 \leq s} m_j m_{j_2} d(x_{j_1}, x_{j_2})^p - \sum_{1 \leq i_1 < i_2 \leq t} n_i n_{i_2} d(y_{i_1}, y_{i_2})^p.
\]

We call \(\gamma^p_D(\omega)\) the \(p\)-simplex gap of \(D(\omega)\) in \((X, d)\).

The significance of normalized simplices and \(p\)-simplex gaps for our purposes is the following theorem.

**Theorem 2.7.** Suppose \(p \geq 0\) and let \((X, d)\) be a metric space with \(p\)-negative type. Then

\[
\Gamma_X(p) = \inf_{D(\omega)} \gamma^p_D(\omega)
\]

where the infimum is taken over all normalized \((s, t)\)-simplices \(D(\omega)\) in \(X\).

The proof of Theorem 2.7 in the case \(p = 1\) is due to Doust and Weston [10] Theorem 4.16. The proof given in [10] may easily be adapted to deal with general \(p \geq 0\).

We conclude this section with an important technical comment concerning Definitions 2.5 and 2.6. Suppose that \((X, d)\) is a finite metric space. The set of normalized simplices in \(X\) can be identified with a compact set \(N \subset \mathbb{R}^n\). The first step is to fix an enumeration of the elements of \(X\); say, \(X = \{z_1, z_2, \ldots, z_n\}\). Let

\[
N = \left\{ \omega = (\omega_j) \in \mathbb{R}^n : \sum \omega_j = 0 \text{ and } \sum |\omega_j| = 2 \right\}.
\]

The set \(N\) inherits the Euclidean metric on \(\mathbb{R}^n\) and is thus a compact metric space. For each \(\omega \in N\) we may construct a normalized simplex in the following manner. If \(\omega_j > 0\), we put \(z_j\) on the \(M\)-team with corresponding weight \(\omega_j\). If \(\omega_j < 0\), we put \(z_j\) on the \(N\)-team with corresponding weight \(\omega_j\). By relabeling, the \(M\) and \(N\)-teams may be enumerated as \(x_1, \ldots, x_s\) and \(y_1, \ldots, y_t\) with corresponding weights \(m_1, \ldots, m_s\) and \(n_1, \ldots, n_t\), respectively. The normalized simplex generated in this fashion will be denoted \(D(\omega)\) without confusion. It is plain that every normalized simplex in \(X\) arises in this manner. Moreover, since \(N\) is a compact metric space, a simple continuity arguments furnishes the following theorem.
Theorem 2.8. Suppose \( p \geq 0 \) and let \((X, d)\) be a finite metric space with \( p \)-negative type. Then there exists a normalized simplex \( D(\omega_0) \) in \( X \) such that

\[
\Gamma_X(p) = \gamma_D^p(\omega_0).
\]

3. Combinatorial properties of ultrametric simplex gaps \( \gamma_D^p(\omega) \)

Throughout this section we assume that \((X, d)\) is a given finite ultrametric space with minimum non-zero distance \( \alpha_1 = 1 \) and associated proximity dendrogram \( D = (X, \{\alpha_0, \alpha_1, \ldots, \alpha_\ell\}, \pi) \). We further assume that \( D(\omega) = [x_i(m_i); y_j(n_j)]_{s,t} \) is a given, fixed, normalized \((s, t)\)-simplex in \( X \). On this basis we formulate the following critical notions.

Definition 3.1. For each node \( v \in T \) we define the partition sums:

\[
M(v) = \sum_{j : (x_j) \in T(v)} m_j, \quad \text{and} \quad N(v) = \sum_{i : (y_i) \in T(v)} n_i.
\]

We say that the subtree \( T(v) \) is simplicially balanced if \( M(v) = N(v) \).

Remark 3.2. If \( v \) is a block of the form \( \{x_j\} \) or \( \{y_i\} \), then \( |M(v) - N(v)| = m_j \) or \( n_i \) (respectively), and so \( T(v) \) is not simplicially balanced.

Now suppose that \( p > 0 \) and let \( \gamma(p) = \gamma_D^p(\omega) \). We wish to study the combinatorial properties of the simplex gap \( \gamma : (0, \infty) \rightarrow \mathbb{R} \). We will see shortly that the range of \( \gamma \) is a subset of \((0, \infty)\). By Definition 2.8

\[
\gamma(p) = \sum_{j,i=1}^{s,t} m_j n_i d^p(x_j, y_i) - \sum_{1 \leq j_1 < j_2 \leq s} m_{j_1} m_{j_2} d^p(x_{j_1}, x_{j_2}) - \sum_{1 \leq i_1 < i_2 \leq t} n_{i_1} n_{i_2} d^p(y_{i_1}, y_{i_2}).
\]

As the non-zero distances in \((X, d)\) are \( \alpha_1, \ldots, \alpha_\ell \) we see that \( \gamma(p) = c_1 \alpha_1^p + \cdots + c_\ell \alpha_\ell^p \) where \( c_1, \ldots, c_\ell \) are constants to be determined.

3.1. Evaluation and properties of the constants \( c_k \). Suppose \( k \in [\ell] \). We wish to determine a formula for \( c_k \). We begin by noting that for all \( x, y \in X \), \( d(x, y) = \alpha_k \) if and only if there exists a level \( k \) node \( v \in T \) such that the geodesic from \( \{x\} \) to \( \{y\} \) in \( T \) passes through the node \( v \) and is contained in the subtree \( T(v) \). In order to better grasp this condition on \( x, y \) and \( v \), let \( u_1, u_2, \ldots, u_b \), where \( b = b(v) \), denote the nodes in \( T(v) \) that are adjacent to \( v \). Then, the geodesic from \( \{x\} \) to \( \{y\} \) in \( T \) passes through the node \( v \) and is contained in the subtree \( T(v) \) if and only if \( x, y \) belong to distinct blocks in the list \( u_1, u_2, \ldots, u_b \). We let \( c_k(v) \) denote the contribution of the level \( k \) node \( v \) to \( c_k \). In other words,

\[
c_k = \sum_{v \in T} c_k(v),
\]

where the sum is taken over all level \( k \) nodes \( v \in T \). With these preliminary comments and notations in mind we are in a position to state our first lemma.

Lemma 3.3. For each \( k \in [\ell] \) and each level \( k \) node \( v \in T \),

\[
2c_k(v) = \sum_{i=1}^{b(v)} (M(u_i) - N(u_i)) \cdot \left( \sum_{j=1 \atop j \neq i}^{b(v)} (N(u_j) - M(u_j)) \right)
\]

\[
= \sum_{i=1}^{b(v)} \{(M(u_i) - N(u_i))^2 - (M(v) - N(v))^2\},
\]

where \( u_1, u_2, \ldots, u_b \) denote the nodes in \( T(v) \) that are adjacent to \( v \).
where \( u_1, u_2, \ldots, u_{b(v)} \) are the nodes in \( T(v) \) that are adjacent to \( v \). If, moreover, the subtree \( T(v) \) is simplicially balanced, then

\[ 2c_k(v) = \sum_{i=1}^{b(v)} (M(u_i) - N(u_i))^2. \]

**Proof.** For simplicity we set \( b = b(v) \). We first note that \( M(v) = \sum_{j=1}^{b} M(u_j) \) and \( N(v) = \sum_{j=1}^{b} N(u_j) \).

Therefore,

\[ \sum_{j=1 \atop j \neq i}^{b} (N(u_j) - M(u_j)) = (M(u_i) - N(u_i)) - (M(v) - N(v)), \]

for each \( i \in [b] \). Now, by definition of \( \gamma(p) \) and \( c_k(v) \), we see that

\[ 2c_k(v) = \sum_{i,j=1 \atop j \neq i}^{b} 2M(u_i)N(u_j) - \sum_{1 \leq i < j \leq b} 2 \{ M(u_i)M(u_j) + N(u_i)N(u_j) \} \]

\[ = \sum_{i=1}^{b} (M(u_i) - N(u_i)) \cdot \left\{ \sum_{j=1 \atop j \neq i}^{b} (N(u_j) - M(u_j)) \right\} \]

\[ = \sum_{i=1}^{b} (M(u_i) - N(u_i)) \cdot \{(M(u_i) - N(u_i)) - (M(v) - N(v))\} \]

\[ = \sum_{i=1}^{b} \{(M(u_i) - N(u_i))^2\} - \sum_{i=1}^{b} (M(u_i) - N(u_i))(M(v) - N(v)) \]

\[ = \sum_{i=1}^{b} \{(M(u_i) - N(u_i))^2\} - (M(v) - N(v))^2. \]

If \( T(v) \) is simplicially balanced, then (by definition) \( M(v) - N(v) = 0 \). The second assertion of the lemma therefore follows immediately from the preceding calculation.

If \( v \) is a level \( k \) node in \( T \) with \( k \in [\ell] \), we let \( \text{Adj}(v) \) denote the set of nodes in \( T(v) \) that are adjacent to \( v \). According to Lemma 3.3

\[ 2c_k(v) = \sum_{u \in \text{Adj}(v)} \{(M(u) - N(u))^2\} - (M(v) - N(v))^2. \]  

(3.1)

For the special case \( k = \ell \) there is only one node at level \( k \); namely, the root \( r \) of the entire dendrogram tree \( T \). In this case \( T = T(r) \) is automatically simplicially balanced by definition of a normalized \((s, t)\)-simplex. As \( c_\ell = c_\ell(r) \) we obtain the following immediate consequence of Lemma 3.3

**Corollary 3.4.**

\[ 2c_\ell = \sum_{u \in \text{Adj}(r)} (M(u) - N(u))^2. \]

In particular, \( c_\ell \) is positive or 0. Moreover, \( c_\ell = 0 \) if and only if \( M(u) = N(u) \) for each \( u \in \text{Adj}(r) \). In other words, \( c_\ell = 0 \) if and only if \( T(u) \) is simplicially balanced for each node \( u \in T \) that is adjacent to \( r \).

Clearly, \( \gamma(p) \) is constant on \((0, \infty)\) if and only if \( c_2 = c_3 = \cdots = c_\ell = 0 \). It is worth examining precisely when this occurs. By Corollary 3.4 \( c_\ell = 0 \) if and only if \( T(v) \) is simplicially balanced for each node \( v \in \text{Adj}(r) \). Suppose that this is indeed the case. Then, by Lemma 3.3 we see that

\[ 2c_\ell-1(v) = \sum_{u \in \text{Adj}(v)} \{(M(u) - N(u))^2\} \]
for each level \( \ell - 1 \) node \( v \in T \). This is because each such node \( v \in \text{Adj}(r) \). Thus, given that \( c_\ell = 0 \), we see that \( c_{\ell - 1} \) is positive or zero. Moreover, \( c_{\ell - 1} = 0 \) if and only if for each level \( \ell - 1 \) node \( v \) and each \( u \in \text{Adj}(v) \), \( T(u) \) is simplicially balanced.

Now suppose that \( c_\ell = c_{\ell - 1} = 0 \). Let \( w \) be a given level \( \ell - 2 \) node. The salient observation is that \( w \) belongs to \( \text{Adj}(v) \) for some level \( \ell - 1 \) node \( v \) or \( \text{Adj}(r) \). Either way, \( T(w) \) is simplicially balanced, and so
\[
2c_{\ell - 2}(w) = \sum_{u \in \text{Adj}(w)} \{(M(u) - N(u))^2\}
\]
by Lemma 3.3. Thus, given that \( c_\ell = c_{\ell - 1} = 0 \), we see that \( c_{\ell - 2} \) is positive or zero. Moreover, \( c_{\ell - 2} = 0 \) if and only if for each level \( \ell - 2 \) node \( w \) and each \( u \in \text{Adj}(w) \), \( T(u) \) is simplicially balanced.

More generally, suppose that it is the case that \( c_\ell = c_{\ell - 1} = \cdots = c_{k+1} = 0 \) for some \( k \geq 2 \). Let \( w \) be a given level \( k \) node. Then there is a unique \( j > k \) such that \( w \in \text{Adj}(v) \). As \( c_\ell = \cdots = c_j = 0 \), the same reasoning as above ensures that \( T(w) \) is simplicially balanced. So
\[
2c_k(w) = \sum_{u \in \text{Adj}(w)} \{(M(u) - N(u))^2\}
\]
by Lemma 3.3. Thus, given that \( c_\ell = c_{\ell - 1} = \cdots = c_{k+1} = 0 \), we see that \( c_k \) is positive or zero. Moreover, \( c_k = 0 \) if and only if for each level \( k \) node \( w \) and each \( u \in \text{Adj}(w) \), \( T(u) \) is simplicially balanced.

In summary, we have established the equivalence of conditions (a), (b) and (c) in the statement of the following theorem.

**Theorem 3.5.** The following conditions are equivalent:

(a) \( \gamma(p) \) is constant on \( (0, \infty) \),
(b) \( c_\ell = c_{\ell - 1} = \cdots = c_2 = 0 \),
(c) for each node \( v \in \Pi_k \) such that \( k \geq 2 \) and each \( u \in \text{Adj}(v) \), \( T(u) \) is simplicially balanced, and
(d) \( T(u) \) is simplicially balanced for each node \( u \in \Pi_1 \) and no \( \{x_j\} \) or \( \{y_i\} \) is adjacent to any level \( k \) node with \( k \geq 2 \).

It is also the case that if \( c_\ell = c_{\ell - 1} = \cdots = c_{k+1} = 0 \) for some \( k \geq 2 \), then \( c_k = c_\ell + c_{\ell - 1} + \cdots + c_k \) is positive or zero.

**Proof.** The comments in the lead up to the statement of the theorem establish that conditions (a), (b) and (c) are equivalent. The same comments also demonstrate that if \( c_\ell = c_{\ell - 1} = \cdots = c_{k+1} = 0 \) for some \( k \geq 2 \), then \( c_k = c_\ell + c_{\ell - 1} + \cdots + c_k \) is positive or zero. It is also clear that condition (c) implies condition (d). This is because every level 1 node is adjacent to some level \( k \) node with \( k \geq 2 \) and, moreover, no \( T(\{x_j\}) \) or \( T(\{y_i\}) \) is simplicially balanced. Now suppose that condition (d) holds. By way of illustration, consider an arbitrary level \( 2 \) node \( v \in T \). Each node \( u \in \text{Adj}(v) \) is a level 1 node or a level 0 node. If \( u \in \text{Adj}(v) \) is a level 1 node, then \( M(u) = N(u) \) by hypothesis. If \( u \in \text{Adj}(v) \) is a level 0 node, then \( u \) carries no weight (because no \( \{x_j\} \) or \( \{y_i\} \) is adjacent to any level 2 node) and so \( M(u) = 0 = N(u) \). Moreover,
\[
M(v) = \sum_{u \in \text{Adj}(v) \cap \Pi_1} M(u) = \sum_{u \in \text{Adj}(v) \cap \Pi_1} N(u) = N(v).
\]
Thus \( T(v) \) is simplicially balanced for every level 2 node \( v \in T \). The argument now proceeds in the obvious fashion to establish condition (c).

Corollary 3.4 is a special instance of a more general phenomenon involving the formulas (3.1).

**Theorem 3.6.** \( \sum_{i=k}^{\ell} c_i \geq 0 \) for each \( k \in [\ell] \). Moreover, \( 2(c_1 + \cdots + c_\ell) = m_1^2 + \cdots + m_2^2 + n_1^2 + \cdots + n_2^2 > 0 \).

**Proof.** As in (3.1), consider a level \( k \) node \( v \) but with the additional assumption that \( k \neq \ell \). Then there is a unique node \( w \in T \) such that \( v \in \text{Adj}(w) \). Now \( w \) is a level \( j \) node for some unique \( j > k \). According to Lemma 3.3, \(-(M(v) - N(v))^2\) is the lone negative term in the expression (3.1) for \( 2c_k(v) \). On the other hand, \((M(v) - N(v))^2\) is a positive term in the corresponding expression for \( 2c_j(w) \). So adding
2c_k(v) to 2c_j(w) will eliminate the terms ±(M(v) − N(v))^2 by cancellation. As a result, due to repeated cancellations of this nature, it follows that

\[ c_k + c_{k+1} + \cdots + c_\ell = \sum \{ c_j(w) : j \geq k \text{ and } w \text{ is a level } j \text{ node in } T \} \geq 0. \]

Now suppose that \( u \in \text{Adj}(v) \) where, again, \( v \) is a level \( k \) node in \( T \) with \( k \in [\ell] \). If \( u \) is an internal node of \( T \) of level \( i \) (say), then \( i \in [k-1] \). In this case, \( + (M(u) - N(u))^2 \) appears as a positive term in the expression (3.1) for \( 2c_k(v) \) while \( -(M(u) - N(u))^2 \) is the lone negative term in the corresponding expression for \( 2c_i(u) \). So adding \( 2c_k(v) \) to \( 2c_i(u) \) will once again eliminate the terms \( ±(M(u) - N(u))^2 \) by cancellation. However, we cannot discount the possibility that \( u \) may be a leaf of \( T \). Indeed, if \( u \) is a level 0 node in \( T \), then the positive term \(+ (M(u) - N(u))^2 \) in the expression for \( 2c_k(v) \) is not cancelled in the following sum:

\[ c_1 + c_2 + \cdots + c_\ell = \sum \{ c_j(w) : j \geq 1 \text{ and } w \text{ is a level } j \text{ node in } T \}. \]

All other terms in the expression for \( 2c_k(v) \) are cancelled by our preceding comments. Thus:

\[
2(c_1 + c_2 + \cdots + c_\ell) = \sum \{(M(u) - N(u))^2 : u \text{ is a leaf of } T\} \\
= \sum \{(M(u) - N(u))^2 : u = \{x_j\} \text{ for some } j \text{ or } u = \{y_i\} \text{ for some } i\} \\
= m_1^2 + \cdots + m_s^2 + n_1^2 + \cdots + n_t^2. \tag{3.2}
\]

Remark 3.7. Let \( n = |X| \). The right hand side of (3.2) is subject to the constraint \( m_1 + \cdots + m_s = 1 = n_1 + \cdots + n_t \) and is easily seen to be at least \( s^{-1} + t^{-1} \) by elementary calculus. Moreover,

\[ s^{-1} + t^{-1} \geq \left( \left\lfloor \frac{n}{2} \right\rfloor^{-1} + \left\lceil \frac{n}{2} \right\rceil^{-1} \right), \]

because \( s + t \leq n \). Hence,

\[
\lim_{p \to 0^+} \gamma(p) = c_1 + \cdots + c_\ell \geq \frac{1}{2} \left( \left\lfloor \frac{n}{2} \right\rfloor^{-1} + \left\lceil \frac{n}{2} \right\rceil^{-1} \right).
\]

Theorem 3.8. The function \( \gamma(p) = c_1 \alpha_1^p + \cdots + c_\ell \alpha_\ell^p \) is constant or strictly increasing on \((0, \infty)\).

Proof. Suppose \( p_2 > p_1 > 0 \). For each \( i \in [\ell] \) we set \( \delta_i = \alpha_i^{p_2} - \alpha_i^{p_1} \). Notice that \( \delta_1 = 0 \) because \( \alpha_1 = 1 \) but \( \delta_i > 0 \) otherwise. Since the function \( g(p) = b^p - a^p \) is strictly increasing on \((0, \infty)\) provided \( 1 \leq a < b \), it follows readily that

\[ \delta_i - \delta_{i-1} = (\alpha_i^{p_2} - \alpha_i^{p_1}) - (\alpha_i^{p_1} - \alpha_i^{p_1}) > 0 \]

for each \( i \geq 2 \). Thus, on rearranging the resulting sum, the difference

\[
\gamma(p_2) - \gamma(p_1) = \sum_{i=1}^\ell c_i \delta_i = (c_1 + c_2 + \cdots + c_\ell) \delta_1 \\
+ (c_2 + c_3 + \cdots + c_\ell) (\delta_2 - \delta_1) \\
+ \cdots \\
+ (c_{\ell-1} + c_\ell) (\delta_{\ell-1} - \delta_{\ell-2}) \\
+ c_\ell (\delta_\ell - \delta_{\ell-1}), \tag{3.3}
\]

is non-negative by Theorem 3.6. Now suppose that \( \gamma \) is non-constant on \((0, \infty)\). Then we may consider the largest \( k \geq 2 \) such that \( c_k \neq 0 \). By the second part of Theorem 3.6, \( c_k = c_k + c_{k+1} + \cdots + c_\ell > 0 \). Hence the term \((c_k + c_{k+1} + \cdots + c_\ell)(\delta_k - \delta_{k-1})\), which appears on the right hand side of (3.3), is positive. Thus \( \gamma(p_2) - \gamma(p_1) > 0 \) in this instance. \( \square \)
4. Flat simplices and the computation of $\Gamma_X(\infty)$

Let $(X,d)$ be a given finite ultrametric space with minimum non-zero distance $\alpha_1 = 1$ and associated proximity dendrogram $D = (X,\{\alpha_0, \alpha_1, \ldots, \alpha_k\}, \pi)$. Fix an enumeration of the elements of $X$: say, $X = \{x_1, x_2, \ldots, x_n\}$. We begin by showing that $\Gamma_X(\infty)$ exists in the real line.

**Theorem 4.1.** The limit

$$\Gamma_X(\infty) = \lim_{p \to \infty} \Gamma_X(p)$$

exists in the real line and is at most 1.

**Proof.** By Theorem 2.7

$$\Gamma_X(p) = \inf_{D(\omega)} \gamma_D^p(\omega)$$

where the infimum is taken over all normalized $(s,t)$-simplices $D(\omega)$ in $X$. Thus, by Theorem 3.8, $\Gamma_X(p)$ is non-decreasing on $(0,\infty)$. To complete the proof it remains to show that $\Gamma_X(p)$ is bounded above independently of $p$. This is easy. We may choose distinct $x, y \in X$ with $d(x,y) = 1$. Then the normalized $(1,1)$-simplex $D(\omega_0) = \{x\{1\}, y\{1\}\}$ satisfies $\gamma_D^p(\omega_0) = 1$ for all $p > 0$. Hence $\Gamma_X(p) \leq 1$ for all $p > 0$. □

We will see presently that the value of $\Gamma_X(\infty)$ is determined by the following class of normalized simplices.

**Definition 4.2.** A normalized simplex $D(\omega)$ in $(X,d)$ is said to be flat if $\gamma_D^p(\omega)$ is constant on $(0,\infty)$.

Recall that the set of all normalized simplices in $X$ may be identified with the compact metric space

$$\mathcal{N} = \{\omega = (\omega_j) \in \mathbb{R}^n : \sum \omega_j = 0 \text{ and } \sum |\omega_j| = 2\}.$$ 

Moreover, given $\omega \in \mathcal{N}$, $D(\omega) = [x_j(m_j); y_j(n_j)]_{x\{1\}}$ denotes the associated normalized simplex. Under these identifications, we let $\mathcal{K} = \{\omega \in \mathcal{N} : D(\omega) \text{ is a flat simplex in } (X,d)\}$. The set $\mathcal{K}$ proves decisive in computing the precise value of $\Gamma_X(\infty)$.

**Lemma 4.3.** $\mathcal{K}$ is a closed set in $\mathbb{R}^n$.

**Proof.** By Theorem 3.8 a vector $\omega$ belongs to $\mathcal{K}$ if and only if $\sum \omega_j = 0$, $\sum |\omega_j| = 2$, $M(u) = N(u)$ for each $u \in \Pi_1$, and all non-coterie weights are 0. So $\omega$ needs to satisfy a finite number of equations. Thus $\mathcal{K}$ is closed in $\mathbb{R}^n$. □

**Theorem 4.4.**

$$\Gamma_X(\infty) = \inf_{\omega \in \mathcal{K}} \gamma_D^\infty(\omega).$$

**Proof.** First of all notice that by Theorem 2.7 and Definition 4.2 we have

$$\Gamma_X(\infty) = \lim_{p \to \infty} \Gamma_X(p)$$

$$= \lim_{p \to \infty} \inf_{\omega \in X} \gamma_D^p(\omega)$$

$$\leq \lim_{p \to \infty} \inf_{\omega \in \mathcal{K}} \gamma_D^p(\omega)$$

$$= \inf_{\omega \in \mathcal{K}} \gamma_D^\infty(\omega).$$

The switch to $p = 1$ in the last line of (4.2) occurs because $\mathcal{K}$ is the collection of flat simplices. We may choose an increasing sequence of positive real numbers $(p_k)$ with $p_1 = 1$ so that $p_k \to \infty$ as $k \to \infty$. For each $k$ we may, by Theorem 2.7 choose a vector $\omega_k \in \mathcal{N}$ so that $\Gamma_X(p_k) = \gamma_D^{p_k}(\omega_k)$. Now, since $\mathcal{N} \subset \mathbb{R}^n$ is compact, $(\omega_k)$ has a convergent subsequence. We may assume, without loss, that we have passed to such a subsequence and that the limit is $\omega_\infty \in \mathcal{N}$.

We claim that $\omega_\infty \in \mathcal{K}$. Indeed, if we assume that $\omega_\infty \notin \mathcal{K}$, then $\gamma_D^{p_k}(\omega_\infty)$ is a strictly increasing function of $p$ (whose limit at $\infty$ is $\infty$) by Theorem 3.8. So, by continuity, there is some $q \in (0,\infty)$ such that $\gamma_D^{p_k}(\omega_\infty) = 2$. Moreover, as $\gamma_D^{p_k}(\cdot)$ is continuous in $\omega$, there must exist some ball $B \subset \mathcal{N} \setminus \mathcal{K}$, centered at $\omega_\infty$ with radius $r > 0$, so that for any $\omega \in B$ we have $\gamma_D^{p_k}(\omega) > 1$. Since $\gamma_D^{p_k}(\cdot)$ never decreases on $\mathcal{N} \setminus \mathcal{K}$ we will have $\gamma_D^{p_k}(\omega) > 1$ for any $\omega \in B$ and any $p \geq q$. Since $(\omega_k)$ converges to $\omega_\infty$, there must exist an $N$ so that $(\omega_k)_{k \geq N} \subset B$. However, we have chosen each $\omega_k$ so that $\gamma_D^{p_k}(\omega_k) = \Gamma_X(p_k)$ and, by Theorem 4.1, $\Gamma_X(p) \leq 1$ for all $p$. So we have obtained a contradiction. It therefore follows that $\omega_\infty \in \mathcal{K}$.
By definition of $K$, $\gamma_D^1(\omega_\infty) = \gamma_D^p(\omega_\infty)$ for all $p > 0$. We claim that
\[
\lim_{k \to \infty} \gamma_D^p(\omega_k) = \gamma_D^1(\omega_\infty). \tag{4.3}
\]
Let $\epsilon > 0$ be given. The function $\gamma_D^1(\cdot)$ is continuous in $\omega$. Hence there exists an $r > 0$ such that if $\omega$ is in the ball $B_{\omega_\infty}(r)$, then $|\gamma_D^1(\omega) - \gamma_D^1(\omega_\infty)| < \epsilon$ and so, in particular, $\gamma_D^1(\omega_\infty) - \gamma_D^1(\omega) < \epsilon$. For each fixed $\omega \in N$, $\gamma_D^1(\omega)$ is a non-decreasing function of $p$. Thus, for all $\omega \in B_{\omega_\infty}(r)$ and all $p \geq 1$, we see that $\gamma_D^1(\omega_\infty) - \gamma_D^p(\omega) \leq \gamma_D^1(\omega_\infty) - \gamma_D^1(\omega) < \epsilon$. Recalling that $(\omega_k)$ converges to $\omega_\infty$, we may choose an $N \in N$ so that $(\omega_k)_{k \geq N} \subset B_{\omega_\infty}(r)$. Hence, for $k \geq N$, we have $\gamma_D^1(\omega_\infty) - \gamma_D^p(\omega_k) \leq \gamma_D^1(\omega_\infty) - \gamma_D^1(\omega_k) < \epsilon$. On the other hand, $\omega_k$ is chosen so that
\[
\gamma_D^p(\omega_k) = \Gamma_X(p_k) = \inf_{\omega \in K} \gamma_D^p(\omega).
\]
Hence $\gamma_D^1(\omega_\infty) - \gamma_D^p(\omega_k) = \gamma_D^p(\omega_\infty) - \gamma_D^p(\omega_k) \geq 0$. So we have that $0 \leq \gamma_D^1(\omega_\infty) - \gamma_D^p(\omega_k) < \epsilon$ for all $k \geq N$. As $\epsilon > 0$ was arbitrary, this establishes (4.3). Thus,
\[
\Gamma_X(\infty) = \lim_{p \to \infty} \Gamma_X(p) = \lim_{k \to \infty} \Gamma_X(p_k) = \lim_{k \to \infty} \gamma_D^p(\omega_k) = \gamma_D^1(\omega_\infty) \geq \inf_{\omega \in K} \gamma_D^1(\omega). \tag{4.4}
\]
The theorem now follows from (4.2) and (4.4).

**Theorem 4.5.** Let $B_1, B_2, \ldots, B_l$ denote the distinct coteries of $(X, d)$. Then,
\[
\Gamma_X(\infty) = \left( \sum_{k=1}^{l} \theta(|B_k|)^{-1} \right)^{-1}. \tag{4.5}
\]

**Proof.** The inherited metric on each coterie $B_k$ is simply the discrete metric on $B_k$. This is because we are assuming that the minimum non-zero distance in our finite ultrametric space $(X, d)$ is $\alpha_1 = 1$. Hence $\Gamma_{B_k}(1) = \Gamma_{B_k}(0) = \theta(|B_k|)$ by Weston [23, Theorem 3.2]. This fact will be used below.

Some additional notation will prove decisive. By Theorem 3.5 we have that $\omega \in K$ if and only if
\[
\sum_{j : x_j \in B_k} m_j = \sum_{i \neq j \in B_k} n_i
\]
for each $k \in [l]$, and no $x_j$ or $y_i$ belongs to $X \setminus (B_1 \cup \cdots \cup B_l)$. Bearing this in mind, consider a fixed $\omega \in K$. For each $k \in [l]$ we form a new vector $\omega_k$ from $\omega$ in the following manner: if $z_j \notin B_k$ redefine $\omega_j$ to be 0. Otherwise, make no change to $\omega_j$. As usual, we let $D(\omega_k)$ denote the (not necessarily normalized) simplex that corresponds to $\omega_k$. The 1-simplex gap $\gamma_D^1(\omega_k)$ is still defined according to the formula given in Definition 3.6 and it is not difficult to see that $\gamma_D^1(\omega) = \gamma_D^1(\omega_1) + \cdots + \gamma_D(\omega_l)$ because distances between the coteries cancel. Moreover, provided
\[
w_k = \sum_{j : x_j \in B_k} m_j > 0,
\]
we may further set $D(\omega_k) = D(\omega_k/w_k)$. (In the event that $w_k = 0$ it suffices to let $D(\omega_k)$ be any normalized simplex in $B_k$.) Then $D(\omega_k)$ is a normalized simplex in $B_k$, $\omega_k \in K$ and it is evident that $\gamma_D^1(\omega_k) = w_k^2 \gamma_D^1(V_k)$. Thus, $\gamma_D^1(\omega) = w_1^2 \gamma_D^1(V_1) + \cdots + w_l^2 \gamma_D^1(V_l)$. 


By Theorem 4.4
\[
\Gamma_X(\infty) = \inf_{\omega \in \mathcal{K}} \gamma_1^D(\omega)
\]
\[
= \inf_{\omega_1, \ldots, \omega_l} \gamma_1^D(\omega_1) + \cdots + \gamma_1^D(\omega_l)
\]
\[
= \inf_{\omega_1, \ldots, \omega_l} w_1^2\gamma_1^D(\omega_1) + \cdots + w_l^2\gamma_1^D(\omega_l)
\]
\[
= \inf \left\{ w_1^2\gamma_1^D(\kappa_1) + \cdots + w_l^2\gamma_1^D(\kappa_l) : D(\kappa_k) \text{ is a normalized simplex in } B_k \text{ and } \sum_{k=1}^l w_k = 1 \right\}
\]
\[
= \inf \left\{ w_1^2T_{B_1}(1) + \cdots + w_l^2T_{B_l}(1) : \sum_{k=1}^l w_k = 1 \right\}
\]
\[
= \inf \left\{ w_1^2\vartheta(|B_1|) + \cdots + w_l^2\vartheta(|B_l|) : \sum_{k=1}^l w_k = 1 \right\}
\]
\[
= \left( \sum_{k=1}^l \vartheta(|B_k|) \right)^{-1}.
\]

The last equality follows from an application of Lagrange’s multiplier theorem. \(\square\)

We note that Theorem 1.5 (and, hence, Corollary 1.6) follow directly from 2.2 and Theorem 4.5.

References

[1] S. Arora, S. Rao and U. Vazirani, Expander flows, geometric embeddings, and graph partitionings, in: 36th Annual Symposium on the Theory of Computing (2004), 222–231.

[2] S. Arora, J. R. Lee and A. Naor, Euclidean distortion and the sparsest cut, J. Amer. Math. Soc. 21 (2008), 1–21.

[3] Y. Bartal, N. Linial, M. Mendel and A. Naor, Low dimensional embeddings of ultrametrics, European J. Comb. 25 (2004) 87–92.

[4] Y. Bartal, B. Bollobás and M. Mendel, Ramsey-type theorems for metric spaces with applications to online problems, J. Comput. Syst. Sci. 72 (2006) 890–921.

[5] G. Carlsson and F. Mémoli, Characterization, Stability and Convergence of Hierarchical Clustering Methods, Journal of Machine Learning Research 11 (2010) 1425–1470.

[6] A. Cayley, On a theorem in the geometry of position, Cambridge Mathematical Journal 11 (1841), 267–271. (Also in The Collected Mathematical Papers of Arthur Cayley (Vol. I), Cambridge University Press, Cambridge (1889), 1–4.)

[7] S. Chawla, A. Gupta and H. Racke, Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut, in: Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms, Vancouver (2005), 102–111.

[8] J. de Groot, Non-Archimedean metrics in topology, Proc. Amer. Math. Soc. 7 (1956), 948–953.

[9] I. Doust and A. Weston, Enhanced negative type for finite metric trees, J. Funct. Anal. 254 (2008), 2336–2364.

[10] I. Doust and A. Weston, Corrigendum to “Enhanced negative type for finite metric trees”, J. Funct. Anal. 255 (2008), 532–533.

[11] P. Enflo, On a problem of Smirnov, Ark. Mat. 8 (1969), 107–109.

[12] T. Faver, K. Kochalski, M. Murugan, H. Verheggen, E. Wesson and A. Weston, Roundness properties of ultrametric spaces, Glasgow Math. J. (to appear).

[13] P. Hjorth, P. Lisoněk, S. Markvorsen and C. Thomassen, Finite metric spaces of strictly negative type, Linear Algebra Appl. 270 (1998), 255–273.

[14] P. G. Hjorth, S. L. Kokkendorff and S. Markvorsen, Hyperbolic spaces are of strictly negative type, Proc. Amer. Math. Soc. 130 (2002), 175–181.

[15] B. Hughes, Trees and ultrametric spaces: a categorical equivalence, Adv. Math. 189 (2004), 148–191.

[16] B. Hughes, Trees, ultrametrics, and noncommutative geometry, Pure and Applied Mathematics Quarterly 8 (2012), 221–312.

[17] S. V. Kozyrev and A. Yu. Khrennikov, Localization in space for a free particle in ultrametric quantum mechanics, Doklady Akademii Nauk. 411 (2006), 319–322.

[18] H. Li and A. Weston, Strict p-negative type of a metric space, Positivity 14 (2010), 529–545.

[19] F. Murtagh, Ultrametric and generalized ultrametric in computational logic and in data analysis, arXiv:1008.3559v1 [cs.LO].

[20] E. Prassidis and A. Weston, Manifestations of non linear roundness in analysis, discrete geometry and topology, in: Limits of Graphs in Group Theory and Computer Science, Research Proceedings of the ´Ecole Polytechnique Fédérale de Lausanne, CRC Press (2009), 141–170.
[21] I. J. Schoenberg, Remarks to Maurice Frechet's article “Sur la définition axiomatique d’une classe d’espaces distanciés vectoriellement applicable sur l’espace de Hilbert.”, Ann. Math. 36 (1935), 724–732.
[22] A. Weston, On the generalized roundness of finite metric spaces, J. Math. Anal. Appl. 192 (1995), 323–334.
[23] A. Weston, Optimal lower bound on the supremal strict $p$-negative type of a finite metric space, Bull. Aust. Math. Soc. 80 (2009), 486–497.
[24] R. Wolf, On the gap of finite metric spaces of $p$-negative type, Lin. Alg. Appl. 436 (2012), 1246–1257.

School of Mathematics and Statistics, University of New South Wales, Sydney, New South Wales 2052, Australia
E-mail address: i.doust@unsw.edu.au

School of Mathematics and Statistics, University of New South Wales, Sydney, New South Wales 2052, Australia
E-mail address: stephen.sanchez@unsw.edu.au

Department of Mathematics and Statistics, Canisius College, Buffalo, NY 14208, USA
E-mail address: westona@canisius.edu

Department of Decision Sciences, University of South Africa, PO Box 392, UNISA 0003, South Africa
E-mail address: westoar@unisa.ac.za