ULTRAMETRIC SUBSETS WITH LARGE HAUSDORFF DIMENSION

MANOR MENDEL AND ASSAF NAOR

Abstract. For every $\varepsilon \in (0,1)$ it is shown that every compact metric space X has a closed subset $Y \subseteq X$ which embeds into an ultrametric space with distortion $O(1/\varepsilon)$, and $\dim_H(Y) \geqslant (1-\varepsilon)\dim_H(X)$, where $\dim_H(\cdot)$ denotes Hausdorff dimension.

Contents

1. Introduction

Given $D \geqslant 1$, a metric space (X,d_X) is said to embed with distortion D into a metric space (Y,d_Y) if there exists $f: X \to Y$ and $\lambda > 0$ such that for all $x,y \in X$ we have

$$\lambda d_X(x,y) \leqslant d_Y(f(x),f(y)) \leqslant D\lambda d_X(x,y). \tag{1}$$

Note that when Y is a Banach space the scaling factor λ can be dropped in the definition (1).

Answering positively a conjecture of Grothendieck [10], Dvoretzky proved [9] that for every $k \in \mathbb{N}$ and $D > 1$ there exists $n = n(k,D) \in \mathbb{N}$ such that every n-dimensional normed

M. M. was partially supported by ISF grant no. 221/07, BSF grant no. 2006009, and a gift from Cisco Research Center. A. N. was partially supported by NSF grant CCF-0832795, BSF grant 2006009, and the Packard Foundation. Part of this work was completed when A. N. was visiting the Discrete Analysis program at the Isaac Newton Institute for Mathematical Sciences.

space has a k-dimensional linear subspace that embeds into Hilbert space with distortion D; see [17, 16, 20] for the best known bounds on $n(k, D)$.

Bourgain, Figiel and Milman [5] studied the following problem as a natural nonlinear variant of Dvoretzky's theorem: given $n \in \mathbb{N}$ and $D > 1$, what is the largest $m \in \mathbb{N}$ such that *any* finite metric space (X, d) of cardinality n has a subset $S \subseteq X$ with $|S| \geqslant m$ such that the metric space (S, d) embeds with distortion D into Hilbert space? Denote this value of m by $R(n, D)$. Bourgain-Figiel-Milman proved [5] that for all $D > 1$ there exists $c(D) \in (0, \infty)$ such that $R(n, D) \geqslant c(D) \log n$, and that $R(n, 1.023) = O(\log n)$. Following several investigations [12, 4, 1] that were motivated by algorithmic applications, a more complete description of the Bourgain-Figiel-Milman phenomenon was obtained in [2].

Theorem 1.1 ([2])**.** *For $D \in (1, \infty)$ there exist $c(D), c'(D) \in (0, \infty)$ and $\delta(D), \delta'(D) \in (0, 1)$ such that for every $n \in \mathbb{N}$,*

- *if $D \in (1, 2)$ then $c(D) \log n \leqslant R(n, D) \leqslant c'(D) \log n$,*
- *if $D \in (2, \infty)$ then $n^{1-\delta(D)} \leqslant R(n, D) \leqslant n^{1-\delta'(D)}$.*

Highlighting the case of large D, which is most relevant for applications, we have the following theorem.

Theorem 1.2 ([2, 15, 19])**.** *For every $\varepsilon \in (0, 1)$ and $n \in \mathbb{N}$, any n-point metric space (X, d) has a subset $S \subseteq X$ with $|S| \geqslant n^{1-\varepsilon}$ that embeds into an ultrametric space with distortion $2e/\varepsilon$. On the other hand, there exists a universal constant $c > 0$ with the following property. For every $n \in \mathbb{N}$ there is an n-point metric space X_n such that for every $\varepsilon \in (0, 1)$ all subsets $Y \subseteq X$ with $|Y| \geqslant n^{1-\varepsilon}$ incur distortion at least c/ε in any embedding into Hilbert space.*

Recall that a metric space (U, ρ) is called an ultrametric space if for every $x, y, z \in U$ we have $\rho(x, y) \leqslant \max\{\rho(x, z), \rho(z, y)\}$. Any separable ultrametric space admits an isometric embedding into Hilbert space [23]. Hence the subset S from Theorem 1.2 also embeds with the stated distortion into Hilbert space, and therefore Theorem 1.2 fits into the Bourgain-Figiel-Milman framework. Note, however, that the stronger statement that S embeds into an ultrametric space is needed for the applications in [2, 15], and that the matching lower bound in Theorem 1.2 is for the weaker requirement of embeddability into Hilbert space. The existence of the metric spaces $\{X_n\}_{n=1}^{\infty}$ from Theorem 1.2 was established in [2]. The estimate $2e/\varepsilon$ on the Euclidean distortion of the subset S from Theorem 1.2 is due to [19], improving by a constant factor over the bound from [15], which itself improves (in an asymptotically optimal way) on the distortion bound of $O\left(\varepsilon^{-1} \log(2/\varepsilon)\right)$ from [2].

In what follows, $\dim_H(X)$ denotes the Hausdorff dimension of a metric space X. Inspired by the above theorems, Terence Tao proposed (unpublished, 2006) another natural variant of the nonlinear Dvoretzky problem: one can keep the statement of Dvoretzky's theorem unchanged in the context of general metric spaces, while interpreting the notion of dimension in the appropriate category. Thus one arrives at the following question.

Question 1.3 (Tao's nonlinear Dvoretzky problem)**.** Given $\alpha > 0$ and $D > 1$, what is the supremum over those $\beta \geqslant 0$ with the following property. Every compact metric space X with $\dim_H(X) \geqslant \alpha$ has a subset $S \subseteq X$ with $\dim_H(S) \geqslant \beta$ that embeds into Hilbert space with distortion D?

The restriction of Question 1.3 to compact metric spaces is not severe. For example, if X is complete and separable then one can first pass to a compact subset of X with the same Hausdorff dimension, and even the completeness of X can be replaced by weaker assumptions; see [7, 11]. We will not address this issue here and restrict our discussion to compact metric spaces, where the crucial subtleties of the problem are already present.

Our purpose here is to provide answers to Question 1.3 in various regimes, the main result being the following theorem.

Theorem 1.4. *There exists a universal constant $C \in (0,\infty)$ such that for every $\varepsilon \in (0,1)$, every compact metric space X has a closed subset $S \subseteq X$ with $\dim_H(S) \geqslant (1-\varepsilon)\dim_H(X)$ that embeds with distortion C/ε into an ultrametric space. In the reverse direction, there is a universal constant $c > 0$ such that for every $\alpha > 0$ there exists a compact metric space X_α with $\dim_H(X_\alpha) = \alpha$ such that if $S \subseteq X$ satisfies $\dim_H(S) \geqslant (1-\varepsilon)\alpha$ then S incurs distortion at least c/ε in any embedding into Hilbert space.*

The construction of the spaces X_α from Theorem 1.4 builds on the examples of [2], which are based on expander graphs. The limiting spaces X_α obtained this way can therefore be called "expander fractals"; their construction is discussed in Section 7. Our main new contribution leading to Theorem 1.4 is the following structural result for general metric measure spaces.

In what follows, by a metric measure space (X,d,μ) we mean a compact metric space (X,d), equipped with a Borel measure μ such that $\mu(X) < \infty$. For $r > 0$ and $x \in X$, the corresponding closed ball is denoted $B(x,r) = \{y \in X : d(x,y) \leqslant r\}$.

Theorem 1.5. *For every $\varepsilon \in (0,1)$ there exists $c_\varepsilon \in (0,\infty)$ with the following property. Every metric measure space (X,d,μ) has a closed subset $S \subseteq X$ such that (S,d) embeds into an ultrametric space with distortion $9/\varepsilon$, and for every $\{x_i\}_{i\in I} \subseteq X$ and $\{r_i\}_{i\in I} \subseteq [0,\infty)$ such that the balls $\{B(x_i,r_i)\}_{i\in I}$ cover S, i.e.,*

$$\bigcup_{i\in I} B(x_i,r_i) \supseteq S, \tag{2}$$

we have

$$\sum_{i\in I} \mu(B(x_i,c_\varepsilon r_i))^{1-\varepsilon} \geqslant \mu(X)^{1-\varepsilon}. \tag{3}$$

Theorem 1.5 contains Theorem 1.2 as a simple special case. Indeed, consider the case when X is finite, say $|X| = n$, the measure μ is the counting measure, i.e., $\mu(A) = |A|$ for all $A \subseteq X$, and all the radii $\{r_i\}_{i\in I}$ vanish. In this case $B(x_i,r_i) = B(x_i,c_\varepsilon r_i) = \{x_i\}$, and therefore the covering condition (2) implies that $\{x_i\}_{i\in I} \supseteq S$. Inequality (3) therefore implies that $|S| \geqslant n^{1-\varepsilon}$, which is the (asymptotically sharp) conclusion of Theorem 1.2, up to a constant multiplicative factor in the distortion.

Theorem 1.5 also implies Theorem 1.4. To see this assume that (X,d) is a compact metric space and $\dim_H(X) > \alpha$. The Frostman lemma (see [11] and [13, Ch. 8]) implies that there exists a constant $K \in (0,\infty)$ and a Borel measure μ such that $\mu(X) > 0$ and $\mu(B(x,r)) \leqslant Kr^\alpha$ for all $r > 0$ and $x \in X$. An application of Theorem 1.5 to the metric measure space (X,d,μ) yields a closed subset $S \subseteq X$ that embeds into an ultrametric space with distortion $O(1/\varepsilon)$ and satisfies the covering condition (3). Thus, all the covers of S by balls $\{B(x_i,r_i)\}_{i\in I}$ satisfy

$$\mu(X)^{1-\varepsilon} \leqslant \sum_{i\in I} \mu(B(x_i,c_\varepsilon r_i))^{1-\varepsilon} \leqslant \sum_{i\in I} \left(Kc_\varepsilon^\alpha r_i^\alpha\right)^{1-\varepsilon}.$$

Hence,

$$\sum_{i\in I} r_i^{(1-\varepsilon)\alpha} \geqslant \frac{\mu(X)^{1-\varepsilon}}{K^{1-\varepsilon}c_\varepsilon^{(1-\varepsilon)\alpha}}.$$

This means that the $(1-\varepsilon)\alpha$-Hausdorff content[1] of S satisfies

$$\mathcal{H}_\infty^{(1-\varepsilon)\alpha}(S) \geqslant \frac{\mu(X)^{1-\varepsilon}}{K^{1-\varepsilon}c_\varepsilon^{(1-\varepsilon)\alpha}} > 0,$$

and therefore $\dim_H(S) = \inf\left\{\beta \geqslant 0 :\ \mathcal{H}_\infty^{\beta}(S) = 0\right\} \geqslant (1-\varepsilon)\alpha$, as asserted in Theorem 1.4.

To summarize the above discussion, the general structural result for metric measure spaces that is contained in Theorem 1.5 implies the sharp Bourgain-Figiel-Milman style nonlinear Dvoretzky theorem when applied to trivial covers of S by singletons. Tao's nonlinear Dvoretzky problem is more subtle since one has to argue about all possible covers of S, and this is achieved by applying Theorem 1.5 to the metric measure space induced by a Frostman measure. In both of these applications the value of the constant c_ε in Theorem 1.5 is irrelevant, but we anticipate that it will play a role in future applications of Theorem 1.5. Our argument yields the bound $c_\varepsilon = e^{O(1/\varepsilon^2)}$, but we have no reason to believe that this dependence on ε is optimal. We therefore pose the following natural problem.

Question 1.6. What is the asymptotic behavior as $\varepsilon \to 0$ of the best possible constant c_ε in Theorem 1.5?

1.1. **An overview of the proof of Theorem 1.5.** Theorem 1.1 was proved in [2] via a deterministic iterative construction of a sufficiently large almost ultrametric subset S of a given finite metric space (X,d). In contrast, Theorem 1.2 was proved in [15] via a significantly shorter probabilistic argument. It is shown in [15] how to specify a distribution (depending on the geometry of X) over *random* subsets $S \subseteq X$ that embed into an ultrametric space with small distortion, yet their expected cardinality is large. The lower bound on the expected cardinality of S is obtained via a lower bound on the probability $\Pr[x \in S]$ for each fixed $x \in X$. Such a probabilistic estimate seems to be quite special, and we do not see how to argue probabilistically about all possible covers of a random subset S, as required in Theorem 1.5. For this reason our proof of Theorem 1.5 is a deterministic construction which uses in some of its steps adaptations of the methods of [2], in addition to a variety of new ingredients that are needed in order to handle a covering condition such as (3). Actually, in order to obtain the sharp $O(1/\varepsilon)$ distortion bound of Theorem 1.5 we also use results of [15, 19] (see Theorem 6.2 below), so in fact Theorem 1.5 is based on a combination of deterministic and probabilistic methods, the deterministic steps being our most substantial new contribution.

The proof of Theorem 1.5 starts with a reduction of the problem to the case of finite metric spaces; see Section 3. Once this is achieved, the argument is a mixture of combinatorial, analytic and geometric arguments, the key objects of interest being *fragmentation maps*. These are maps that are defined on rooted combinatorial trees and assign to each vertex of the tree a subset of the metric space (X,d) in a way that respects the tree structure, i.e., the set corresponding to an offspring of a vertex is a subset of the set corresponding to the

[1] Recall that for $\beta \geqslant 0$ the β-Hausdorff content of a metric space (Z,d) is defined to be the infimum of $\sum_{j\in J} r_j^\beta$ over all possible covers of Z by balls $\{B(z_j,r_j)\}_{j\in J}$; see [13].

vertex itself, and vertices lying on distinct root-leaf paths are assigned to disjoint subsets of X. We also require that leaves are mapped to singletons.

Each fragmentation map corresponds to a subset of X (the images of the leaves), and our goal is to produce a fragmentation map that corresponds to a subset of X which satisfies the conclusion of Theorem 1.5. To this end, we initiate the iteration via a bottom-up construction of a special fragmentation map; see Section 4. We then proceed to iteratively "prune" (or "sparsify") this initial tree so as to produce a smaller tree whose leaves satisfy the conclusion of Theorem 1.5. At each step we argue that there must exist sufficiently many good pruning locations so that by a pigeonhole argument we can make the successive pruning locations align appropriately; see Section 5 and Section 6. The exact details of these steps are quite involved, and it is impossible to describe them without introducing a significant amount of notation and terminology, and specifying rather complicated inductive hypotheses. We therefore refer to the relevant sections for a detailed description.

1.2. **The low distortion regime.** We have thus far focused on Question 1.3 in the case of high distortion embeddings into ultrametric spaces. There is also a Hausdorff dimensional variant of the phase transition at distortion 2 that was described in Theorem 1.1.

Theorem 1.7 (Distortion $2+\delta$)**.** *There exists a universal constant $c \in (0,\infty)$ such that for every $\delta \in (0,1/2)$, any compact metric space (X,d) has a closed subset $S \subseteq X$ that embeds with distortion $2+\delta$ in an ultrametric space, and*

$$\dim_H(S) \geqslant \frac{c\delta}{\log(1/\delta)} \dim_H(X). \tag{4}$$

For distortion strictly less than 2 the following theorem shows that there is no nonlinear Dvoretzky phenomenon in terms of Hausdorff dimension.

Theorem 1.8. *For every $\alpha > 0$ there exists a compact metric space (X,d) of Hausdorff dimension α, such that if $S \subseteq X$ embeds into Hilbert space with distortion strictly smaller than 2 then $\dim_H(S) = 0$.*

As in the case of the finite nonlinear Dvoretzky theorem, Question 1.3 at distortion 2 remains open. In the same vein, the correct asymptotic dependence on δ in (4) is unknown.

Theorem 1.7 follows from the following result in the spirit of Theorem 1.5, via the same Frostman measure argument.

Theorem 1.9. *There exists a universal constant $c \in (0,\infty)$ such that for every $\delta \in (0,1/2)$ there exists $c'_\delta \in (0,\infty)$ with the following property. Every metric measure space (X,d,μ) has a closed subset $S \subseteq X$ such that (S,d) embeds into an ultrametric space with distortion $2+\delta$, and for every $\{x_i\}_{i\in I} \subseteq X$ and $\{r_i\}_{i\in I} \subseteq [0,\infty)$ such that the balls $\{B(x_i,r_i)\}_{i\in I}$ cover Y, we have*

$$\sum_{i\in I} \mu(B(x_i,c'_\delta r_i))^{\frac{c\delta}{\log(1/\delta)}} \geqslant \mu(X)^{\frac{c\delta}{\log(1/\delta)}}. \tag{5}$$

1.3. **Talagrand's nonlinear Dvoretzky theorem.** Despite the fact that Theorem 1.5 satisfactorily encompasses both the Bourgain-Figiel-Milman nonlinear Dvoretzky phenomenon and Tao's nonlinear Dvoretzky phenomenon, there is an important nonlinear Dvoretzky theorem due to Talagrand that does not seem to fall into this framework. To describe it we recall the definition of Talagrand's γ_2 functional, which is defined for every compact metric space (X,d) as follows.

$$\gamma_2(X,d) = \inf_{\substack{\{A_s\}_{s=0}^\infty \subseteq 2^X \\ \forall s\geqslant 0,\ |A_s|\leqslant 2^{2^s}}} \sup_{x\in X} \sum_{s=0}^{\infty} 2^{s/2} d(x,A_s).$$

In 1987 Talagrand proved [21] the following nonlinear Dvoretzky theorem, where the notion of dimension of (X,d) is interpreted to be $\gamma_2(X,d)$.

Theorem 1.10 ([21])**.** *There are universal constants $c, D \in (0,\infty)$ such that every finite metric space (X,d) has a subset $S \subseteq X$ that embeds into an ultrametric space with distortion D and $\gamma_2(S,d) \geqslant c\gamma_2(X,d)$.*

Theorem 1.10 is stated slightly differently in [21], but it easily follows from a combination of [21, Lem. 6], [21, Thm. 11], and [21, Prop. 13]. Also, in [21] a different definition of γ_2 is used, but it is equivalent to the definition described above, as explained in [22].

Talagrand's nonlinear Dvoretzky theorem is of major importance since it easily implies the majorizing measures theorem. Specifically, assume that $\{G_x\}_{x\in X}$ is a Gaussian process and for $x, y \in X$ we have $d(x,y) = \sqrt{\mathbb{E}[(G_x - G_y)^2]}$. Talagrand's majorizing measures theorem asserts that $\mathbb{E}[\sup_{x\in X} G_x] \geqslant K\gamma_2(X,d)$, where $K \in (0,\infty)$ is a universal constant. Let $S \subseteq X$ be the subset obtained from an application of Theorem 1.10 to (X,d). Since ultrametric spaces are isometric to subsets of Hilbert space, there is a Gaussian process $\{H_x\}_{x\in S}$ such that $\rho(x,y) = \sqrt{\mathbb{E}[(H_x - H_y)^2]}$ is an ultrametric on S and $d(x,y) \leqslant \rho(x,y) \leqslant Dd(x,y)$ for all $x,y \in S$. Hence $\gamma_2(S,\rho) \geqslant \gamma_2(S,d) \geqslant c\gamma_2(X,d)$, and a standard application of Slepian's lemma (see [21, Prop. 5]) yields $\mathbb{E}[\sup_{x\in X} G_x] \geqslant \mathbb{E}[\sup_{x\in S} G_x] \geqslant D^{-1}\mathbb{E}[\sup_{x\in S} H_x]$. This shows that due to Theorem 1.10 it suffices to prove the majorizing measures theorem when (X,d) is an ultrametric space itself. Ultrametric spaces have a natural tree structure (based on nested partitions into balls; see e.g. [2, Sec. 3.1]), and they can be embedded into Hilbert space so that disjoint subtrees are orthogonal. For Gaussian processes orthogonality means independence, which indicates how the ultrametric structure can be harnessed to yield a direct and short proof of the majorizing measures theorem for ultrametrics; see [21, Prop. 13]. This striking application of a metric Dvoretzky-type theorem is of great importance to several areas; we refer to [22] for an exposition of some of its many applications.

We do not know how to use our approach to Theorem 1.5 to deduce Theorem 1.10, nor do we see a way to use Talagrand's methods to prove Theorem 1.5. The nonlinear Dvoretzky theorems that are currently known, including variants of Theorem 1.10 for other functionals that are defined similarly to γ_2, differ from each other in the notion of "dimension", or "largeness", of a metric space that they use. It is natural to ask if there is a common theme to these results, perhaps even a general nonlinear Dvoretzky theorem that contains the Bourgain-Figiel-Milman, Talagrand and Tao phenomena as special cases. We leave this as an open-ended and somewhat vague direction for future research. Additional discussion along these lines is deferred to our forthcoming survey [14].

Question 1.11. What are the notions of "dimension" of metric spaces that yield a nonlinear Dvoretzky theorem in the sense that every (compact) metric space can be shown to contain a subset of proportional dimension that well-embeds into an ultrametric space? At present we know this for the following notions of dimension: $\log|X|$, $\dim_H(X)$, $\gamma_2(X)$ (and some natural variants of these notions). Is there an overarching principle here?

2. Preliminaries and notation

In this section we describe the basic concepts and definitions that will be used throughout the ensuing sections. As we have already noted earlier, given a metric space (X,d) and $r>0$, the closed ball centered at $x\in X$ of radius r is denoted $B(x,r)=\{y\in X:\ d(x,y)\leqslant r\}$. Open balls are denoted by $B^\circ(x,r)=\{y\in X:\ d(x,y)<r\}$. We shall use this notation whenever the metric space in question will be clear from the context of the discussion, but when we will need to argue about several metrics at once we will add a subscript indicating the metric with respect to which balls are taken. Thus, we will sometimes use the notation $B_d(x,r), B_d^\circ(x,r)$. Similar conventions hold for diameters of subsets of X: given a nonempty $A\subseteq X$ we denote $\mathrm{diam}(A)=\sup_{x,y\in A} d(x,y)$ whenever the underlying metric is clear from the context, and otherwise we denote this quantity by $\mathrm{diam}_d(A)$.

Given two nonempty subsets $A,B\subseteq X$ we denote as usual

$$d(A,B)=\inf_{\substack{x\in A\\ y\in B}} d(x,y), \tag{6}$$

and we will also use the standard notation $d(x,A)=d(\{x\},A)$. The Hausdorff distance between A and B is denoted

$$d_H(A,B)=\max\left\{\sup_{x\in A} d(x,B),\sup_{y\in B} d(y,A)\right\}. \tag{7}$$

2.1. **Combinatorial trees.** The ensuing arguments rely on a variety of constructions involving combinatorial trees. We will work only with finite rooted trees, i.e., finite graph-theoretical trees T with a distinguished vertex $r(T)$ called the root of T. We will slightly abuse notation by identifying T with its vertex set, i.e., when we write $v\in T$ we mean that v is a vertex of T. We shall say that $u\in T$ is an ancestor of $v\in T\smallsetminus\{u\}$ if u lies on the path joining v and $r(T)$. In this case we also say that v is a descendant of u. We say that v is a weak descendant (respectively weak ancestor) of u if it is either a descendant (respectively ancestor) of u or $v=u$. If u is either a weak ancestor of v or a weak descendant of v we say that u and v are comparable, and otherwise we say that they are incomparable. The leaves of T, denoted $\mathcal{L}(T)\subseteq T$, is the set of vertices of T that do not have descendants.

Definition 2.1 (Cut set)**.** Let T be a rooted tree. A subset $S\subseteq T$ is called a cut set of T if any root-leaf path in T intersects S. Equivalently, S is a cut set of T if every $u\in T$ is comparable to a vertex in S. See [18, Ch. 4 & Sec. 12.4].

If $v\in T\smallsetminus\{r(T)\}$ then we denote by $\mathbf{p}(v)=\mathbf{p}_T(v)$ its parent in T, i.e., the vertex adjacent to v on the path joining v and $r(T)$. We say that $v\in T\smallsetminus\{r(t)\}$ is a child of $u\in T$ if $\mathbf{p}(v)=u$, and the set $\mathbf{p}^{-1}(u)=\{v\in T:\ \mathbf{p}(v)=u\}$ is the set of children of u. Thus $\mathcal{L}(T)=\{u\in T:\ \mathbf{p}^{-1}(u)=\emptyset\}$. If $u,v\in T\smallsetminus\{r(T)\}$ are distinct and satisfy $\mathbf{p}(u)=\mathbf{p}(v)$ then we say that u and v are siblings in T.

The depth of $u\in T$, denoted $\mathrm{depth}_T(u)$, is the number of edges on the path joining u and $r(T)$. Thus $\mathrm{depth}_T(r(T))=0$. The least common ancestor of $u,v\in T$, denoted $\mathrm{lca}(u,v)=\mathrm{lca}_T(u,v)$, is the vertex of maximal depth that is an ancestor of both u and v.

Given $u\in T$, we denote by $T_u\subseteq T$ the subtree rooted at u, i.e., the tree consisting of all the weak descendants of u in T, with the edges inherited from T. Thus $r(T_u)=u$.

Definition 2.2. Let T be a rooted tree and $A \subseteq T$. For $u \in T$ define $D_T(u, A) \subseteq T$ to be the set of all $v \in A$ such that v is a descendant of u and no ancestor of v is also in A and is a descendant of u. Note that $D_T(u, A) = \emptyset$ if u has no descendants in A, and $D_T(u, A)$ is a cut set of the subtree T_u if $A \cap T_u$ is a cut-set in T_u (this happens in particular if A contains the leaves of T_u). We also define

$$D_T^*(u, A) = \begin{cases} D_T(u, A) & \text{if } u \in T \smallsetminus A, \\ \{u\} & \text{if } u \in A. \end{cases} \tag{8}$$

2.2. **Fragmentation maps.** Trees interact with metric spaces via the notion of fragmentation maps.

Definition 2.3 (Fragmentation map)**.** Let (X, d) be a finite metric space. A fragmentation map of X is a function $\mathcal{F} : T \to 2^X$, where T is a finite rooted tree, that satisfies the following conditions.

- $\mathcal{F}(r(T)) = X$.
- If $v \in \mathcal{L}(T)$ is a leaf of T then $\mathcal{F}(v)$ is a singleton, i.e., $\mathcal{F}(v) = \{x\}$ for some $x \in X$.
- If $v \in T \smallsetminus \{r(T)\}$ the $\mathcal{F}(v) \subseteq \mathcal{F}(\mathbf{p}(v))$.
- If $u, v \in T$ are incomparable then $\mathcal{F}(u) \cap \mathcal{F}(v) = \emptyset$.

In what follows, given a fragmentation map $\mathcal{F} : T \to 2^X$, we will use the notation $\mathcal{F}_u = \mathcal{F}(u)$.

Definition 2.4 (Boundary of a fragmentation map)**.** The boundary of a fragmentation map $\mathcal{F} : T \to 2^X$ is a new map $\partial\mathcal{F} : T \to 2^X$ defined as follows. For $u \in T$ the set $\partial\mathcal{F}(u) = \partial\mathcal{F}_u$ is the subset of X corresponding to the image under $\mathcal{F}$ of the leaves of the subtree T_u, i.e.,

$$\partial\mathcal{F}_u = \bigcup_{v \in \mathcal{L}(T_u)} \mathcal{F}_v.$$

Note that we always have $\partial\mathcal{F}_u \subseteq \mathcal{F}_u$.

Definition 2.5 (Partition map)**.** A partition map is a fragmentation map $\mathcal{F} : T \to 2^X$ such that $\partial F_{r(T)} = X$.

Up to this point the metric on X did not play any role. The following two crucial definitions measure the interaction between partition maps of X and the geometry of X.

Definition 2.6 (Lacunary fragmentation map)**.** Given $K, \gamma \in (0, \infty)$, a fragmentation map $\mathcal{F} : T \to 2^X$ is (K, γ)-lacunary if for every $q \in T$ and every $u \in T$ such that u is a weak descendant of q and u has at least two children, i.e., $|\mathbf{p}^{-1}(u)| > 1$, we have

$$\operatorname{diam}(\mathcal{F}_q) \leqslant K\gamma^{\operatorname{depth}_T(u) - \operatorname{depth}_T(q)} \cdot \min_{\substack{v, w \in \mathbf{p}^{-1}(u) \\ v \neq w}} d(\partial\mathcal{F}_v, \partial\mathcal{F}_w). \tag{9}$$

Definition 2.7 (Separated fragmentation map)**.** Given $\beta > 0$ and a fragmentation map $\mathcal{F} : T \to 2^X$, a vertex $u \in T$ is called β-separated if for every $x \in \left(\partial\mathcal{F}_{r(T)}\right) \smallsetminus (\partial\mathcal{F}_u)$ we have

$$d(x, \mathcal{F}_u) \geqslant \beta \cdot \operatorname{diam}(\mathcal{F}_u). \tag{10}$$

The map $\mathcal{F} : T \to 2^X$ is called β-separated if all the vertices $u \in T$ are β-separated.

Lemma 2.8. *Let $\mathcal{F} : T \to X$ be a (K, γ)-lacunary fragmentation map of a metric space (X, d). Then $(\partial\mathcal{F}_{r(T)}, d)$ embeds with distortion K into an ultrametric space.*

Proof. For $x, y \in \partial \mathcal{F}_{r(T)}$ there are $a, b \in \mathcal{L}(T)$ such that $\mathcal{F}_a = \{x\}$ and $\mathcal{F}_b = \{y\}$. Define

$$\rho(x,y) = \operatorname{diam}\left(\mathcal{F}_{\mathrm{lca}(a,b)}\right).$$

Since $x, y \in \mathcal{F}_{\mathrm{lca}(a,b)}$ we have $d(x,y) \leqslant \rho(x,y)$. Assume that $x \neq y$. Then $\mathrm{lca}(a,b)$ has distinct children $v, w \in \mathbf{p}^{-1}(\mathrm{lca}(a,b))$ such that $x \in \partial \mathcal{F}_v$ and $y \in \partial \mathcal{F}_w$. An application of (9) to $q = u = \mathrm{lca}(a,b)$ shows that $\rho(x,y) \leqslant K d(x,y)$. It remains to note that ρ is an ultrametric. Indeed, take $a, b, c \in \mathcal{L}(T)$ and write $\mathcal{F}_a = \{x\}$, $\mathcal{F}_b = \{y\}$, $\mathcal{F}_c = \{w\}$. If $\mathrm{lca}(a,b)$ is a weak descendant of $\mathrm{lca}(b,c)$ then $\mathcal{F}_{\mathrm{lca}(a,b)} \subseteq \mathcal{F}_{\mathrm{lca}(b,c)}$, implying that $\rho(x,y) \leqslant \rho(y,z)$. Otherwise $\mathrm{lca}(a,b) \in \{\mathrm{lca}(a,c), \mathrm{lca}(b,c)\}$, implying that $\rho(x,y) = \max\{\rho(x,z), \rho(y,z)\}$. □

3. Reduction to finite metric spaces

The following simple compactness lemma shows that it suffices to prove Theorem 1.5 and Theorem 1.9 when (X,d) is a finite metric space.

Lemma 3.1. *Fix $D, c \geqslant 1$ and $\theta \in (0,1]$. Suppose that any finite metric measure space (X,d,μ) has a subset $S \subseteq X$ that embeds with distortion D into an ultrametric space, such that every family of balls $\{B(x_i,r_i)\}_{i \in I}$ that covers S satisfies*

$$\sum_{i \in I} \mu\left(B(x_i, c r_i)\right)^{\theta} \geqslant \mu(X)^{\theta}. \tag{11}$$

Then any metric measure space (X,d,μ) has a closed subset $S \subseteq X$ that embeds with distortion D into an ultrametric space, such that every family of balls $\{B(x_i,r_i)\}_{i \in I}$ that covers S satisfies (11).

Proof. Let (X,d,μ) be a metric measure space and let X_n be a $\frac{1}{n}$-net in X, i.e., $d(x,y) > \frac{1}{n}$ for all distinct $x, y \in X_n$, and $d(x, X_n) \leqslant \frac{1}{n}$ for all $x \in X$. Since X is compact, X_n is finite. Write $X_n = \{x_1^n, x_2^n, \ldots, x_{k_n}^n\}$ and for $j \in \{1, \ldots, k_n\}$ define

$$j_n(x) \stackrel{\mathrm{def}}{=} \min\left\{i \in \{1, \ldots, k_n\} :\ d\left(x, x_i^n\right) = d(x, X_n)\right\}.$$

Consider the Voronoi tessellation $\{V_1^n, \ldots, V_{k_n}^n\} \subseteq 2^X$ given by

$$V_j^n \stackrel{\mathrm{def}}{=} \left\{x \in X :\ j_n(x) = j\right\}.$$

Thus $\{V_1^n, \ldots, V_{k_n}^n\}$ is a Borel partition of X, and we can define a measure μ_n on X_n by $\mu_n(x_j^n) = \mu(V_j)$. Note that by definition $\mu_n(X_n) = \mu(X)$.

The assumption of Lemma 3.1 applied to the finite metric measure space (X_n, d, μ_n) yields a subset $S_n \subseteq X_n$, an ultrametric space (U_n, ρ_n) and a mapping $f_n : S_n \to U_n$ such that $d(x,y) \leqslant \rho_n(f_n(x), f_n(y)) \leqslant D d(x,y)$ for all $x, y \in S_n$. Moreover, if $\{z_j\}_{j \in J} \subseteq X_n$ and $\{r_j\}_{j \in J} \subseteq [0,\infty)$ satisfy $\bigcup_{j \in J} B(z_j, r_j) \supseteq S_n$ then

$$\sum_{j \in J} \mu_n\left(X_n \cap B(z_j, c r_j)\right)^{\theta} \geqslant \mu_n(X_n)^{\theta} = \mu(X)^{\theta}. \tag{12}$$

Let $\mathscr{U}$ be a free ultrafilter over $\mathbb{N}$. Since the Hausdorff metric d_H (recall (7)) on the space of closed subsets of X is compact (e.g. [6, Thm. 7.3.8]), there exists a closed subset $S \subseteq X$ such that $\lim_{n \to \mathscr{U}} d_H(S, S_n) = 0$.

Define $\rho : S \times S \to [0,\infty)$ as follows. For $x, y \in S$ there are $(x_n)_{n=1}^{\infty}, (y_n)_{n=1}^{\infty} \in \prod_{n=1}^{\infty} S_n$ such that $\lim_{n \to \mathscr{U}} d(x, x_n) = \lim_{n \to \mathscr{U}} d(y, y_n) = 0$. Set $\rho(x,y) = \lim_{n \to \mathscr{U}} \rho_n(f_n(x_n), f_n(y_n))$.

This is well defined, i.e., $\rho(x,y)$ does not depend on the choice of $(x_n)_{n=1}^\infty,(y_n)_{n=1}^\infty$. Indeed, if $(x_n')_{n=1}^\infty,(y_n')_{n=1}^\infty\in\prod_{n=1}^\infty S_n$ also satisfy $\lim_{n\to\mathscr{U}} d(x,x_n')=\lim_{n\to\mathscr{U}} d(y,y_n')=0$ then

$$\begin{aligned}
&\lim_{n\to\mathscr{U}}\rho_n(f_n(x_n),f_n(y_n))\\
&\qquad\leqslant \lim_{n\to\mathscr{U}}\rho_n(f_n(x_n'),f_n(y_n'))+\lim_{n\to\mathscr{U}}\rho_n(f_n(x_n),f_n(x_n'))+\lim_{n\to\mathscr{U}}\rho_n(f_n(y_n),f_n(y_n'))\\
&\qquad\leqslant \lim_{n\to\mathscr{U}}\rho_n(f_n(x_n'),f_n(y_n'))+D\lim_{n\to\mathscr{U}}d(x_n,x_n')+D\lim_{n\to\mathscr{U}}d(y_n,y_n')\\
&\qquad= \lim_{n\to\mathscr{U}}\rho_n(f_n(x_n'),f_n(y_n')),
\end{aligned}$$

so that by symmetry $\lim_{n\to\mathscr{U}}\rho_n(f_n(x_n),f_n(y_n))=\lim_{n\to\mathscr{U}}\rho_n(f_n(x_n'),f_n(y_n'))$. It is immediate to check that $d(x,y)\leqslant\rho(x,y)\leqslant Dd(x,y)$ for all $x,y\in S$ and that ρ is an ultrametric on S.

Now, let $\{x_i\}_{i=1}^\infty\subseteq X$ and $\{r_i\}_{i=1}^\infty\subseteq[0,\infty)$ satisfy $\bigcup_{i=1}^\infty B(x_i,r_i)\supseteq S$. Fix $\eta>0$. For every $i\in\mathbb{N}$ there is $\varepsilon_i\in(0,1)$ such that

$$\mu\left(B(x_i,cr_i+c\varepsilon_i)\right)\leqslant\mu\left(B(x_i,cr_i)\right)+\left(\frac{\eta}{2^i}\right)^{1/\theta}. \tag{13}$$

Since S is compact, there exists a finite subset $I_\eta\subseteq\mathbb{N}$ such that $\bigcup_{i\in I_\eta}B(x_i,r_i+\varepsilon_i/2)\supseteq S$. Denote $\varepsilon=\min_{i\in I_\eta}\varepsilon_i$. By definition of S there exists $n\in\mathbb{N}$ such that $n>8/\varepsilon$ and $d_H(S_n,S)<\varepsilon/8$. For every $i\in I_\eta$ let $z_i^n\in X_n$ satisfy $d(z_i^n,x_i)=d(x_i,X_n)\leqslant 1/n<\varepsilon/8$. Now, $\bigcup_{i\in I_\eta}B(z_i^n,r_i+3\varepsilon_i/4)\supseteq S_n$ because $d_H(S_n,S)<\varepsilon/8$ and $\bigcup_{i\in I_\eta}B(x_i,r_i+\varepsilon_i/2)\supseteq S$. An application of (12) now yields the bound

$$\sum_{i\in I_\eta}\mu_n\left(X_n\cap B\left(z_i^n,cr_i+\frac{3c\varepsilon_i}{4}\right)\right)^\theta\geqslant\mu(X)^\theta. \tag{14}$$

By the definition of μ_n,

$$\begin{aligned}
&\mu_n\left(X_n\cap B\left(z_i^n,cr_i+\frac{3c\varepsilon_i}{4}\right)\right)\\
&\qquad=\mu\left(\bigcup\left\{V_j^n:\ j\in\{1,\ldots,k_n\}\ \wedge\ x_j^n\in B\left(z_i^n,cr_i+\frac{3c\varepsilon_i}{4}\right)\right\}\right)\\
&\qquad\leqslant\mu\left(B\left(z_i^n,cr_i+\frac{3c\varepsilon_i}{4}+\frac{1}{n}\right)\right)\\
&\qquad\leqslant\mu\left(B\left(x_i,cr_i+\frac{3c\varepsilon_i}{4}+\frac{2}{n}\right)\right)\\
&\qquad\leqslant\mu\left(B\left(x_i,cr_i+c\varepsilon_i\right)\right)\\
&\qquad\overset{(13)}{\leqslant}\mu\left(B(x_i,cr_i)\right)+\left(\frac{\eta}{2^i}\right)^{1/\theta}.
\end{aligned}\tag{15}$$

Hence,

$$\begin{aligned}
\mu(X)^\theta&\overset{(14)\wedge(15)}{\leqslant}\sum_{i\in I_\eta}\left(\mu\left(B(x_i,cr_i)\right)+\left(\frac{\eta}{2^i}\right)^{1/\theta}\right)^\theta\\
&\leqslant\sum_{i\in I_\eta}\left(\mu\left(B(x_i,cr_i)\right)^\theta+\frac{\eta}{2^i}\right)\leqslant\eta+\sum_{i=1}^\infty\mu\left(B(x_i,cr_i)\right)^\theta.
\end{aligned}\tag{16}$$

Since (16) holds for all $\eta > 0$, the proof of Lemma 3.1 is complete. $\square$

4. The initial fragmentation map

Due to Lemma 3.1 we may assume from now on that (X, d, μ) is a finite metric measure space. By restricting to the support of μ, we assume throughout that $\mu(\{x\}) > 0$ for all $x \in X$. By rescaling the metric, assume also that $\operatorname{diam}(X) = 1$. Our proof of Theorem 1.5 in this case will start with an initial fragmentation map $\mathcal{F} : T \to 2^X$ which we will now construct in a bottom-up fashion. Once this is achieved, our goal will be to iteratively modify this fragmentation map so as to gradually arrive at the conclusion of Theorem 1.5.

Observe that any fragmentation map $\mathcal{F} : T \to 2^X$ induces a weighting $w : T \to (0, \infty)$ of the vertices of T given by $w(u) = \mu(\mathcal{F}_u)$. The following definition describes a way to measure the size of subtrees of a weighted tree. The exact notion of "size" here is tailored to be compatible with the ensuing arguments.

Definition 4.1 (Modified weight function)**.** Fix integers $h, k \geqslant 2$ and let T be a finite graph-theoretical rooted tree, all of whose leaves are at the same depth, which is divisible by h. Assume that we are given $w : T \to (0, \infty)$. Define a new function $w_h^k : T \to (0, \infty)$ as follows. If $u \in \mathcal{L}(T)$ then

$$w_h^k(u) = w(u)^{\frac{k-1}{k}}.$$

Continue defining $w_h^k(u)$ by reverse induction on $\operatorname{depth}_T(u)$ as follows.

$$w_h^k(u) = \begin{cases} w(u)^{\frac{k-1}{k}} & \text{if } h \mid \operatorname{depth}_T(u), \\ \sum_{v \in \mathbf{p}^{-1}(u)} w_h^k(v) & \text{if } h \nmid \operatorname{depth}_T(u). \end{cases} \tag{17}$$

Equivalently, if $u \in T$ and $(j-1)h < \operatorname{depth}_T(u) \leqslant jh$ for some integer j then

$$w_h^k(u) = \sum_{\substack{v \in T_u \\ \operatorname{depth}_T(v) = jh}} w(v)^{\frac{k-1}{k}}. \tag{18}$$

The following lemma shows how to construct an initial fragmentation map of (X, d, μ) which has some desirable properties. In particular, the initial tree contains "pointers" to the "largest" child, where size is measured in terms of the modified weight function w_h^k of Definition 4.1. We will need this information in the ensuing pruning steps, since we will want to avoid discarding (as much as possible) these large "designated children".

Lemma 4.2. *Let (X, d, μ) be a finite metric measure space of diameter 1 and $\tau \in (0, 1/3)$. For every triple of integers $m, h, k \geqslant 2$ there exists a fragmentation map $\mathcal{F} : T \to 2^X$ with the following properties.*

- *All the leaves of the tree T are at depth mh.*
- *$\mathcal{F}$ is a partition map, i.e., $\partial F_{r(T)} = X$.*
- *For every $u \in T$ we have*
$$\operatorname{diam}(\mathcal{F}_u) \leqslant \tau^{\operatorname{depth}_T(u)}. \tag{19}$$
- *Every non-leaf vertex $u \in T \smallsetminus \mathcal{L}(T)$ has a "designated child" $\mathbf{c}(u) \in \mathbf{p}^{-1}(u)$ such that*
$$w_h^k(\mathbf{c}(u)) = \max_{v \in \mathbf{p}^{-1}(u)} w_h^k(v), \tag{20}$$
where $w : T \to (0, \infty)$ is given by $w(u) = \mu(\mathcal{F}_u)$ and $w_h^k : T \to (0, \infty)$ is the modified weight function from Definition 4.1.

- *Suppose that $u, v \in T \smallsetminus \mathcal{L}(T)$ satisfy* $\operatorname{depth}_T(u) = \operatorname{depth}_T(v)$ *and $u \neq v$. Then*

$$d\left(\mathcal{F}_{\mathbf{c}(u)}, \mathcal{F}_{\mathbf{c}(v)}\right) > \frac{1-3\tau}{2} \cdot \tau^{\operatorname{depth}_T(u)}. \tag{21}$$

Proof. The construction will be in a bottom-up fashion: the tree T will be decomposed as a disjoint union on "levels" $V_0, V_1, \ldots, V_{mh}$, where V_i are the vertices at depth i. We will construct these levels V_i and the mappings $\mathcal{F} : V_i \to 2^X$ and $w_h^k : V_i \to (0, \infty)$ by reverse induction on i, and describe inductively for each $v \in V_{i+1}$ its parent $u \in V_i$, as well as the designated child $\mathbf{c}(u)$. At the end of this construction V_0 will consist of a single vertex, the root of T.

Define $\ell_{mh} = |X|$ and write $X = \{x_1, \ldots, x_{\ell_{mh}}\}$. The initial level V_{mh} consists of the leaves of T, and it is defined to be $V_{mh} = \{v_j^{mh}\}_{j=1}^{\ell_{mh}}$. We also set

$$\mathcal{F}_{v_j^{mh}} = \{x_j\} \quad \text{and} \quad w_h^k(v_j^{mh}) = w(x_j)^{\frac{k-1}{k}} = \mu(\{x_j\})^{\frac{k-1}{k}}$$

for all $j \in \{1, \ldots, \ell_{mh}\}$.

Assume inductively that for $i \in \{1, \ldots, mh-1\}$ we have already defined

$$V_{i+1} = \left\{v_1^{i+1}, v_2^{i+1}, \ldots, v_{\ell_{i+1}}^{i+1}\right\},$$

and the mappings $\mathcal{F} : V_{i+1} \to 2^X$ and $w_h^k : V_{i+1} \to (0, \infty)$.

Choose $j(1) \in \{1, \ldots, \ell_{i+1}\}$ such that

$$w_h^k\left(v_{j(1)}^{i+1}\right) = \max_{j \in \{1, \ldots, \ell_{i+1}\}} w_h^k\left(v_j^{i+1}\right).$$

Define

$$A_1^i = \left\{s \in \{1, \ldots, \ell_{i+1}\} :\ d\left(\mathcal{F}_{v_{j(1)}^{i+1}}, \mathcal{F}_{v_s^{i+1}}\right) \leqslant \frac{1-3\tau}{2}\tau^i\right\}.$$

Create a new vertex $v_1^i \in V_i$ and define

$$\mathcal{F}_{v_1^i} = \bigcup_{s \in A_1^i} \mathcal{F}_{v_s^{i+1}}.$$

Also, declare the vertices $\{v_s^{i+1}\}_{s \in A_1^i} \subseteq V_{i+1}$ to be the children of v_1^i, and in accordance with (17) define

$$w_h^k\left(v_1^i\right) = \begin{cases} w\left(v_1^i\right)^{\frac{k-1}{k}} & \text{if } h \mid i, \\ \sum_{s \in A_1^i} w_h^k\left(v_s^{i+1}\right) & \text{if } h \nmid i. \end{cases}$$

Finally, set

$$\mathbf{c}\left(v_1^i\right) = v_{j(1)}^{i+1}.$$

Continuing inductively, assume that we have defined $v_1^i, v_2^i, \ldots, v_r^i \in V_i$, together with nonempty disjoint sets

$$A_1^i, \ldots, A_r^i \subseteq \{1, \ldots, \ell_{i+1}\}.$$

If $\bigcup_{t=1}^r A_t^i = \{1, \ldots, \ell_{i+1}\}$ then define $\ell_i = r$ and $V_i = \{v_1^i, v_2^i, \ldots, v_r^i\}$. Otherwise, choose $j(r+1) \in \{1, \ldots, \ell_{i+1}\} \smallsetminus \bigcup_{t=1}^r A_t^i$ such that

$$w_h^k\left(v_{j(r+1)}^{i+1}\right) = \max_{j \in \{1, \ldots, \ell_{i+1}\} \smallsetminus \bigcup_{t=1}^r A_t^i} w_h^k\left(v_j^{i+1}\right), \tag{22}$$

and define

$$A^i_{r+1}=\left\{s\in\{1,\ldots,\ell_{i+1}\}\setminus\bigcup_{t=1}^{r}A^i_t:\ d\left(\mathcal{F}_{v^{i+1}_{j(r+1)}},\mathcal{F}_{v^{i+1}_s}\right)\leqslant\frac{1-3\tau}{2}\tau^i\right\}. \tag{23}$$

Create a new vertex $v^i_{r+1}\in V_i$ and define

$$\mathcal{F}_{v^i_{r+1}}=\bigcup_{s\in A^i_{r+1}}\mathcal{F}_{v^{i+1}_s}. \tag{24}$$

Also, declare the vertices $\{v^{i+1}_s\}_{s\in A^i_{r+1}}\subseteq V_{i+1}$ to be the children of v^i_{r+1} and define

$$w^k_h\left(v^i_{r+1}\right)=\begin{cases} w\left(v^i_{r+1}\right)^{\frac{k-1}{k}} & \text{if } k\mid i,\\ \sum_{s\in A^i_{r+1}}w^k_h\left(v^{i+1}_s\right) & \text{if } k\nmid i.\end{cases}$$

Finally, set

$$\mathbf{c}\left(v^i_{r+1}\right)=v^{i+1}_{j(r+1)}. \tag{25}$$

The above recursive procedure must terminate, yielding the level i set V_i. We then proceed inductively until the set V_1 has been defined. We conclude by defining V_0 to be a single new vertex $r(T)$ (the root) with all the vertices in V_1 its children. The designated child of the root, $\mathbf{c}(r(T))$, is chosen to be a vertex $u\in V_1$ such that $w^k_h(u)=\max_{v\in V_1}w^k_h(v)$. We also set $\mathcal{F}_{r(T)}=X$ and $w^k_h(r(T))=\mu(X)^{\frac{k-1}{k}}$.

The resulting fragmentation map $\mathcal{F}:T\to 2^X$ is by definition a partition map, since $\mathcal{F}(V_{mh})=\mathcal{F}(\mathcal{L}(T))=X$. Also, the construction above guarantees the validity of (20) due to (22) and (25).

We shall now prove (19) by reverse induction on $\mathrm{depth}_T(u)$. If $\mathrm{depth}_T(u)=mh$ then $\mathrm{diam}(\mathcal{F}_u)=0$ and there is nothing to prove. Assuming the validity of (19) whenever $\mathrm{depth}_T(u)=i+1$, suppose that $\mathrm{depth}_T(u)=i$ and moreover that $u=v^i_{r+1}$ in the above construction. By virtue of (23) and (24) we know that

$$\mathrm{diam}\left(\mathcal{F}_u\right)=\mathrm{diam}\left(\mathcal{F}_{v^i_{r+1}}\right)\leqslant 3\max_{s\in A^i_{r+1}}\mathrm{diam}\left(\mathcal{F}^{i+1}_{v_s}\right)+2\frac{1-3\tau}{2}\tau^i\leqslant 3\tau^{i+1}+(1-3\tau)\tau^i=\tau^i.$$

Since (19) is also valid for $i=0$ (because $\mathrm{diam}(X)=1$), this concludes the proof of (19).

It remains to prove (21). Since we are assuming that $u\neq v$ are non-leaf vertices and $\mathrm{depth}_T(u)=\mathrm{depth}_T(v)$, we may write $u=v^i_s$ and $v=v^i_t$ for some $i\in\{1,\ldots,mh-1\}$ and $s<t$. Then by the above construction $\mathbf{c}\left(v^i_s\right)=v^{i+1}_{j(s)}$, $\mathbf{c}\left(v^i_t\right)=v^{i+1}_{j(t)}$ and

$$j(t)\in\{1,\ldots,\ell_{i+1}\}\setminus\bigcup_{\ell=1}^{t-1}A^i_\ell\subseteq\{1,\ldots,\ell_{i+1}\}\setminus\bigcup_{\ell=1}^{s-1}A^i_\ell,$$

yet $j(t)\notin A^i_s$. The validity of (21) now follows from the definition of A^i_s; see (23). □

5. An iterated Hölder argument for trees

Lemma 4.2 yields a partition map that will serve as the base case of an inductive construction. This construction will iteratively remove parts of the tree T of Lemma 4.2 while maintaining certain properties of the resulting tree. To simplify matters, it is worthwhile to formulate some of the ensuing steps as combinatorial statements for weighted trees.

As we discussed earlier, the partition map $\mathcal{F} : T \to 2^X$ of Lemma 4.2 induces a weighting $w : T \to (0, \infty)$ of the vertices of T given by $w(u) = \mu(\mathcal{F}_u)$. In this section we will prove inequalities for weighted trees that will be used crucially in our proof of Theorem 1.5. Lemma 5.4 below is inspired by an argument of [2, Lem. 3.25], though our assumptions, proof, and conclusion are different. These changes are crucial for later applications. The appropriate setting for the discussion is a certain class of weighted trees, which we now introduce; the definition below contains the tree of Lemma 4.2 as a special case.

Definition 5.1 (Subadditive weighted tree with designated children)**.** Fix integers $h, k \geqslant 2$. A subadditive weighted tree with designated children is a triple $(T, w, \mathbf{c})$ consisting of a finite rooted graph-theoretical tree T, a mapping $w : T \to (0, \infty)$ and for every non-leaf vertex $u \in T \smallsetminus \mathcal{L}(T)$ a "designated child" $\mathbf{c}(u) \in \mathbf{p}^{-1}(u)$, such that the following conditions hold true.

- All the leaves of T are at the same depth, which is divisible by h.
- For every non-leaf vertex $u \in T \smallsetminus \mathcal{L}(T)$,
$$w(u) \leqslant \sum_{v \in \mathbf{p}^{-1}(u)} w(v). \tag{26}$$
- For every non-leaf vertex $u \in T \smallsetminus \mathcal{L}(T)$,
$$w_h^k(\mathbf{c}(u)) = \max_{v \in \mathbf{p}^{-1}(u)} w_h^k(v), \tag{27}$$
  where $w_h^k : T \to (0, \infty)$ is the modified weight function of Definition 4.1.

Definition 5.2 (Subtree)**.** Let T be a finite rooted tree. A subtree T' of T is a connected rooted subgraph of T whose set of leaves is a subset of the leaves of T, i.e., $\mathcal{L}(T') \subseteq \mathcal{L}(T)$.

Definition 5.3 (Sparsified tree)**.** Fix integers $h, k \geqslant 2$ and let $(T, w, \mathbf{c})$ be a subadditive weighted tree with designated children. For $i \in \mathbb{Z}$ define a subtree $T^{(i)}$ of T as follows.

$$T^{(i)} = T \smallsetminus \left(\bigcup_{\substack{u \in T \\ \mathrm{depth}_T(u) = i-1}} \bigcup_{v \in \mathbf{p}^{-1}(u) \smallsetminus \{\mathbf{c}(u)\}} T_v \right). \tag{28}$$

Thus $T^{(i)}$ is obtained from T by removing all the subtrees rooted at vertices of depth i that are not designated children. Note that by definition $T^{(i)} = T$ if either $i \leqslant 0$ or T has no vertices at depth i.

Lemma 5.4. *Fix $h, k \in \mathbb{N}$ satisfying $h \geqslant k \geqslant 2$ and let $(T, w, \mathbf{c})$ be a subadditive weighted tree with designated children. Assume that all the leaves of T are at depth h. Then there exists $L \subseteq \{1, \ldots, h\}$ with $|L| \geqslant h - k + 1$ such that for every $i \in L$ we have*

$$\sum_{\ell \in \mathcal{L}\left(T^{(i)}\right)} w(\ell)^{\frac{k-1}{k}} \geqslant w(r)^{\frac{k-1}{k}},$$

where $r = r(T)$ is the root of T and $T^{(i)}$ is as in Definition 5.3.

Proof. For $i \in \{1,\dots,h\}$ and $u \in T$ define $f_i(u) \in (0,\infty)$ by reverse induction on $\mathrm{depth}_T(u)$ as follows. If $\mathrm{depth}_T(u) = h$, i.e., u is a leaf of T, set

$$f_i(u) = w(u)^{\frac{k-1}{k}}. \tag{29}$$

If $\mathrm{depth}_T(u) < h$ define recursively

$$f_i(u) = \begin{cases} \max_{v\in \mathbf{p}^{-1}(u)} f_i(v) & \text{if } i = \mathrm{depth}_T(u)+1, \\ \sum_{v\in \mathbf{p}^{-1}(u)} f_i(v) & \text{if } i \neq \mathrm{depth}_T(u)+1. \end{cases} \tag{30}$$

We observe that for all $i \in \{1,\dots,h\}$ and $u \in T$ we have

$$i \leqslant \mathrm{depth}_T(u) \implies f_i(u) = \sum_{\ell\in\mathcal{L}(T_u)} w(\ell)^{\frac{k-1}{k}}, \tag{31}$$

and

$$i > \mathrm{depth}_T(u) \implies f_i(u) = \sum_{\ell\in\mathcal{L}\left((T_u)^{(i)}\right)} w(\ell)^{\frac{k-1}{k}}, \tag{32}$$

where analogously to (28) we define

$$(T_u)^{(i)} = T_u \smallsetminus \left(\bigcup_{\substack{a\in T\\ \mathrm{depth}_T(a)=i-1}} \ \bigcup_{b\in \mathbf{p}^{-1}(a)\smallsetminus\{\mathbf{c}(a)\}} T_b \right). \tag{33}$$

In other words, recalling that T_u is the subtree of T rooted at u, the subtree $(T_u)^{(i)}$ is obtained from T_u by deleting all the subtrees rooted at vertices of depth i that are not designated children (here depth is measured in T, i.e., the distance from the original root $r(T)$).

Identities (31) and (32) follow by reverse induction on $\mathrm{depth}_T(u)$ from the recursive definition of $f_i(u)$. Indeed, if $\mathrm{depth}_T(u) = h$ then (32) is vacuous and (31) follows from (29). Assume that $u \in T$ is not a leaf of T and that (31) and (32) hold true for the children of u. If $i \leqslant \mathrm{depth}_T(u)$ then by (30) and the inductive hypothesis we have

$$f_i(u) = \sum_{v\in\mathbf{p}^{-1}(u)} f_i(v) = \sum_{v\in\mathbf{p}^{-1}(u)} \sum_{\ell\in\mathcal{L}(T_v)} w(\ell)^{\frac{k-1}{k}} = \sum_{\ell\in\mathcal{L}(T_u)} w(\ell)^{\frac{k-1}{k}}.$$

If $i = \mathrm{depth}_T(u)+1$ then since we are assuming that (31) holds for each $v \in \mathbf{p}^{-1}(u)$,

$$\begin{aligned} f_i(u) &\overset{(30)}{=} \max_{v\in\mathbf{p}^{-1}(u)} f_i(v) = \max_{v\in\mathbf{p}^{-1}(u)} \sum_{\ell\in\mathcal{L}(T_v)} w(\ell)^{\frac{k-1}{k}} \overset{(18)}{=} \max_{v\in\mathbf{p}^{-1}(u)} w_h^k(v) \\ &\overset{(27)}{=} w_h^k(\mathbf{c}(u)) \overset{(18)}{=} \sum_{\ell\in\mathcal{L}(T_{\mathbf{c}(u)})} w(\ell)^{\frac{k-1}{k}} \overset{(33)}{=} \sum_{\ell\in\mathcal{L}\left((T_u)^{(i)}\right)} w(\ell)^{\frac{k-1}{k}}. \end{aligned}$$

Finally, if $i > \mathrm{depth}_T(u)+1$ then we are assuming that (32) holds for each $v \in \mathbf{p}^{-1}(u)$, and therefore

$$f_i(u) \overset{(30)}{=} \sum_{v\in\mathbf{p}^{-1}(u)} f_i(v) \overset{(32)}{=} \sum_{v\in\mathbf{p}^{-1}(u)} \sum_{\ell\in\mathcal{L}\left((T_v)^{(i)}\right)} w(\ell)^{\frac{k-1}{k}} \overset{(33)}{=} \sum_{\ell\in\mathcal{L}\left((T_u)^{(i)}\right)} w(\ell)^{\frac{k-1}{k}}.$$

This completes the inductive verification of the identities (31) and (32).

Our next goal is to prove by reverse induction on $\mathrm{depth}_T(u)$ that for every $H \subseteq \{1, \ldots, h\}$ with $|H| = k$ we have,

$$\prod_{i \in H} f_i(u) \geqslant w(u)^{k-1}. \tag{34}$$

Indeed, if $\mathrm{depth}_T(u) = h$ then (34) holds as equality due to (29). Assume inductively that $\mathrm{depth}_T(u) < h$ and that (34) holds for all the children of u. We claim that there exists $j \in H$ such that

$$\prod_{i \in H} f_i(u) \geqslant \left(\max_{v \in \mathbf{p}^{-1}(u)} f_j(v) \right) \prod_{i \in H \smallsetminus \{j\}} \left(\sum_{v \in \mathbf{p}^{-1}(u)} f_i(v) \right). \tag{35}$$

Indeed, if $\mathrm{depth}_T(u) + 1 \in H$ then take $j = \mathrm{depth}_T(u) + 1$ and note that (35) holds as equality due to (30). On the other hand, if $\mathrm{depth}_T(u) + 1 \notin H$ then let j be an arbitrary element of H, and note that due to (30) we have

$$\begin{aligned} \prod_{i \in H} f_i(u) &= \left(\sum_{v \in \mathbf{p}^{-1}(u)} f_j(v) \right) \prod_{i \in H \smallsetminus \{j\}} \left(\sum_{v \in \mathbf{p}^{-1}(u)} f_i(v) \right) \\ &\geqslant \left(\max_{v \in \mathbf{p}^{-1}(u)} f_j(v) \right) \prod_{i \in H \smallsetminus \{j\}} \left(\sum_{v \in \mathbf{p}^{-1}(u)} f_i(v) \right), \end{aligned}$$

as required. Now,

$$\begin{aligned} \prod_{i \in H} f_i(u) &\overset{(*)}{\geqslant} \left(\max_{v \in \mathbf{p}^{-1}(u)} f_j(v) \right) \left(\sum_{v \in \mathbf{p}^{-1}(u)} \prod_{i \in H \smallsetminus \{j\}} f_i(v)^{\frac{1}{k-1}} \right)^{k-1} \\ &\geqslant \left(\sum_{v \in \mathbf{p}^{-1}(u)} \prod_{i \in H} f_i(v)^{\frac{1}{k-1}} \right)^{k-1} \overset{(**)}{\geqslant} \left(\sum_{v \in \mathbf{p}^{-1}(u)} w(v) \right)^{k-1} \overset{(26)}{\geqslant} w(u)^{k-1}, \end{aligned}$$

where in $(*)$ we used (35) and Hölder's inequality, and in $(**)$ we used the inductive hypothesis. This concludes the proof of (34).

We are now in position to complete the proof of Lemma 5.4. Set $H_1 = \{1, \ldots, k\}$. Then

$$\max_{i \in H_1} \sum_{\ell \in \mathcal{L}\left(T^{(i)}\right)} w(\ell)^{\frac{k-1}{k}} \geqslant \left(\prod_{i \in H_1} \sum_{\ell \in \mathcal{L}\left(T^{(i)}\right)} w(\ell)^{\frac{k-1}{k}} \right)^{\frac{1}{k}} \overset{(32)}{=} \left(\prod_{i \in H_1} f_i(r) \right)^{\frac{1}{k}} \overset{(34)}{\geqslant} w(r)^{\frac{k-1}{k}}.$$

Hence there is $i_1 \in H_1$ satisfying

$$\sum_{\ell \in \mathcal{L}\left(T^{(i_1)}\right)} w(\ell)^{\frac{k-1}{k}} \geqslant w(r)^{\frac{k-1}{k}}.$$

Now define $H_2 = (H_1 \smallsetminus \{i_1\}) \cup \{k+1\}$ and repeat the above argument with H_2 replacing H_1. We deduce that there exists $i_2 \in H_2$ satisfying

$$\sum_{\ell \in \mathcal{L}\left(T^{(i_2)}\right)} w(\ell)^{\frac{k-1}{k}} \geqslant w(r)^{\frac{k-1}{k}}.$$

We may repeat this process inductively $h-k+1$ times, and let $L=\{i_1,i_2,\dots,i_{h-k+1}\}$. $\square$

Definition 5.5 (Subtree sparsified at a subset)**.** Fix two integers $h,k\geqslant 2$ and let let $(T,w,\mathbf{c})$ be a subadditive weighted tree with designated children (recall Definition 5.1). Let T' be a subtree of T (see Definition 5.2) and $S\subseteq T'$. We say that the subtree T' is sparsified at S if every $v\in S$ is a designated child and has no siblings in T'. For the purpose of this definition we declare the root $r(T)$ to be a designated child, i.e., we allow $r(T)\in S$. Thus $v\in T$ is a designated child if it is either the root of T or $\mathbf{c}(\mathbf{p}(v))=v$.

Lemma 5.6. *Fix two integers $h,k\geqslant 2$ with $h\geqslant 2k^2$. Let $(T,w,\mathbf{c})$ be a subadditive weighted tree with designated children as in Definition 5.1 (thus all the leaves of T are at the same depth, which is divisible by h, and the designated child map $\mathbf{c}$ satisfies* (27)*). Then there exists a subtree T' of T with the same root as T, and two subsets $R,S\subseteq T'$, both containing the root of T', with the following properties:*

- *For any non-leaf $u\in T'$ we have $\mathbf{c}(u)\in T'$.*
- $R=\{v\in T' : h\mid \mathrm{depth}_T(v)\}$.
- *For any non-leaf vertex $u\in R$,*

$$\sum_{v\in D_{T'}(u,R)} w(v)^{\left(1-\frac{1}{k}\right)^2}\geqslant w(u)^{\left(1-\frac{1}{k}\right)^2}. \tag{36}$$

  Recall that $D_{T'}(\cdot,\cdot)$ is given in Definition 2.2.
- *For every $u,v\in R$ such that* $\mathrm{depth}_T(v)=\mathrm{depth}_T(u)+h$ *and v is a descendant of u, there is one and only one $w\in S$ such that w lies on the path joining u and v and* $\mathrm{depth}_T(u)<\mathrm{depth}_T(w)\leqslant\mathrm{depth}_T(v)$.
- *For any $u\in T'$ such that $D_{T'}(u,S)\neq\emptyset$, all the vertices of $D_{T'}(u,S)$ are at the same depth in T'_u, which is an integer between 1 and $2h$.*
- *T' is sparsified at the subset S.*

Proof. The proof of Lemma 5.6 is by induction on the height of the tree, but we need the following strengthening of the inductive hypothesis so as to deal with multiple trees. Suppose that we are given a collection of (disjoint) subadditive weighted trees with designated children $(T_1,w,\mathbf{c})\dots,(T_\ell,w,\mathbf{c})$, each T_i is rooted at r_i and all of them having the same height, which is divisible by h (formally we should denote the weighting of T_i by w_i, but since the trees are disjoint, denoting all the weightings by w will not create any confusion). We will prove that there exists a subset $C\subseteq\{1,\dots,\ell\}$ with the following properties.

- For every $i\in C$ there is a subtree T_i' of T_i rooted at r_i, and subsets $S_i,R_i\subseteq T_i'$, both containing the root of T_i', such that for any non-leaf $u\in T_i'$ we have $\mathbf{c}(u)\in T_i'$.
- $R_i=\{v\in T_i' :\ h\mid \mathrm{depth}_{T_i}(v)\}$.
- For any non-leaf vertex $u\in R_i$,

$$\sum_{v\in D_{T_i'}(u,R_i)} w(v)^{\left(1-\frac{1}{k}\right)^2}\geqslant w(u)^{\left(1-\frac{1}{k}\right)^2}. \tag{37}$$

- For every $u,v\in R_i$ such that $\mathrm{depth}_{T_i}(v)=\mathrm{depth}_{T_i}(u)+h$ and v is a descendant of u, there is one and only one $w\in S_i$ such that w lies on the path joining u and v and $\mathrm{depth}_{T_i}(u)<\mathrm{depth}_{T_i}(w)\leqslant\mathrm{depth}_{T_i}(v)$.
- For any $u\in T_i'$ such that $D_{T_i'}(u,S_i)\neq\emptyset$, all the vertices of $D_{T_i'}(u,S_i)$ are at the same depth in $(T_i')_u$, which is an integer between 1 and $2h$.

- T_i' is sparsified at the subset S_i.
- The vertices in $\bigcup_{i\in C} D_{T_i'}(r_i,S_i)$ have the same depth (in their respective tree), regardless of i, and

$$\sum_{i\in C} w(r_i)^{\left(1-\frac{1}{k}\right)^2} \geqslant \left(\sum_{i=1}^{\ell} w(r_i)^{1-\frac{1}{k}}\right)^{1-\frac{1}{k}}. \tag{38}$$

Note that Lemma 5.6 is the case $\ell = 1$ of this statement, but it will be beneficial to prove the more general statement as formulated above.

When the height of all the T_i is 0 we simply set $C = \{1,\ldots,\ell\}$ and $T_i' = S_i = R_i = T_i$. Most of the above conditions hold vacuously in this case (there are no non-leaf vertices and $D_{T_i'}(u,S_i) = \emptyset$). Condition (38) follows from subadditivity of the function $(0,\infty) \ni t \mapsto t^{1-\frac{1}{k}}$.

Assume next that the leaves of $\{T_i\}_{i=1}^{\ell}$ are all at depth mh for some $m \in \mathbb{N}$. Let $\widehat{T}_j$ be the subgraph of T_j induced on all the vertices of depth at most h in T_j. Note that by (18) the restriction of the induced weight function w_h^k to $\widehat{T}_j$ coincides with the corresponding weight function induced by the weighted tree $\left(\widehat{T}_j, w\right)$, and consequently the same can be said about the designated child map $\mathbf{c}|_{\widehat{T}_j}$. Therefore, an application of Lemma 5.4 to $\left(\widehat{T}_j, w, \mathbf{c}\right)$ yields a subset $L_j \subseteq \{1,\ldots,h\}$ with $|L_j| = h-k+1$ such that for all $i \in L_j$ we have

$$\sum_{\substack{u\in \widehat{T}_j^{(i)}\\ \mathrm{depth}_{T_j}(u)=h}} w(u)^{1-\frac{1}{k}} \geqslant w(r_j)^{1-\frac{1}{k}}, \tag{39}$$

where $\widehat{T}_j^{(i)}$ is the subtree of $\widehat{T}_j$ that is obtained by sparsifying the ith level as in Definition 5.3.

Let $j_0 \in \{1,\ldots,\ell\}$ satisfy

$$w(r_{j_0}) = \max_{j\in\{1,\ldots,\ell\}} w(r_j). \tag{40}$$

For $j \in \{1,\ldots,\ell\}$ denote $L_j' = L_j \cap L_{j_0}$. Then

$$|L_j'| = |L_j| + |L_{j_0}| - |L_j \cup L_{j_0}| \geqslant (h-k+1) + (h-k+1) - h = h - 2(k-1). \tag{41}$$

If $\ell = 1$ let $s_0 \in \{1,\ldots,h\}$ be an arbitrary integer in L_{j_0}. If $\ell \geqslant 2$ let $s_0 \in \{1,\ldots,h\}$ satisfy

$$\sum_{\substack{j\in\{1,\ldots,\ell\}\smallsetminus\{j_0\}\\ s_0\in L_j'}} w(r_j)^{1-\frac{1}{k}} = \max_{s\in\{1,\ldots,h\}} \sum_{\substack{j\in\{1,\ldots,\ell\}\smallsetminus\{j_0\}\\ s\in L_j'}} w(r_j)^{1-\frac{1}{k}}. \tag{42}$$

By averaging we see that

$$\begin{aligned}\sum_{\substack{j\in\{1,\ldots,\ell\}\smallsetminus\{j_0\}\\ s_0\in L_j'}} w(r_j)^{1-\frac{1}{k}} &\geqslant \frac{1}{h-k+1}\sum_{s\in L_{j_0}} \sum_{\substack{j\in\{1,\ldots,\ell\}\smallsetminus\{j_0\}\\ s\in L_j'}} w(r_j)^{1-\frac{1}{k}}\\ &= \frac{1}{h-k+1}\sum_{j\in\{1,\ldots,\ell\}\smallsetminus\{j_0\}} |L_j'| w(r_j)^{1-\frac{1}{k}} \overset{(41)}{\geqslant} \frac{h-2(k-1)}{h-(k-1)} \sum_{j\in\{1,\ldots,\ell\}\smallsetminus\{j_0\}} w(r_j)^{1-\frac{1}{k}}. \end{aligned}\tag{43}$$

Now define

$$C = \{j \in \{1,\ldots,\ell\} : s_0 \in L_j\}. \tag{44}$$

We know that $j_0 \in C$, since this is trivial if $\ell = 1$ and if $\ell \geqslant 2$ then $s_0 \in L'_j \subseteq L_{j_0}$ for some $j \in \{1, \ldots, \ell\} \smallsetminus \{j_0\}$. This also means that if $\ell \geqslant 2$ then $|C| \geqslant 2$. We shall now prove (38). By the choice of j_0 in (40) we know that for all $j \in \{1, \ldots, \ell\} \smallsetminus \{j_0\}$ we have

$$w\left(r_j\right)^{1-\frac{1}{k}} \leqslant \frac{1}{2} \sum_{i=1}^{\ell} w(r_i)^{1-\frac{1}{k}}. \tag{45}$$

This implies that for all $j \in \{1, \ldots, \ell\} \smallsetminus \{j_0\}$,

$$w(r_j)^{\left(1-\frac{1}{k}\right)^2} \geqslant \frac{h-(k-1)}{h-2(k-1)} \cdot \frac{w(r_j)^{1-\frac{1}{k}}}{\left(\sum_{i=1}^{\ell} w(r_i)^{1-\frac{1}{k}}\right)^{\frac{1}{k}}}. \tag{46}$$

To check (46) note that it is equivalent to the inequality

$$\frac{h-(k-1)}{h-2(k-1)} \left(\frac{w(r_j)^{1-\frac{1}{k}}}{\sum_{i=1}^{\ell} w(r_i)^{1-\frac{1}{k}}}\right)^{\frac{1}{k}} \leqslant 1.$$

By (45) and the fact that $h \geqslant 2k^2$, it suffices to show that $(2k^2-k+1)/(2k^2-2k+2) \leqslant 2^{1/k}$. Since $2^{1/k} \geqslant 1 + 1/(2k)$ it suffices to check that $2k(2k^2-k+1) \leqslant (2k+1)(2k^2-2k+2)$, which is immediate to verify.

Having established (46), we proceed as follows.

$$\begin{aligned}
\sum_{j\in C} w(r_j)^{\left(1-\frac{1}{k}\right)^2} \quad &= \quad \frac{w(r_{j_0})^{1-\frac{1}{k}}}{w(r_{j_0})^{\left(1-\frac{1}{k}\right)\frac{1}{k}}} + \sum_{j\in C\smallsetminus\{j_0\}} w(r_j)^{\left(1-\frac{1}{k}\right)^2} \\
&\overset{(46)}{\geqslant} \quad \frac{w(r_{j_0})^{1-\frac{1}{k}}}{\left(\sum_{i=1}^{\ell} w(r_i)^{1-\frac{1}{k}}\right)^{\frac{1}{k}}} + \frac{h-(k-1)}{h-2(k-1)} \cdot \frac{\sum_{j\in C\smallsetminus\{j_0\}} w(r_j)^{1-\frac{1}{k}}}{\left(\sum_{i=1}^{\ell} w(r_i)^{1-\frac{1}{k}}\right)^{\frac{1}{k}}} \\
&\overset{(43)\wedge(44)}{\geqslant} \quad \frac{w(r_{j_0})^{1-\frac{1}{k}}}{\left(\sum_{i=1}^{\ell} w(r_i)^{1-\frac{1}{k}}\right)^{\frac{1}{k}}} + \frac{\sum_{j\in\{1,\ldots,\ell\}\smallsetminus\{j_0\}} w(r_j)^{1-\frac{1}{k}}}{\left(\sum_{i=1}^{\ell} w(r_i)^{1-\frac{1}{k}}\right)^{\frac{1}{k}}} \\
&= \quad \left(\sum_{i=1}^{\ell} w(r_i)^{1-\frac{1}{k}}\right)^{1-\frac{1}{k}},
\end{aligned}$$

completing the proof of (38).

We can now complete the proof of Lemma 5.6 by applying the inductive hypothesis. For every $j \in C$ let $u_1^j, u_2^j, \ldots, u_{\ell_j}^j$ be the leaves of the tree $\widehat{T}_j^{(s_0)}$, i.e., the subtree of $\widehat{T}_j$ that was sparsified at level s_0. Consider the subtrees of T_j that are rooted at $u_1^j, u_2^j, \ldots, u_{\ell_j}^j$, i.e., $(T_j)_{u_1^j}, (T_j)_{u_2^j}, \ldots, (T_j)_{u_{\ell_j}^j}$. By the inductive hypothesis applied to these trees there exists a subset $C_j \subseteq \{1, \ldots, \ell_j\}$ such that for each $i \in C_j$ there is a subtree $(T_j)'_{u_i^j}$ of $(T_j)_{u_i^j}$ and subsets $S_{ji}, R_{ji} \subseteq (T_j)'_{u_i^j}$ that satisfy the inductive hypotheses.

Denote for $j \in C$,

$$T_j' = \left(\bigcup_{i \in C_j} (T_j)'_{u_i^j} \right) \bigcup \left(\bigcup_{i \in C_j} \left\{ u \in \widehat{T}_j^{(s_0)} :\ u \text{ ancestor of } u_i^j \right\} \right).$$

Thus T_j' is obtained by taking the subtree of $\widehat{T}_j^{(s_0)}$ whose leaves are $\left\{u_i^j\right\}_{i \in C_j}$, and replacing every leaf u_i^j by the tree $(T_j)'_{u_i^j}$. We also define $R_j = \left(\bigcup_{i=1}^{\ell_j} R_{ji}\right) \bigcup \{r_j\}$, and

$$S_j = \left(\bigcup_{i=1}^{\ell_j} \left(S_{ji} \smallsetminus \{u_i^j\} \right) \right) \bigcup \left\{ u \in \widehat{T}_j^{(s_0)} :\ \operatorname{depth}_{T_j}(u) = s_0 \right\} \bigcup \{r_j\}.$$

Note by the definition of $\widehat{T}_j^{(s_0)}$ every $u \in \widehat{T}_j^{(s_0)}$ with $\operatorname{depth}_{T_j}(u) = s_0$ has no siblings in $\widehat{T}_j^{(s_0)}$.

All the desired properties of T_j', R_j, S_j follow immediately for the construction; only (37) when $u = r_j$ requires justification as follows.

$$\sum_{v \in D_{T_j'}(r_j, R_j)} w(v)^{\left(1-\frac{1}{k}\right)^2} = \sum_{i \in C_j} w\left(u_i^j\right)^{\left(1-\frac{1}{k}\right)^2} \overset{(38)}{\geqslant} \left(\sum_{i=1}^{\ell_j} w\left(u_i^j\right)^{1-\frac{1}{k}} \right)^{1-\frac{1}{k}} \overset{(39)}{\geqslant} w(r_j)^{\left(1-\frac{1}{k}\right)^2}. \quad \square$$

6. Construction of ultrametric subsets

An application of Lemma 5.6 to the initial fragmentation tree of Lemma 4.2 yields a fragmentation tree with useful geometric properties that will be exploited heavily in the subsequent steps of the proof of Theorem 1.5.

Corollary 6.1. *Fix $\tau \in (0, 1/3)$ and integers $m, h, k \geqslant 2$ with $h \geqslant 2k^2$. Let (X, d, μ) be a finite metric measure space of diameter 1. Then there exists a fragmentation map $\mathcal{F} : T \to 2^X$ with the following properties. universal $c > 0$)*

{*1*} *All the leaves of the tree T are at depth mh.*

{*2*} *For every $u \in T$ we have*

$$\operatorname{diam}(\mathcal{F}_u) \leqslant \tau^{\operatorname{depth}_T(u)}. \tag{47}$$

{*3*} *Denote by $R \subseteq T$ the set of vertices at depths which are integer multiples of h. Then for every non-leaf $u \in R$,*

$$\sum_{v \in D_T(u,R)} \mu\left(\mathcal{F}_v\right)^{\left(1-\frac{1}{k}\right)^2} \geqslant \mu\left(\mathcal{F}_u\right)^{\left(1-\frac{1}{k}\right)^2}. \tag{48}$$

{*4*} *There is a subset $S \subseteq T$ containing the root of T such that every $u \in S$ is either the root or has no sibling in T. Moreover, R and S are "alternating" in the following sense. For every $u, v \in R$ such that $\operatorname{depth}_T(v) = \operatorname{depth}_T(u) + h$ and v is a descendant of u, there is one and only one $w \in S$ such that w lies on the path joining u and v and $\operatorname{depth}_T(u) < \operatorname{depth}_T(w) \leqslant \operatorname{depth}_T(v)$.*

{*5*} *The vertices of S are $\frac{1-3\tau}{2\tau}$-separated (recall Definition 2.7).*

{*6*} *For every $u \in R$, all the vertices in $D_T(u, S)$ have the same depth in T, which is at most $\operatorname{depth}_T(u) + h$.*

{*7*} *$\mathcal{F}$ is $\left(\frac{2}{1-3\tau}\tau^{-2h}, \tau^{-1}\right)$-lacunary (recall Definition 2.6).*

Proof. Let $\mathcal{F}^0 : T^0 \to 2^X$ be the partition map of Lemma 4.2, constructed with parameters m, h, k, and having the associated designated child map $\mathbf{c}$ from Lemma 4.2. Let T be the tree obtained by applying Lemma 5.6 to $(T^0, w, \mathbf{c})$, where $w : T^0 \to (0,\infty)$ is given by $w(v) = \mu(\mathcal{F}^0_v)$. Define a fragmentation map $\mathcal{F} : T \to 2^X$ by $\mathcal{F} = \mathcal{F}^0|_T$, i.e., by restricting $\mathcal{F}^0$ to the subtree T. Properties $\{1\}$, $\{2\}$ are satisfied by $\mathcal{F}^0$ due to Lemma 4.2, and therefore they are also satisfied by $\mathcal{F}$ since T has the same root as T^0. Properties $\{3\}$, $\{4\}$, $\{6\}$ are part of the conclusion of Lemma 5.6. It remains to prove properties $\{5\}$ and $\{7\}$.

Assume that $u \in S$. Take $y \in \left(\partial\mathcal{F}_{r(T)}\right) \smallsetminus (\partial\mathcal{F}_u)$. In order to prove property $\{5\}$ it suffices to show that $d(y, \mathcal{F}_u) \geqslant \frac{1-3\tau}{2\tau}\operatorname{diam}(\mathcal{F}_u)$. By property $\{4\}$ it follows that $u = \mathbf{c}(\mathbf{p}(u))$. Let $w = \mathrm{lca}_T(u, y)$ and take $u', y' \in D_T(w, S)$ such that u' is a weak ancestor of u and y' is a weak ancestor of y. By Lemma 5.6 we know that $\operatorname{depth}_T(u') = \operatorname{depth}_T(y')$, and therefore by conclusion (21) of Lemma 4.2 and using the fact that $\mathbf{c}(\mathbf{p}(u')) = u'$ and $\mathbf{c}(\mathbf{p}(y')) = y'$ (because $u', y' \in S$),

$$d\left(y, \mathcal{F}_u\right) \geqslant d\left(\mathcal{F}_{y'}, \mathcal{F}_{u'}\right) \geqslant \frac{1-3\tau}{2}\tau^{\operatorname{depth}_T(u')-1} \geqslant \frac{1-3\tau}{2\tau}\tau^{\operatorname{depth}_T(u)} \overset{(47)}{\geqslant} \frac{1-3\tau}{2\tau}\operatorname{diam}(\mathcal{F}_u).$$

It remains to prove property $\{7\}$. Take $q \in T$ and let $u \in T$ be a weak descendent of q that has at least two children in T, i.e., $v, w \in \mathbf{p}^{-1}(u) \cap T$, $v \neq w$. Our goal is to show that

$$\operatorname{diam}\left(\mathcal{F}_q\right) \leqslant \frac{2\tau^{-2h}}{1-3\tau} \cdot \tau^{\operatorname{depth}_T(q)-\operatorname{depth}_T(u)} \cdot d\left(\partial\mathcal{F}_v, \partial\mathcal{F}_w\right). \tag{49}$$

Since v and w are siblings in T we know by $\{4\}$ that $\{u, v\} \cap S = \emptyset$. Note that

$$\partial F_v = \bigcup_{y \in D(v,S)} \partial\mathcal{F}_y \quad \text{and} \quad \partial F_w = \bigcup_{x \in D(w,S)} \partial\mathcal{F}_x,$$

and therefore

$$d\left(\partial\mathcal{F}_v, \partial F_w\right) = \min_{\substack{y \in D(v,S)\\ x \in D(w,S)}} d\left(\partial F_y, \partial F_x\right). \tag{50}$$

Note that since $\{v, w\} \cap S = \emptyset$ we have $D(v,S) \cup D(w,S) \subseteq D(u,S)$. By Lemma 5.6 it follows that all the vertices in $D(v,S) \cup D(w,S)$ are at the same depth in T. Denote this depth by ℓ. Due to Lemma 5.6 we know that $\ell \leqslant \operatorname{depth}_T(u) + 2h$. By conclusion (21) of Lemma 4.2 we deduce that for all $y \in D(v,S)$ and $x \in D(w,S)$ we have

$$d\left(\partial\mathcal{F}_y, \partial\mathcal{F}_x\right) \geqslant d\left(\mathcal{F}_y, \mathcal{F}_x\right) = d\left(\mathcal{F}_{\mathbf{c}(\mathbf{p}(y))}, \mathcal{F}_{\mathbf{c}(\mathbf{p}(x))}\right) > \frac{1-3\tau}{2}\tau^{\ell} \geqslant \frac{1-3\tau}{2}\tau^{\operatorname{depth}_T(u)+2h}. \tag{51}$$

Hence,

$$d\left(\partial\mathcal{F}_v, \partial F_w\right) \overset{(50)\wedge(51)}{>} \frac{1-3\tau}{2}\tau^{\operatorname{depth}_T(u)+2h} \overset{(47)}{\geqslant} \frac{1-3\tau}{2}\tau^{\operatorname{depth}_T(u)-\operatorname{depth}_T(q)+2h}\operatorname{diam}(\mathcal{F}_q),$$

proving (49). $\square$

In order to proceed with the proof of Theorem 1.5 we will need to use a weighted version of the results of [15, 19]. Theorem 6.2 below follows from a slight modification of the argument in [19], though it is not stated there explicitly. We will therefore explain how [19] can be modified to deduce this statement. Alternatively, a similar statement (with slightly worse distortion bound) can be obtained by natural modifications of the argument in [15].

Theorem 6.2. *Let (X,d) be a finite metric space and $w_1, w_2 : X \to [0,\infty)$ two nonnegative weight functions. Then for every $\varepsilon \in (0,1)$ there exists a subset $S \subseteq X$ that embeds into an ultrametric space with distortion*

$$D = \frac{2}{\varepsilon(1-\varepsilon)^{\frac{1-\varepsilon}{\varepsilon}}}, \tag{52}$$

and satisfying

$$\left(\sum_{x\in S} w_1(x)\right)\left(\sum_{x\in X} w_2(x)\right)^{\varepsilon} \geqslant \sum_{x\in X} w_1(x) w_2(x)^{\varepsilon}. \tag{53}$$

Remark 6.3. Theorem 6.2 implies the finite nonlinear Dvoretzky theorem (Theorem 1.2) in the special case $w_1 = w_2 = 1$. Another special case of interest is when $w_1 = w$ and $w_2 = w^{1-\varepsilon}$. In this case conclusion (53) becomes

$$\sum_{x\in S} w(x)^{1-\varepsilon} \geqslant \left(\sum_{x\in X} w(x)\right)^{1-\varepsilon}. \tag{54}$$

This type of requirement was studied in [2] under the name of "the weighted metric Ramsey problem", where is was shown that there always exists $S \subseteq X$ satisfying (54) that embeds into an ultrametric space with distortion $O\left(\varepsilon^{-1}\log(2/\varepsilon)\right)$.

Proof of Lemma 6.2. The beginning of the argument is most natural to state in the context of general metric measure spaces (X,d,μ). So, assume that (X,d,μ) is a metric measure space; we will later specialize the discussion to the case of finite spaces.

Let $f : X \to [0,\infty)$ be a nonnegative Borel measurable function. Lemma 2.1 of [19] states that for every compact $S \subseteq X$ and every $R > r > 0$ there exists a compact subset $T \subseteq S$ satisfying

$$\int_T \frac{\mu(B(x,R))}{\mu(B(x,r))} f(x)\,d\mu(x) \geqslant \int_S f\,d\mu, \tag{55}$$

such that T can be partitioned as $T = \bigcup_{n=1}^{\infty} T_n$, where each (possibly empty) T_n is compact and contained in a ball of radius r, and any two non-empty T_n, T_m are separated by a distance of at least $R - r$.

Fix a nonnegative Borel measurable $w \in L_1(\mu)$. Iterate the above statement as follows; the same iteration is carried out for the special case $w = 1$ in Lemma 2.2 of [19]. Assume that we are given a non-increasing sequence of positive numbers $R = r_0 \geqslant r_1 \geqslant r_2 \geqslant \cdots > 0$ converging to zero. Assume also that $\mathrm{diam}(X) \leqslant 2R$. For $n \in \mathbb{N}$ define $f_n : X \to [0,\infty)$ by

$$f_n(x) = \left(\prod_{m=n}^{\infty} \frac{\mu(B(x,r_m))}{\mu\left(B\left(x, r_m + \frac{2r_{m-1}}{D}\right)\right)}\right) w(x), \tag{56}$$

Where D be given by (52). Note that $0 \leqslant f_n \leqslant w$ for all $n \in \mathbb{N}$. Assume that we already defined a compact subset $S_{n-1} \subseteq X$. An application of Lemma 2.1 of [19], with radii $r_n + \frac{2r_{n-1}}{D} > r_n$ and weight function f_n, yields a compact subset $S_n \subseteq S_{n-1}$ satisfying

$$\int_{S_n} f_{n+1}\,d\mu = \int_{S_n} \frac{\mu\left(B\left(x, r_n + \frac{2r_{n-1}}{D}\right)\right)}{\mu(B(x,r_n))} f_n(x)\,d\mu(x) \overset{(55)}{\geqslant} \int_{S_{n-1}} f_n\,d\mu.$$

Hence for all $n \in \mathbb{N}$ we have,

$$\int_{S_n} w d\mu \geqslant \int_X f_1 d\mu.$$

Consider the compact subset $S = \bigcap_{n=1}^\infty S_n$. By the dominated convergence theorem,

$$\int_S w d\mu \geqslant \int_X f_1 d\mu. \tag{57}$$

In [19, Lem. 2.2] it is shown that S embeds with distortion D into an ultrametric space.

Assume now that the radii $1 = r_0 \geqslant r_1 \geqslant r_2 \geqslant \cdots > 0$ are random variables satisfying $\lim_{n\to\infty} r_n = 0$ and for every real number $r > 0$

$$\Pr\left[r_n < r \leqslant r_n + \frac{2r_{n-1}}{D}\right] \leqslant \varepsilon. \tag{58}$$

For the existence of such random variables, as well as the optimality for this purpose of the choice of D in (52), see [19, Thm. 1.5]. Specializing to the case of a finite metric measure space (X, d, μ) of diameter at most 2, apply (57) when $w(x) = w_1(x)/\mu(\{x\})$ and $w_2(x) = \mu(\{x\})$, and the radii are the random radii chosen above. By taking expectation of the resulting (random) inequality and using Jensen's inequality we arrive at the following estimate.

$$\mathbb{E}\left[\sum_{x\in S} w_1(x)\right] \overset{(56)\wedge(57)}{\geqslant} \sum_{x\in X} w_1(x) \exp\left(\mathbb{E}\left[\sum_{n=1}^\infty \log\left(\frac{\mu(B(x, r_m))}{\mu\left(B\left(x, r_m + \frac{2r_{m-1}}{D}\right)\right)}\right)\right]\right). \tag{59}$$

For every $x \in X$ let $0 = t_1(x) < t_2(x) < \ldots < t_{k(x)}(x)$ be the radii at which $\mu(B(x,t))$ jumps, i.e., $\mu(\{x\}) = \mu(B(x, t_1(x))) < \mu(B(x, t_2(x))) < \ldots < \mu(B(x, t_{k(x)}(x))) = \mu(X)$, and $B(x,t) = B(x, t_j(x))$ if $t_j(x) \leqslant t < t_{j+1}(x)$ (where we use the convention $t_{k(x)+1}(x) = \infty$). Then we have the following straightforward identity (see equation (15) in [19]), which holds for every $x \in X$.

$$\begin{aligned}\mathbb{E}&\left[\sum_{n=1}^\infty \log\left(\frac{\mu(B(x, r_m))}{\mu\left(B\left(x, r_m + \frac{2r_{m-1}}{D}\right)\right)}\right)\right] \\ &= -\sum_{j=2}^{k(x)}\left(\sum_{n=1}^\infty \Pr\left[r_n < t_j(x) \leqslant r_n + \frac{2r_{n-1}}{D}\right]\right)\log\left(\frac{\mu\left(B(x, t_j(x))\right)}{\mu\left(B(x, t_{j-1}(x))\right)}\right).\end{aligned} \tag{60}$$

Hence,

$$\begin{aligned}\mathbb{E}\left[\sum_{x\in S} w_1(x)\right] &\overset{(58)\wedge(59)\wedge(60)}{\geqslant} \sum_{x\in X} w_1(x) \exp\left(-\varepsilon \sum_{j=2}^{k(x)} \log\left(\frac{\mu\left(B(x, t_j(x))\right)}{\mu\left(B(x, t_{j-1}(x))\right)}\right)\right) \\ &= \sum_{x\in X} w_1(x)\left(\frac{\mu(\{x\})}{\mu(X)}\right)^\varepsilon = \frac{\sum_{x\in X} w_1(x) w_2(x)^\varepsilon}{\left(\sum_{y\in X} w_2(y)\right)^\varepsilon}.\end{aligned}$$

We have shown that the required estimate (53) holds in expectation for our random subset $S \subseteq X$, completing the proof of Theorem 6.2. □

Definition 6.4. As in [19], given $D \in (2,\infty)$ we let $\theta(D) \in (0,1)$ denote the unique solution of the equation

$$\frac{2}{D} = (1-\theta)\theta^{\frac{\theta}{1-\theta}}. \tag{61}$$

By Remark 6.3 we then know that for every $D > 2$, every finite metric space (X,d) and every $w : X \to (0,\infty)$, there exists $S \subseteq X$ that embeds with distortion D into an ultrametric space and satisfies,

$$\sum_{x\in S} w(x)^{\theta(D)} \geqslant \left(\sum_{x\in X} w(x)\right)^{\theta(D)}. \tag{62}$$

It is elementary to check that

$$\forall D \in (2,\infty), \quad \theta(D) \geqslant 1 - \frac{2e}{D}, \tag{63}$$

and

$$\forall \delta \in (0,1/2), \quad \theta(2+\delta) \geqslant \frac{c\delta}{\log(1/\delta)}, \tag{64}$$

where $c \in (0,\infty)$ is a universal constant.

The following very simple lemma exploits the fact that the class of ultrametrics is closed under truncation. This fact will be a useful normalization in the ensuing arguments.

Lemma 6.5. *Let (X,d) be a bounded metric space that embeds with distortion D into an ultrametric space. Then there exists an ultrametric ρ on X satisfying*

- *$d(x,y) \leqslant \rho(x,y) \leqslant Dd(x,y)$ for all $x,y \in X$,*
- *$\operatorname{diam}_d(X) = \operatorname{diam}_\rho(X)$.*

Proof. We are assuming that there exist $A, B > 0$ and an ultrametric ρ_0 on X that satisfies $Ad(x,y) \leqslant \rho_0(x,y) \leqslant Bd(x,y)$ for all $x,y \in X$, where $B/A \leqslant D$. We can therefore define $\rho = \min\{\rho_0/A, \operatorname{diam}_d(X)\}$. $\square$

Lemma 6.6. *Fix $D \in (2,\infty)$ and $\beta \in (0,\infty)$. Let (X,d) be a finite metric space. Suppose that $\mathcal{F} : T \to 2^X$ is a fragmentation map which is β-separated. Suppose also that there is a weight function $w : T \to (0,\infty)$ which is subadditive, i.e., that for every non-leaf vertex $u \in T \smallsetminus \mathcal{L}(T)$,*

$$\sum_{v\in \mathbf{p}_T^{-1}(u)} w(v) \geqslant w(u), \tag{65}$$

Then there exists a substree T' of T with the same root such that the restricted fragmentation map $\mathcal{G} = \mathcal{F}|_{T'}$ satisfies the following properties.

- *$\left(\partial\mathcal{G}_{r(T')}, d\right)$ embeds into an ultrametric space with distortion $D\left(1+\frac{2}{\beta}\right)$.*
- *Every non-leaf vertex $u \in T' \smallsetminus \mathcal{L}(T')$ satisfies*

$$\sum_{v\in \mathbf{p}_{T'}^{-1}(u)} w(v)^{\theta(D)} \geqslant w(u)^{\theta(D)}. \tag{66}$$

Proof. We use the "metric composition technique" of [2]. For every vertex $u \in T \smallsetminus \mathcal{L}(T)$ let $C_u = \mathbf{p}_T^{-1}(u)$ be the set of children of u in T. Let $\overline{d}_u$ be a metric defined on C_u as follows

$$\overline{d}_u(x,y) = \begin{cases} \operatorname{diam}_d\left((\partial\mathcal{F}_x)\cup(\partial\mathcal{F}_y)\right) & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases} \tag{67}$$

The validity of the triangle inequality for $\overline{d}_u$ is immediate to verify. By the definition of $\theta(D)$ there exists a subset $S_u \subseteq C_u$ such that

$$\sum_{x\in S_u^1} w(x)^{\theta(D)} \overset{(62)}{\geqslant} \left(\sum_{x\in C_u} w(x)\right)^{\theta(D)} \overset{(65)}{\geqslant} w(u)^{\theta(D)}. \tag{68}$$

and $(S_u, \overline{d}_u)$ embeds with distortion D into an ultrametric space. By Lemma 6.5 there exists an ultrametric ρ_u on S_u such that every $x,y \in S_u$ satisfy

$$\begin{aligned}\overline{d}_u(x,y) \leqslant \rho_u(x,y) &\leqslant \min\left\{\operatorname{diam}_{\overline{d}_u}(S_u), D\overline{d}_u(x,y)\right\} \\ &\overset{(67)}{\leqslant} \min\left\{\operatorname{diam}_d\left(\bigcup_{x\in S_u}\partial\mathcal{F}_x\right), D\overline{d}_u(x,y)\right\}. \end{aligned}\tag{69}$$

The subtree $T' \subseteq T$ is now defined inductively in a top-down fashion as follows: declare $r(T) \in T'$ and if $u \in T$ is a non-leaf vertex that was already declared to be in T', add the vertices in S_u to T' as well. Inequality (66) follows from (68). It remains to prove that $\left(\partial\mathcal{G}_{r(T')}, d\right)$ embeds into an ultrametric space with distortion $D\,(1+2/\beta)$. To this end fix $p,q \in \partial\mathcal{G}_{r(T')}$ and choose the corresponding $a,b \in \mathcal{L}(T')$ such that $\mathcal{G}_a = \{p\}$ and $\mathcal{G}_b = \{q\}$. Let $u = \operatorname{lca}_{T'}(a,b) = \operatorname{lca}_T(a,b)$ and choose $x,y \in S_u$ that are weak ancestors of a and b, respectively. Define $\rho(p,q) = \rho_u(x,y)$. Now,

$$d(p,q) = d(\mathcal{G}_a,\mathcal{G}_b) \leqslant \operatorname{diam}_d\left((\partial\mathcal{F}_x)\cup(\partial\mathcal{F}_y)\right) \overset{(67)}{=} \overline{d}_u(x,y) \overset{(69)}{\leqslant} \rho_u(x,y) = \rho(p,q).$$

The corresponding lower bound on $d(p,q)$ is proved as follows, using the assumption that the fragmentation tree $\mathcal{F}$ is β-separated.

$$\frac{\rho(p,q)}{D} \overset{(69)}{\leqslant} \overline{d}_u(x,y) \overset{(67)}{\leqslant} \operatorname{diam}_d(\partial\mathcal{F}_x) + \operatorname{diam}_d(\partial\mathcal{F}_y) + d\left(\partial\mathcal{F}_x,\partial\mathcal{F}_y\right) \overset{(10)}{\leqslant} \left(1+\frac{2}{\beta}\right)d(p,q).$$

We now argue that ρ is an ultrametric on $\partial\mathcal{G}_{r(T')}$. This is where we will use the truncation in (69), i.e., that for all $u \in T \smallsetminus \mathcal{L}(T)$ we have $\operatorname{diam}_{\rho_u}(S_u) \leqslant \operatorname{diam}_d\left(\bigcup_{x\in S_u}\partial\mathcal{F}_x\right)$. Take $p_1,p_2,p_3 \in \partial\mathcal{G}_{r(T')}$ and choose the corresponding $a_1,a_2,a_3 \in \mathcal{L}(T')$ such that $\mathcal{G}_{a_i} = \{p_i\}$ for $i \in \{1,2,3\}$. By relabeling the points if necessary, we may assume that $u = \operatorname{lca}_T(a_1,a_2)$ is a weak descendant of $v = \operatorname{lca}_T(a_2,a_3)$. If $u = v$ take $x_1,x_2,x_3 \in S_u$ that are weak ancestors of a_1,a_2,a_3, respectively. Since ρ_u is an ultrametric, it follows that

$$\rho(p_1,p_2) = \rho_u(x_1,x_2) \leqslant \max\left\{\rho_u(x_1,x_3),\rho_u(x_3,x_2)\right\} = \max\left\{\rho_u(p_1,p_3),\rho_u(p_3,p_2)\right\}.$$

If, on the other hand, u is a proper descendant of v then choose $x_1,x_2 \in S_u$ that are weak ancestors of a_1,a_2 (respectively), and choose $s,t \in S_v$ that are weak ancestors of u,a_3 (respectively). Then,

$$\rho(p_1,p_2)=\rho_u(x,y)\overset{(69)}{\leqslant}\operatorname{diam}_d\left(\bigcup_{w\in S_u}\partial\mathcal{F}_w\right)\leqslant \operatorname{diam}_d(\partial\mathcal{F}_s)$$
$$\leqslant \operatorname{diam}_d\left((\partial\mathcal{F}_s)\cup(\partial\mathcal{F}_t)\right)\overset{(67)}{=}\overline{d}_v(s,t)\overset{(69)}{\leqslant}\rho_v(s,t)=\rho(p_1,p_3)=\rho(p_2,p_3).$$

This establishes the ultratriangle inequality for ρ, completing the proof of Lemma 6.6. □

Lemma 6.7. *Fix $D\in(2,\infty)$, an integer $k\geqslant 2$, and $\tau\in\left(0,\frac{D-2}{3D+2}\right)$. Let (X,d,μ) be a finite metric measure space of diameter 1. Then there exists a fragmentation map $\mathcal{G}:T\to 2^X$ with the following properties.*

- *Every leaf $\ell\in\mathcal{L}(T)$ has no siblings, and furthermore $\mathcal{G}_{\mathbf{p}_T(\ell)}=\mathcal{G}_\ell$.*
- *$\mathcal{G}$ is $\frac{1-3\tau}{2\tau}$-separated and $\left(\frac{2}{1-3\tau}\tau^{-4k^2},\tau^{-4k^2}\right)$-lacunary.*
- *$\left(\partial\mathcal{G}_{r(T)},d\right)$ embeds into an ultrametric space with distortion D.*
- *Every cut-set $G\subseteq T$ (recall Definition 2.1) satisfies*

$$\sum_{v\in G}\mu\left(\mathbf{p}_T(v)\right)^{\left(1-\frac{1}{k}\right)^2\theta\left(\frac{1-3\tau}{1+\tau}D\right)}\geqslant \mu(X)^{\left(1-\frac{1}{k}\right)^2\theta\left(\frac{1-3\tau}{1+\tau}D\right)},\tag{70}$$

where $\mathbf{p}_T(v)$ is the parent of v in T if v is not the root, and the root if v is the root.

Proof. Denote $h=2k^2$ and fix $m\in\mathbb{N}$ satisfying

$$\min_{\substack{x,y\in X\\ x\neq y}} d(x,y)>\tau^{(m-2)h+1}.\tag{71}$$

Apply Corollary 6.1 with parameters τ,m,h,k as above, obtaining a fragmentation tree $\mathcal{F}:T^1\to 2^X$ with corresponding subsets $S,R\subseteq T^1$. Let T^2 be the tree induced by T^1 on S, i.e., join $u,v\in S$ by an edge of T^2 if u is an ancestor of v in T^1 and any $w\in S$ that is an ancestor of v in T^1 is a weak ancestor of u. This is the same as the requirement $v\in D_{T^1}(u,S)$. Let $\mathcal{S}:T^2\to 2^X$ be the fragmentation map obtained by restricting $\mathcal{F}$ to S. To check that $\mathcal{S}$ is indeed a fragmentation map we need to verify that if $u\in\mathcal{L}(T^2)$ then $\mathcal{F}_u$ is a singleton. To see this let $v=\mathbf{p}_{T^2}(u)$ be the parent in T^2 of u. Since all the leaves of T are at depth mh, we have $\operatorname{depth}_{T^1}(v)\geqslant(m-2)h+1$. Using (47) we deduce that $\operatorname{diam}(\mathcal{F}_u)\leqslant\operatorname{diam}(\mathcal{F}_v)\leqslant\tau^{(m-2)h+1}$, implying that $\mathcal{S}_u$ and $\mathcal{S}_v$ are both singletons due to (71).

Since by Corollary 6.1 we know that the vertices in S are $\frac{1-3\tau}{2\tau}$-separated in the fragmentation map $\mathcal{F}$, it follows that the fragmentation map $\mathcal{S}$ is $\frac{1-3\tau}{2\tau}$-separated. Corollary 6.1 also ensures that $\mathcal{F}$ is $\left(\frac{2}{1-3\tau}\tau^{-2h},\tau^{-1}\right)$-lacunary. This implies that $\mathcal{S}$ is $\left(\frac{2}{1-3\tau}\tau^{-2h},\tau^{-2h}\right)$-lacunary. Indeed, due to Corollary 6.1 we know that if $u\in S$ and $v\in D_{T^1}(u,S)$ is a child of u in T^2 then $\operatorname{depth}_{T^1}(u)\leqslant\operatorname{depth}_{T^1}(u)+2h-1$. This implies that if $q,u\in S$ are such that u is a weak descendant of q in T^2 then $\operatorname{depth}_{T^1}(u)-\operatorname{depth}_{T^1}(q)\leqslant 2h\left(\operatorname{depth}_{T^2}(u)-\operatorname{depth}_{T^2}(q)\right)$. Hence, if $v,w\in D_{T^1}(u,S)$ are distinct children of u in T^2, choose distinct $x,y\in T^1$ that are children of u in T^1 and weak ancestors of v,w (respectively), and use the fact that $\mathcal{F}$ is $\left(\frac{2}{1-3\tau}\tau^{-2h},\tau^{-1}\right)$-lacunary to deduce that

$$\operatorname{diam}_d(\mathcal{S}_q) = \operatorname{diam}_d(\mathcal{F}_q) \leqslant \frac{2\tau^{-2h}}{1-3\tau} \cdot \tau^{-\left(\operatorname{depth}_{T^1}(u)-\operatorname{depth}_{T^1}(q)\right)} d\left(\partial\mathcal{F}_x, \partial\mathcal{F}_y\right)$$
$$\leqslant \frac{2\tau^{-2h}}{1-3\tau} \cdot \tau^{-2h\left(\operatorname{depth}_{T^2}(u)-\operatorname{depth}_{T^2}(q)\right)} d\left(\partial\mathcal{S}_x, \partial\mathcal{S}_y\right).$$

Define $w_R : R \to (0,\infty)$ by top-down induction as follows. Set

$$w_R(r) = \mu(X)^{\left(1-\frac{1}{k}\right)^2}, \tag{72}$$

where r is the root of T^1. If $u \in R$ is not a leaf and $v \in D_{T^1}(u,R)$ then define

$$w_R(v) = \frac{w_R(u)}{\sum_{z\in D_{T^1}(u,R)} \mu(\mathcal{F}_z)^{\left(1-\frac{1}{k}\right)^2}} \cdot \mu(\mathcal{F}_v)^{\left(1-\frac{1}{k}\right)^2}. \tag{73}$$

Thus for every non-leaf $u \in R$ we have

$$w_R(u) = \sum_{v\in D_{T^1}(u,R)} w_R(v). \tag{74}$$

Moreover, it follows from the recursive definition (73) combined with (48) that

$$\forall\ u \in R, \quad w_R(u) \leqslant \mu(\mathcal{F}_u)^{\left(1-\frac{1}{k}\right)^2}. \tag{75}$$

Recalling the notation $D^*_T(x,A)$ as given in (8), by summing (74) we see that for all $u \in S \smallsetminus \mathcal{L}(T^2)$ we have

$$\sum_{x\in D^*_{T^1}(u,R)} \sum_{y\in D_{T^1}(x,R)} w_R(y) = \sum_{x\in D^*_{T^1}(u,R)} w_R(x). \tag{76}$$

Notice that

$$\bigcup_{x\in D^*_{T^1}(u,R)} D_{T^1}(x,R) = \bigcup_{v\in D_{T^1}(u,S)} D^*_{T^1}(v,R), \tag{77}$$

where the unions on both sides of (77) are disjoint. Hence,

$$\sum_{x\in D^*_{T^1}(u,R)} w_R(x) \overset{(76)\wedge(77)}{=} \sum_{v\in D_{T^1}(u,S)} \sum_{z\in D^*_{T^1}(v,R)} w_R(z). \tag{78}$$

Define $w_S : S \to (0,\infty)$ by

$$w_S(u) = \sum_{z\in D^*_{T^1}(u,R)} w_R(z). \tag{79}$$

Then for all $u \in S \smallsetminus \mathcal{L}(T^2)$ we have

$$\sum_{v\in \mathbf{p}^{-1}_{T^2}(u)} w_S(v) = \sum_{v\in D_{T^1}(u,S)} w_S(v) \overset{(78)\wedge(79)}{=} \sum_{x\in D^*_{T^1}(u,R)} w_R(x) \overset{(79)}{=} w_S(u).$$

This establishes condition (65) of Lemma 6.6 for the weighting w_S of T^2. Before applying Lemma 6.6 we record one more useful fact about w_S. Recall that for $u \in S$ the vertex $\mathbf{p}_{T^2}(u)$

is r if $u = r$, and otherwise it is the first proper ancestor of u in T^1 which is in S. Take $u' \in D^*_{T^1}(\mathbf{p}_{T^2}(u), R)$ which is a weak ancestor of u. Then $\mathcal{F}_{\mathbf{p}_{T^2}(u)} \supseteq \mathcal{F}_{u'}$, and therefore

$$\mu\left(\mathcal{F}_{\mathbf{p}_{T^2}(u)}\right)^{\left(1-\frac{1}{k}\right)^2} \geqslant \mu\left(\mathcal{F}_{u'}\right)^{\left(1-\frac{1}{k}\right)^2} \overset{(75)}{\geqslant} w_R(u') \overset{(74)}{=} \sum_{x\in D_{T^1}(u',R)} w_R(x) = \sum_{x\in D^*_{T^1}(u,R)} w_R(x) \overset{(79)}{=} w_S(u). \quad (80)$$

Apply Lemma 6.6 to $\mathcal{S} : T^2 \to 2^X$ and $w_S : T^2 \to (0,\infty)$, with $\beta = (1-3\tau)/(2\tau)$ and the parameter D of Lemma 6.6 replaced by $D/(1+2/\beta) = D(1-3\tau)/(1+\tau)$. Note that our assumption $\tau < (D-2)/(3D+2)$ guarantees that this new value of D is bigger than 2, so we are indeed allowed to use Lemma 6.6. We therefore obtain a subtree $T \subseteq T^2$ with the same root, such that the restricted fragmentation map $\mathcal{G} = \mathcal{S}|_T$ satisfies the following properties.

- $(\partial\mathcal{G}_{r(T)}, d)$ embeds into an ultrametric space with distortion D.
- Every non-leaf vertex $u \in T \smallsetminus \mathcal{L}(T)$ satisfies

$$\sum_{v\in\mathbf{p}_T^{-1}(u)} w_S(v)^{\theta\left(\frac{1-3\tau}{1+\tau}D\right)} \overset{(66)}{\geqslant} w_S(u)^{\theta\left(\frac{1-3\tau}{1+\tau}D\right)}. \quad (81)$$

Let $G \subseteq T$ be a cut-set of T. Define G_0 to be a subset of G which is still a cut-set and is minimal with respect to inclusion. Assume inductively that we defined a cut-set G_i of T which is minimal with respect to inclusion. Let $v \in G_i$ be such that $\mathrm{depth}_T(v)$ is maximal. Let $u = \mathbf{p}_T(v)$. By the maximality of $\mathrm{depth}_T(v)$, since G_i is a minimal cut-set of T we necessarily have $\mathbf{p}_T^{-1}(u) \subseteq G_i$, i.e., all the siblings of v in T are also in G_i. Note that $G_i' = (G_i \cup \{u\}) \smallsetminus \mathbf{p}_T^{-1}(u)$ is also a cut-set of T, so let G_{i+1} be a subset of G_i' which is still a cut-set of T and is minimal with respect to inclusion. Then,

$$\sum_{v\in G_i} w_S(v)^{\theta\left(\frac{1-3\tau}{1+\tau}D\right)} = \sum_{v\in G_i\smallsetminus\mathbf{p}_T^{-1}(u)} w_S(v)^{\theta\left(\frac{1-3\tau}{1+\tau}D\right)} + \sum_{v\in\mathbf{p}_T^{-1}(u)} w_S(v)^{\theta\left(\frac{1-3\tau}{1+\tau}D\right)} \overset{(81)}{\geqslant} \sum_{v\in G_i'} w_S(v)^{\theta\left(\frac{1-3\tau}{1+\tau}D\right)} \geqslant \sum_{v\in G_{i+1}} w_S(v)^{\theta\left(\frac{1-3\tau}{1+\tau}D\right)}. \quad (82)$$

After finitely many iterations of the above process we will arrive at $G_j = \{r\}$. By concatenating the inequalities (82) we see that

$$\sum_{v\in G} w_S(v)^{\theta\left(\frac{1-3\tau}{1+\tau}D\right)} \geqslant \sum_{v\in G_0} w_S(v)^{\theta\left(\frac{1-3\tau}{1+\tau}D\right)} \geqslant w_S(r)^{\theta\left(\frac{1-3\tau}{1+\tau}D\right)} \overset{(72)\wedge(79)}{=} \mu(X)^{\left(1-\frac{1}{k}\right)^2\theta\left(\frac{1-3\tau}{1+\tau}D\right)}. \quad (83)$$

The desired inequality (70) follows from (83) and (80). Since $\mathcal{S}$ is $\frac{1-3\tau}{2\tau}$-separated and $\left(\frac{2}{1-3\tau}\tau^{-4k^2}, \tau^{-4k^2}\right)$-lacunary (recall that $h = 2k^2$), the same holds true for $\mathcal{G}$ since it is the restriction of $\mathcal{S}$ to the subtree of T^2. □

Lemma 6.8. *Fix $K, \gamma \in (0,\infty)$ and $\theta \in (0,1)$. Let (X,d,μ) be a finite metric measure space. Assume that there exists a (K,γ)-lacunary fragmentation map $\mathcal{G} : T \to 2^X$ such that every leaf $\ell \in \mathcal{L}(T)$ has no siblings, and furthermore $\mathcal{G}_{\mathbf{p}(\ell)} = \mathcal{G}_\ell$. Suppose also that for any* *cut-set G of T we have*

$$\sum_{v\in G}\mu(\mathcal{G}_{\mathbf{p}(v)})^{\theta}\geqslant\mu(X)^{\theta},$$

where as usual, if $v\in T$ is the root then we set $\mathbf{p}(v)=v$. Then for any $\{x_i\}_{i\in I}\subseteq X$ and $\{r_i\}_{i\in I}\subseteq[0,\infty)$ such that the d-balls $\{B_d(x_i,r_i)\}_{i\in I}$ cover $\partial\mathcal{G}_{r(T)}$, we have

$$\sum_{i\in I}\mu\left(B_d\left(x_i,\left(1+2K^2\gamma\right)r_i\right)\right)^{\theta}\geqslant\mu(X)^{\theta}. \tag{84}$$

Proof. Without loss of generality assume that $\partial\mathcal{G}_{r(T)}\cap B_d(x_i,r_i)\neq\emptyset$ for all $i\in I$. Let ρ be the ultrametric induced by $\mathcal{G}$ on $\partial\mathcal{G}_{r(T)}$, as constructed in the proof of Lemma 2.8. Thus for $x,y\in\partial\mathcal{G}_{r(T)}$ we have $\rho(x,y)=\mathrm{diam}_d\left(\mathcal{G}_{\mathrm{lca}(a,b)}\right)$, where $a,b\in\mathcal{L}(T)$ satisfy $\mathcal{G}_a=\{x\}$ and $\mathcal{G}_b=\{y\}$. Note that this definition implies that

$$\forall v\in T,\quad \mathrm{diam}_{\rho}(\partial\mathcal{G}_v)=\mathrm{diam}_d(\mathcal{G}_v). \tag{85}$$

By Lemma 2.8 we know that

$$\forall x,y\in\partial\mathcal{G}_{r(T)},\quad d(x,y)\leqslant\rho(x,y)\leqslant Kd(x,y). \tag{86}$$

For every $i\in I$ choose $y_i\in\partial\mathcal{G}_{r(T)}$ satisfying

$$d(x_i,y_i)=\min_{y\in\partial\mathcal{G}_{r(T)}}d(x_i,y)\leqslant r_i. \tag{87}$$

Then $B_d(y_i,2r_i)\supseteq B_d(x_i,r_i)$. Hence the balls $\{B_d(y_i,2r_i)\}_{i\in I}$ also cover $\partial\mathcal{G}_{r(T)}$. By (86) we have $B_\rho(y_i,2Kr_i)\supseteq B_d(y_i,2r_i)\cap\partial\mathcal{G}_{r(T)}$, so we also know that the ρ-balls $\{B_\rho(y_i,2Kr_i)\}_{i\in I}$ cover $\partial\mathcal{G}_{r(T)}$.

For $i\in I$ choose $v_i\in T$ as follows. If $B_\rho(y_i,2Kr_i)$ is a singleton then v_i is defined to be the leaf of T such that $\mathcal{G}_{v_i}=\{y_i\}$. Otherwise pick v_i to be the highest ancestor of y_i in T such that

$$\mathrm{diam}_{\rho}(\partial\mathcal{G}_{v_i})\leqslant 2Kr_i \tag{88}$$

and v_i has at least two children. Then $B_\rho(y_i,2Kr_i)=\partial\mathcal{G}_{v_i}$. Hence, since $\{B_\rho(y_i,2Kr_i)\}_{i\in I}$ cover $\partial\mathcal{G}_{r(T)}$, we have $\bigcup_{i\in I}\mathcal{L}\left(T_{v_i}\right)=\mathcal{L}(T)$. If we set $G=\{v_i\}_{i\in I}$ then we conclude that G is a cut-set of T. Our assumption therefore implies that

$$\sum_{i\in I}\mu\left(\mathcal{G}_{\mathbf{p}(v_i)}\right)^{\theta}\geqslant\mu(X)^{\theta}. \tag{89}$$

When v_i has at least two children we deduce from the fact that $\mathcal{G}$ is (K,γ)-lacunary that

$$\mathrm{diam}_d\left(\mathcal{G}_{\mathbf{p}(v_i)}\right)\leqslant K\gamma\,\mathrm{diam}_d(\partial\mathcal{G}_{v_i}). \tag{90}$$

(Recall Definition 2.6 with $q=\mathbf{p}(v_i)$ and $u=v_i$.) When v_i is a leaf our assumptions imply that $\mathcal{G}_{v_i}$ and $\mathcal{G}_{\mathbf{p}(v_i)}$ are both singletons, and therefore $\mathrm{diam}_d\left(\mathcal{G}_{\mathbf{p}(v_i)}\right)=\mathrm{diam}_d\left(\partial\mathcal{G}_{v_i}\right)=0$, so (90) holds in this case as well. Hence for every $i\in I$ and $z\in\mathcal{G}_{\mathbf{p}(v_i)}$ we have

$$\begin{aligned}d(z,x_i)\leqslant d(x_i,y_i)+d(z,y_i)&\overset{(87)}{\leqslant} r_i+d(z,y_i)\overset{(\clubsuit)}{\leqslant} r_i+\mathrm{diam}_d\left(\mathcal{G}_{\mathbf{p}(v_i)}\right)\\&\overset{(90)}{\leqslant} r_i+K\gamma\,\mathrm{diam}_d(\partial\mathcal{G}_{v_i})\overset{(86)}{\leqslant} r_i+K\gamma\,\mathrm{diam}_\rho(\partial\mathcal{G}_{v_i})\overset{(88)}{\leqslant} r_i+2K^2\gamma r_i,\end{aligned} \tag{91}$$

where (♣) follows from the fact that $y_i \in \partial\mathcal{G}_{v_i} \subseteq \mathcal{G}_{v_i} \subseteq \mathcal{G}_{\mathbf{p}(v_i)}$. The validity of (91) for all $z \in \mathcal{G}_{\mathbf{p}(v_i)}$ is the same as the inclusion $\mathcal{G}_{\mathbf{p}(v_i)} \subseteq B_d\left(x_i, (1+2K^2\gamma)\, r_i\right)$. Now (84) follows from (89). □

Proof of Theorem 1.5. By Lemma 3.1 we may assume that (X,d,μ) is a finite metric measure space. Fix an integer $\frac{10}{\varepsilon} \leqslant k \leqslant \frac{11}{\varepsilon}$ and set $\tau = \frac{1}{20}$. Then $(1-\varepsilon)\left(1-\frac{1}{k}\right)^{-2} \in (0,1)$, so we can define

$$D = \frac{1+\tau}{1-3\tau}\cdot\theta^{-1}\left(\frac{1-\varepsilon}{\left(1-\frac{1}{k}\right)^2}\right) = \frac{11}{7}\theta^{-1}\left(\frac{1-\varepsilon}{\left(1-\frac{1}{k}\right)^2}\right). \tag{92}$$

Equivalently,

$$\left(1-\frac{1}{k}\right)^2\theta\left(\frac{1-3\tau}{1+\tau}D\right) = 1-\varepsilon.$$

Due to (62), for every $s \in (0,1)$ we have $\theta^{-1}(s) = 2(1-s)^{-1}s^{-s/(1-s)} > 2$. Hence it follows from (92) that $D > 2(1+\tau)/(1-3\tau)$, or equivalently $\tau < (D-2)/(3D+2)$. By (63) we have $\theta^{-1}(s) \leqslant 2e/(1-s)$. Therefore,

$$D \leqslant \frac{11}{7}\cdot\frac{2e}{1-\frac{1-\varepsilon}{(1-\varepsilon/10)^2}} = \frac{42e(10-\varepsilon)^2}{17\varepsilon(8+\varepsilon)} \leqslant \frac{9}{\varepsilon}, \tag{93}$$

where the last inequality in (93) is elementary. The required conclusion now follows from Lemma 6.7 and Lemma 6.8. Note that we get the bound $c_\varepsilon = \tau^{-O(k^2)} = e^{O(1/\varepsilon^2)}$. □

Proof of Theorem 1.9. Again, using Lemma 3.1 we may assume that (X,d,μ) is a finite metric measure space. Apply Lemma 6.7 with $D = 2+\delta$, $k=2$ and $\tau = \delta/9$. Denote the exponent in (70) by $s = \frac{1}{2}\theta\left((9-3\delta)(2+\delta)/(9+\delta)\right)$. By (63) there is a universal constant $c \in (0,\infty)$ such that $s \geqslant t$, where $t = c\delta/\log(1/\delta)$. Let $\mathcal{G}: T \to 2^X$ be fragmentation tree obtained using Lemma 6.7, and let G be a cut-set in T. Then by (70) we have,

$$\left(\sum_{v\in G}\mu\left(\mathbf{p}_T(v)\right)^t\right)^{1/t} \geqslant \left(\sum_{v\in G}\mu\left(\mathbf{p}_T(v)\right)^s\right)^{1/s} \geqslant \mu(X).$$

We can therefore apply Lemma 6.8 with $\theta = t$, $K = 2\tau^{-16}/(1-3\tau)$ and $\gamma = \tau^{-16}$, obtaining Theorem 1.9. Note that this shows that c'_δ can be taken to be a constant multiple of δ^{-16}. □

7. Impossibility results

The purpose of this section is to prove the second part of Theorem 1.4 and Theorem 1.8. In both cases the goal is to construct a metric space having the property that all its "almost Euclidean" subsets have small Hausdorff dimension. We will do so by gluing together the finite examples from [2]: in the high distortion regime corresponding to Theorem 1.4 these building blocks are expander graphs, and in the low distortion regime corresponding to Theorem 1.8 these building blocks arise from dense random graphs. The gluing procedure, which is an infinitary variant of the "metric composition" method from [2], starts with a sequence of finite metric spaces and joins them in a tree-like fashion. The details of the construction are contained in Section 7.1 below, and the specializations to prove Theorem 1.4 and Theorem 1.8 are described in Section 7.2 and Section 7.3, respectively.

7.1. **Trees of metric spaces.** Fix $\{n_k\}_{k=0}^\infty \subseteq \mathbb{N}$ with $n_0 = 1$ and $n_k > 1$ for $k \geqslant 1$. Fix also $\{\delta_k\}_{k=1}^\infty \subseteq (0,\infty)$. Assume that for each $k \in \mathbb{N}$ we are given a metric d_k on $\{1,\ldots,n_k\}$ with

$$\operatorname{diam}_{d_k}(\{1,\ldots,n_k\}) = 1 \quad \text{and} \quad \min_{i,j\in\{1,\ldots,n_k\}} d_k(i,j) = \delta_k. \tag{94}$$

For distinct $x = (x_k)_{k=1}^\infty, y = (y_k)_{k=1}^\infty \in \prod_{k=1}^\infty \{1,\ldots,n_k\}$ let $k(x,y)$ be the smallest $k \in \mathbb{N}$ such that $x_k \neq y_k$. For $\alpha \in (0,\infty)$ define

$$\rho_\alpha(x,y) = \frac{d_{k(x,y)}\left(x_{k(x,y)}, y_{k(x,y)}\right)}{\prod_{i=0}^{k(x,y)-1} n_i^{1/\alpha}}. \tag{95}$$

Also, set $\rho_\alpha(x,x) = \rho_\alpha(y,y) = 0$.

Remark 7.1. One can visualize the above construction as follows. Let T be the infinite rooted tree such for $i \geqslant 0$ each vertex at depth i in T has exactly n_{i+1} children. Then $\prod_{k=1}^\infty \{1,\ldots,n_k\}$ can be identified with the set of all infinite branches of T. With this identification, the distance ρ_α has the following meaning: given two infinite branches in T, find the vertex $v \in T$ at which they split (i.e., their deepest common vertex). Say that the depth of v is $i-1$. The metric d_i induces a metric space structure on the n_i children of v, and the distance between the two given branches is a multiple of the distance between the two children of v that belong to these branches.

Lemma 7.2. *$(\prod_{k=1}^\infty \{1,\ldots,n_k\}, \rho_\alpha)$ is a compact metric space provided that*

$$\forall k \in \mathbb{N}, \quad \delta_k > \frac{1}{n_k^{1/\alpha}}. \tag{96}$$

Proof. Take $x,y,z \in \prod_{k=1}^\infty \{1,\ldots,n_k\}$. If $k(x,y) = k(y,z) = k(x,z) = k$ then because d_k satisfies the triangle inequality, $\rho_\alpha(x,z) \leqslant \rho_\alpha(x,y) + \rho_\alpha(y,z)$. If $k(x,y) > k(x,z)$ then necessarily $k(x,z) = k(y,z) = k$ and $x_k = y_k$. Hence $\rho_\alpha(x,z) = \rho_\alpha(x,y) \leqslant \rho_\alpha(x,y) + \rho_\alpha(y,z)$. The remaining case $k(x,z) > k(x,y)$ is dealt with as follows.

$$\rho_\alpha(x,z) = \frac{d_{k(x,z)}\left(x_{k(x,z)}, z_{k(x,z)}\right)}{\prod_{i=0}^{k(x,z)-1} n_i^{1/\alpha}} \overset{(94)}{\leqslant} \frac{1}{\prod_{i=0}^{k(x,z)-1} n_i^{1/\alpha}} \overset{(96)}{\leqslant} \frac{\delta_{k(x,y)}}{\prod_{i=0}^{k(x,y)-1} n_i^{1/\alpha}} \overset{(94)}{\leqslant} \rho_\alpha(x,y).$$

This proves the triangle inequality. Compactness follows from Tychonoff's theorem since ρ_α induces the product topology on $\prod_{k=1}^\infty \{1,\ldots,n_k\}$. $\square$

Lemma 7.3. *Assume that in addition to* (96) *we have*

$$\lim_{k\to\infty} \frac{\log(1/\delta_k)}{\sum_{i=1}^{k-1} \log n_i} = 0. \tag{97}$$

Then

$$\dim_H\left(\prod_{k=1}^\infty \{1,\ldots,n_k\}, \rho_\alpha\right) = \alpha.$$

Proof. Define $\Delta = \dim_H(\prod_{k=1}^\infty \{1,\ldots,n_k\}, \rho_\alpha)$. The fact that $\Delta \leqslant \alpha$ is simple. Indeed, for $k \in \mathbb{N}$ consider the sets $\{B_x^k\}_{x\in\prod_{i=1}^k \{1,\ldots,n_i\}}$ given by

$$B_x^k = \left(\prod_{i=1}^k \{x_i\}\right) \times \left(\prod_{i=k+1}^\infty \{1,\ldots,n_i\}\right). \tag{98}$$

Then $\operatorname{diam}_{\rho_\alpha}(B_x^k) = \prod_{i=1}^k n_i^{-1/\alpha}$ and $\{B_x^k\}_{x\in\prod_{i=1}^k\{1,\ldots,n_i\}}$ cover $\prod_{i=1}^\infty\{1,\ldots,n_i\}$. Hence for $\beta > \alpha$ the β-Hausdorff content of $(\prod_{i=1}^\infty\{1,\ldots,n_i\},\rho_\alpha)$ can be estimated as follows.

$$\mathcal{H}_\infty^\beta\left(\prod_{i=1}^\infty\{1,\ldots,n_i\},\rho_\alpha\right) \leqslant \inf_{k\in\mathbb{N}} \sum_{x\in\prod_{i=1}^k\{1,\ldots,n_i\}} \frac{1}{\prod_{i=0}^k n_i^{\beta/\alpha}} = \inf_{k\in\mathbb{N}} \frac{1}{\prod_{i=0}^k n_i^{(\beta-\alpha)/\alpha}} = 0.$$

Thus $\Delta \leqslant \alpha$.

We now pass to the proof of $\Delta \geqslant \alpha$. We first prove the following preliminary statement. Assume that $x^1,\ldots,x^m \in \prod_{i=1}^\infty\{1,\ldots,n_i\}$ and $k_1,\ldots,k_m \in \mathbb{N}\cup\{0\}$ are such that $\{B_{x^j}^{k_j}\}_{j=1}^m$ cover $\prod_{i=1}^\infty\{1,\ldots,n_i\}$, where B_x^k is given in (98) and we use the convention $B_x^0 = \prod_{i=1}^\infty\{1,\ldots,n_i\}$. We claim that this implies that

$$\sum_{j=1}^m \frac{1}{\prod_{i=0}^{k_j} n_i} \geqslant 1. \tag{99}$$

The proof is by induction on m. If $m = 1$ then $k_1 = 0$ and (99) follows. Assume that $m \geqslant 2$, no subset of $\{B_{x^j}^{k_j}\}_{j=1}^m$ covers $\prod_{i=1}^\infty\{1,\ldots,n_i\}$, and that $k_1 \leqslant k_2 \leqslant \cdots \leqslant k_m$. For every $y \in \{1,\ldots,n_{k_m}\}$ let $x^{k_m}(y) \in \prod_{i=1}^\infty\{1,\ldots,n_i\}$ have y in the k_m'th coordinate, and coincide with x^{k_m} in all other coordinates. The sets $\{B_{x^{k_m}(y)}^{k_m}\}_{y\in\{1,\ldots,n_{k_m}\}}$ are pairwise disjoint, and are either contained in or disjoint from $B_{x^j}^{k_j}$ for each $j \in \{1,\ldots,m\}$. Hence, by the minimality of the cover $\{B_{x^j}^{k_j}\}_{j=1}^m$ we have $k_m = k_{m-1} = \cdots = k_{m-n_{k_m}+1}$ and $\{B_{x^j}^{k_j}\}_{j=m-n_{k_m}+1}^{k_m} = \{B_{x^{k_m}(y)}^{k_m}\}_{y\in\{1,\ldots,n_{k_m}\}}$. This implies that

$$\sum_{j=1}^m \frac{1}{\prod_{i=0}^{k_j} n_i} = \sum_{j=1}^{m-n_{k_m}} \frac{1}{\prod_{i=0}^{k_j} n_i} + \sum_{y\in\{1,\ldots,n_{k_m}\}} \frac{1}{\prod_{i=0}^{k_m} n_i} = \sum_{j=1}^{m-n_{k_m}} \frac{1}{\prod_{i=0}^{k_j} n_i} + \frac{1}{\prod_{i=0}^{k_m-1} n_i}.$$

The induction hypothesis applied to $\{B_{x^j}^{k_j}\}_{j=1}^{m-n_{k_m}} \cup \{B_{x^m}^{k_m-1}\}$ concludes the proof of (99).

Fix $\beta \in (0,\alpha)$. Due to (97) there exists $C \in (0,\infty)$ such that for all $k \in \mathbb{N}$,

$$\frac{1}{\delta_k} \leqslant C\prod_{i=1}^{k-1} n_i^{(\alpha-\beta)/\alpha^2}. \tag{100}$$

Let $\{B_{\rho_\alpha}(x^j,r_j)\}_{j\in J}$ be a family of balls that covers $\prod_{i=1}^\infty\{1,\ldots,n_i\}$. We will show that

$$\sum_{j\in J} r_j^\beta \geqslant \frac{1}{C^\alpha}. \tag{101}$$

This would mean that $\mathcal{H}_\infty^\beta(\prod_{i=1}^\infty\{1,\ldots,n_i\},\rho_\alpha) > 0$ for all $\beta \in (0,\alpha)$, proving that $\Delta \geqslant \alpha$. By compactness of $(\prod_{i=1}^\infty\{1,\ldots,n_i\},\rho_\alpha)$ it suffices to prove (101) when J is finite. For every $j \in J$ choose $k_j \in \mathbb{N}$ such that $\prod_{i=0}^{k_j} n_i^{-1/\alpha} < r_j \leqslant \prod_{i=0}^{k_j-1} n_i^{-1/\alpha}$. Define

$$r_j^* = \begin{cases} \prod_{i=1}^{k_j-1} n_i^{-1/\alpha} & \text{if } \delta_{k_j}\prod_{i=1}^{k_j-1} n_i^{-1/\alpha} \leqslant r_j \leqslant \prod_{i=1}^{k_j-1} n_i^{-1/\alpha}, \\ \prod_{i=1}^{k_j} n_i^{-1/\alpha} & \text{if } \prod_{i=1}^{k_j} n_i^{-1/\alpha} < r_j < \delta_{k_j}\prod_{i=1}^{k_j-1} n_i^{-1/\alpha}. \end{cases} \tag{102}$$

If $\delta_{k_j}\prod_{i=1}^{k_j-1} n_i^{-1/\alpha} \leqslant r_j \leqslant \prod_{i=1}^{k_j-1} n_i^{-1/\alpha}$ then $r_j^* \geqslant r_j$, hence $B_{\rho_\alpha}(x^j,r_j) \subseteq B_{\rho_\alpha}(x^j,r_j^*) = B_{x^j}^{k_j-1}$. Also, observe that ρ_α does not take values in the interval $\left(\prod_{i=1}^{k_j} n_i^{-1/\alpha}, \delta_{k_j}\prod_{i=1}^{k_j-1} n_i^{-1/\alpha}\right)$.

This implies that $B_{\rho_\alpha}(x^j,r_j)=B_{\rho_\alpha}(x^j,r_j^*)=B_{x^j}^{k_j}$ when $\prod_{i=1}^{k_j} n_i^{-1/\alpha}<r_j<\delta_{k_j}\prod_{i=1}^{k_j-1} n_i^{-1/\alpha}$. We deduce that the balls $B_{\rho_\alpha}(x^j,r_j^*)$ cover $\prod_{i=1}^\infty\{1,\ldots,n_i\}$, and they are all sets of the form B_x^k. It therefore follows from (99) that

$$1\leqslant\sum_{j\in J}\left(r_j^*\right)^\alpha\overset{(102)}{\leqslant}\sum_{j\in J}\left(\frac{r_j}{\delta_{k_j}}\right)^\alpha\overset{(100)}{\leqslant}C^\alpha\sum_{j\in J}r_j^\alpha\prod_{i=0}^{k_j-1}n_i^{(\alpha-\beta)/\alpha}\leqslant C^\alpha\sum_{j\in J}r_j^\beta, \tag{103}$$

where in the last inequality of (103) we used the fact that $r_j\leqslant\prod_{i=0}^{k_j-1}n_i^{-1/\alpha}$. $\square$

Remark 7.4. A less direct way to prove the bound $\Delta\geqslant\alpha$ in Lemma 7.3 is to define $\mu(B_x^k)=\prod_{i=0}^k n_i^{-1}$ and to argue that the Carathéodory extension theorem applies here and yields an extension of μ to a Borel measure on $\prod_{i=1}^\infty\{1,\ldots,n_i\}$. One can then show analogously to (103) that this measure is a β-Frostman measure for $(\prod_{i=1}^\infty\{1,\ldots,n_i\},\rho_\alpha)$.

In what follows we say that a property $\mathscr{P}$ of metric spaces is a metric property if whenever $(X,d_X)\in P$ and (Y,d_Y) is isometric to (X,d_X) then also $(Y,d_Y)\in\mathscr{P}$. We say that $\mathscr{P}$ is hereditary if whenever $(X,d)\in\mathscr{P}$ and $Y\subseteq X$ then also $(Y,d)\in\mathscr{P}$. Finally, we say that $\mathscr{P}$ is dilation-invariant if whenever $(X,d)\in\mathscr{P}$ and $\lambda\in(0,\infty)$ also $(X,\lambda d)\in\mathscr{P}$.

Theorem 7.5. *Fix $\alpha>0$. Fix also $\{n_k\}_{k=0}^\infty\subseteq\mathbb{N}$ with $n_0=1$ and $n_k>1$ for $k\geqslant1$, a sequence $\{\delta_k\}_{k=1}^\infty\subseteq(0,\infty)$, and for each $k\in\mathbb{N}$ a metric d_k on $\{1,\ldots,n_k\}$. Assume that* (94)*,* (96) *and* (97) *hold true. Then there exists a metric space (Y,ρ) with $\dim_H(Y,\rho)=\alpha$ that satisfies the following property. Let $\{\mathscr{P}_k\}_{i=1}^\infty$ is a non-decreasing (with respect to inclusion) sequence of hereditary dilation-invariant metric properties and for every $k\in\mathbb{N}$ let m_k be the cardinality of the largest subset S of $\{1,\ldots,n_k\}$ such that (S,d_k) has the property $\mathscr{P}_k$. Then every $Z\subseteq Y$ that has the property $\bigcup_{k=1}^\infty\mathscr{P}_k$ satisfies*

$$\dim_H(Z,\rho)\leqslant\limsup_{k\to\infty}\frac{\alpha\log m_k}{\log n_k}.$$

Proof. Take $(Y,\rho)=(\prod_{k=1}^\infty\{1,\ldots,n_k\},\rho_\alpha)$, where ρ_α is given in (95). By Lemma 7.3 and Lemma 7.2 we know that (Y,ρ) is a compact metric space of Hausdorff dimension α.

Assume that $Z\subseteq Y$ and $(Z,\rho)\in\mathscr{P}_K$ for some $K\in\mathbb{N}$. Since $\{\mathscr{P}_k\}_{i=1}^\infty$ are non-decreasing properties, we know that $(Z,\rho)\in\mathscr{P}_k$ for all $k\geqslant K$. For every $x\in Y$ and $k\in\mathbb{N}$ denote

$$S_x^k=\left\{j\in\{1,\ldots,n_k\}:\ Z\cap\left(\left(\prod_{i=1}^{k-1}\{x_i\}\right)\times\{j\}\times\left(\prod_{i=k+1}^\infty\{1,\ldots,n_i\}\right)\right)\neq\emptyset\right\}.$$

If $j\in S_x^k$ choose $x^k(j)\in Z$ whose first $k-1$ coordinates coincide with the corresponding coordinates of x, and whose k'th coordinate equals j. Then $\{x^k(j)\}_{j\in S_x^k}$ is a subset of Z whose metric is isometric to a dilation of the metric d_k on S_x^k. Since $\mathscr{P}_k$ is a hereditary dilation-invariant metric property, it follows that $S_x^k\in\mathscr{P}_k$ for all $k\geqslant K$. Hence $|S_x^k|\leqslant m_k$. Let Z_k be the projection of the set Z onto the first k coordinates, i.e., the set of all $x\in\prod_{i=1}^k\{1,\ldots,n_i\}$ such that $B_x^k\cap Z\neq\emptyset$, where B_x^k is given in (98). Then it follows by induction that for every $k\geqslant K$ we have $|Z_k|\leqslant\prod_{i=0}^{K-1}n_i\cdot\prod_{i=K}^k m_i$. Denote $\gamma=\limsup_{k\to\infty}\frac{\alpha\log m_k}{\log n_k}$. If $\beta>\gamma$ then there exists $K'\geqslant K$ such that for every $k\geqslant K'$ we have $m_k\leqslant n_k^{(\beta+\gamma)/(2\alpha)}$. Since the sets

$\{B_x^k\}_{x\in Z_k}$ cover Z and have ρ-diameter $\prod_{i=0}^k n_i^{-1/\alpha}$,

$$\mathcal{H}_\infty^\beta(Z,\rho)\leqslant \inf_{k\geqslant K'}\sum_{x\in Z_k}\frac{1}{\prod_{i=0}^k n_i^{\beta/\alpha}}\leqslant \inf_{k\geqslant K'}\frac{\prod_{i=0}^{K-1}n_i\cdot\prod_{i=K}^k m_i}{\prod_{i=0}^k n_i^{\beta/\alpha}}$$
$$\leqslant \inf_{k\geqslant K'}\frac{\prod_{i=0}^{K'-1}n_i\cdot\prod_{i=K'}^k n_i^{(\beta+\gamma)/(2\alpha)}}{\prod_{i=0}^k n_i^{\beta/\alpha}}\leqslant \inf_{k\geqslant K'}\frac{\prod_{i=0}^{K'-1}n_i}{\prod_{i=K'}^k n_i^{(\beta-\gamma)/(2\alpha)}}=0.$$

Hence $\dim_H(Z,\rho)\leqslant\gamma$. □

Corollary 7.6. *Fix an integer $n\geqslant 2$. Let (X,d) be an n-point metric space, and assume that $\Psi\in(0,\infty)$ satisfies*
$$\Psi\cdot\min_{\substack{x,y\in X\\ x\neq y}} d(x,y)>\operatorname{diam}(X).$$
Then there exists a compact metric space (Y,ρ) with $\dim_H(Y,\rho)=\log_\Psi n$ that has the following property. Fix $m\in\{1,\dots,n-1\}$ and assume that $\mathscr{P}$ is a hereditary dilation-invariant metric property such that the largest subset of X having the property $\mathscr{P}$ is of size m. Then $\dim_H(Z,\rho)\leqslant\log_\Psi m$ for every $Z\subseteq Y$ with property $\mathscr{P}$.

Proof. By rescaling assume that $\operatorname{diam}(X)=1$. Now apply Theorem 7.5 with $X_i=X$, $n_i=n$, $\mathscr{P}_i=\mathscr{P}$ and $\alpha=\log_\Psi n$. □

7.2. **Expander fractals.** It is shown in [2] that there is $c\in(0,\infty)$ such that for any $n\in\mathbb{N}$ there exists an n-point metric space X_n such that for every $\varepsilon\in(0,1)$ all the subsets of X_n of cardinality greater than $n^{1-\varepsilon}$ incur distortion greater than c/ε in any embedding into Hilbert space. In fact, the spaces X_n are the shortest-path metrics on expander graphs, implying that $\operatorname{diam}X_n\leqslant C\log n$ for some $C\in(0,\infty)$ and all $n\in\mathbb{N}$ (see [8]). We will apply Corollary 7.6 to these spaces, thus obtaining compact metric spaces that can be called "expander fractals". The property $\mathscr{P}$ that will be used is "X embeds with distortion c/ε into Hilbert space", which is clearly a hereditary dilation-invariant metric property.

Proof of the second part of Theorem 1.4. Let c,C,X_n be as above. Fix $\alpha>0$ and choose an integer $n\geqslant 2$ such that $n^{1/\alpha}>C\log n$. We may therefore use Corollary 7.6 with $X=X_n$ and $\Psi=n^{1/\alpha}$. The resulting compact metric space (Y,ρ) will then have Hausdorff dimension equal to α. For every $\varepsilon\in(0,1)$ let $\mathscr{P}_\varepsilon$ be the property "X embeds with distortion c/ε into Hilbert space". Then all the subsets Z of Y that embed into Hilbert space with distortion c/ε satisfy $\dim_H(Z,\rho)\leqslant\log_\Psi\left(n^{1-\varepsilon}\right)=(1-\varepsilon)\alpha=(1-\varepsilon)\dim_H(Y,\rho)$. □

7.3. $G\,(n,1/2)$ **fractals.** It is shown in [2] that there exists $K\in(1,\infty)$ such that for any $n\in\mathbb{N}$ there exists an n-point metric space W_n such that for every $\delta\in(0,1)$ any subset of W_n of size larger than $2\log_2 n+K\left(\delta^{-2}\log(2/\delta)\right)^2$ must incur distortion at least $2-\delta$ when embedded into Hilbert space. The space W_n comes from a random construction: consider a random graph G on n vertices, drawn from the Erdős-Reyni model $G(n,1/2)$ (thus every edge is present independently with probability 1/2). The space W_n is obtained from G by declaring two vertices that are joined by an edge to be at distance 1, and two distinct vertices that are not joined by an edge are declared to be at distance 2. Therefore the positive distances in W_n are either 1 or 2. This description of W_n is implicit in [2] but follows immediately from the proof of [2]; see [3] for an alternative proof of this fact (yielding a worse asymptotic dependence on δ that is immaterial for our purposes). We will apply Corollary 7.6 to $\{W_n\}_{n=2}^{\infty}$, thus obtaining compact metric spaces that can be called "$G(n,1/2)$ fractals".

Proof of Theorem 1.8. Let K and $\{W_n\}_{n=2}^{\infty}$ be as in the above discussion. Set $X_n = W_{n+\lceil 3^\alpha \rceil}$. Hence $|X_n|^{1/\alpha} > 2$. Let $\mathscr{P}_n$ be the property "X embeds with distortion $2 - 1/\log\log n$ into Hilbert space". Then $\{P_n\}_{n\geqslant 20}$ is a non-decreasing sequence of hereditary dilation-invariant metric properties. Moreover, $\bigcup_{n\geqslant 20} \mathscr{P}_n$ is the property "X embeds into Hilbert space with distortion smaller than 2". By the above discussion, letting m_n be the size of the largest subset of X_n that has the property $\mathscr{P}_n$, we have $m_n \leqslant 2\log_2 n + O\left((\log\log)^2(\log\log\log n)^2\right)$. Hence $\limsup_{n\to\infty} \frac{\alpha \log m_n}{\log |X_n|} = 0$. By Theorem 7.5 it follows there exists a compact metric space (Y,ρ) with Hausdorff dimension α such that all of its subsets with positive Hausdorff dimension do not have property $\bigcup_{n\geqslant 20} \mathscr{P}_n$, namely any embeding of such a subset into Hilbert space must incur distortion at least 2. $\square$

Acknowledgements. We are grateful to Terence Tao for sharing with us his initial attempts to solve Question 1.3.

References

[1] Y. Bartal, B. Bollobás, and M. Mendel. Ramsey-type theorems for metric spaces with applications to online problems. *J. Comput. System Sci.*, 72(5):890–921, 2006.

[2] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric Ramsey-type phenomena. *Ann. of Math. (2)*, 162(2):643–709, 2005.

[3] Y. Bartal, N. Linial, M. Mendel, and A. Naor. Some low distortion metric Ramsey problems. *Discrete Comput. Geom.*, 33(1):27–41, 2005.

[4] A. Blum, H. Karloff, Y. Rabani, and M. Saks. A decomposition theorem for task systems and bounds for randomized server problems. *SIAM J. Comput.*, 30(5):1624–1661 (electronic), 2000.

[5] J. Bourgain, T. Figiel, and V. Milman. On Hilbertian subsets of finite metric spaces. *Israel J. Math.*, 55(2):147–152, 1986.

[6] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[7] L. Carleson. *Selected problems on exceptional sets*. Van Nostrand Mathematical Studies, No. 13. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London, 1967.

[8] F. R. K. Chung. Diameters and eigenvalues. *J. Amer. Math. Soc.*, 2(2):187–196, 1989.

[9] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, pages 123–160. Jerusalem Academic Press, Jerusalem, 1961.

[10] A. Grothendieck. Sur certaines classes de suites dans les espaces de Banach et le théorème de Dvoretzky-Rogers. *Bol. Soc. Mat. São Paulo*, 8:81–110 (1956), 1953.

[11] J. D. Howroyd. On dimension and on the existence of sets of finite positive Hausdorff measure. *Proc. London Math. Soc. (3)*, 70(3):581–604, 1995.

[12] H. Karloff, Y. Rabani, and Y. Ravid. Lower bounds for randomized k-server and motion-planning algorithms. *SIAM J. Comput.*, 23(2):293–312, 1994.

[13] P. Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Fractals and rectifiability.

[14] M. Mendel and A. Naor. Nonlinear Dvoretzky theory. In preparation.

[15] M. Mendel and A. Naor. Ramsey partitions and proximity data structures. *J. Eur. Math. Soc.*, 9(2):253–275, 2007.

[16] V. Milman and G. Schechtman. An "isomorphic" version of Dvoretzky's theorem. II. In *Convex geometric analysis (Berkeley, CA, 1996)*, volume 34 of *Math. Sci. Res. Inst. Publ.*, pages 159–164. Cambridge Univ. Press, Cambridge, 1999.

[17] V. D. Milman. A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies. *Funkcional. Anal. i Priložen.*, 5(4):28–37, 1971.

[18] P. Mörters and Y. Peres. *Brownian motion*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010. With an appendix by Oded Schramm and Wendelin Werner.

[19] A. Naor and T. Tao. Scale-oblivious metric fragmentation and the nonlinear Dvoretzky theorem, 2010. Preprint available at `http://arxiv.org/abs/1003.4013`. To appear in Israel J. Math.

[20] G. Schechtman. Two observations regarding embedding subsets of Euclidean spaces in normed spaces. *Adv. Math.*, 200(1):125–135, 2006.

[21] M. Talagrand. Regularity of Gaussian processes. *Acta Math.*, 159(1-2):99–149, 1987.

[22] M. Talagrand. *The generic chaining*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005. Upper and lower bounds of stochastic processes.

[23] I. A. Vestfrid and A. F. Timan. A universality property of Hilbert spaces. *Dokl. Akad. Nauk SSSR*, 246(3):528–530, 1979.

Mathematics and Computer Science Department, Open University of Israel, 1 University Road, P.O. Box 808 Raanana 43107, Israel
E-mail address: `mendelma@gmail.com`

Courant Institute, New York University, 251 Mercer Street, New York NY 10012, USA
E-mail address: `naor@cims.nyu.edu`