Local Embeddings of Metric Spaces

Ittai Abraham* Yair Bartal† Ofer Neiman‡

March 24, 2013

Abstract

In many application areas, complex data sets are often represented by some metric space and metric embedding is used to provide a more structured representation of the data. In many of these applications much greater emphasis is put on the preserving the *local* structure of the original space than on maintaining its complete structure. This is also the case in some networking applications where "small world" phenomena in communication patterns has been observed. Practical study of embedding has indeed involved with finding embeddings with this property. In this paper[1] we initiate the study of *local embeddings* of metric spaces and provide embeddings with distortion depending solely on the local structure of the space.

1 Introduction

The field of metric space embedding studies embeddings that "faithfully" preserve distances of the source space in the host space. There are many ways to formally measure the "faithfulness" of an embedding. In this paper we suggest a new and quite natural paradigm of *local distortion embeddings*: I.e. embeddings that preserve the local structure of the space, distances of close neighbors are preserved better than those of distant neighbors.

Metric embedding has emerged as powerful tool in several applications areas. Typically, an embedding takes a "complex" metric space and maps it into a "simpler" one. For example embedding of metric spaces into trees and ultrametrics found a large number of algorithmic applications (e.g. [Ind01]). In many fields that use high dimensional data (e.g. computer vision, computational biology, machine learning, networking, statistics, and mathematical psychology), embeddings are used to map complex data sets into simpler and more compact representations [BN03]. In distributed network settings, embedding has been used to map the Internet latencies into a simpler metric structure. Often, the embedding can then be distributed as a labeling scheme in a distributed system [Pel00, CCRK04].

In many important applications of embedding, preserving the distances of nearby points is much more important than preserving all distances. Indeed, it is sometimes the case in distance estimation, that determining the distance of nearby objects can be done easily, while far away objects may just be labeled as "far" and only a rough estimate of the distance between them will be given. Thus large distances may already incorporate an inherently larger error factor. In such scenarios it is natural to seek local embeddings that maintain only distances of close by neighbors. Indeed both [BN03] and [XSB06] study low dimensional embeddings that maintain distances only to the k nearest neighbors.

The revolution of large scale Social Networking in the Internet has increased the interest in new areas of research that emerged from issues in the border of Sociology and Network theory. One aspect studied by Kleinberg [Kle00] is the algorithmic aspects of the "small world" phenomena: how messages are greedily routed in networks that arise from a social and geographical structure. In this model the network is assumed to have a *local* property: the probability of

*Microsoft Research SVC. Email: ittaia@microsoft.com.

†Hebrew University. Email: yair@cs.huji.ac.il. Supported in part by a grant from the Israeli Science Foundation (1609/11).

‡Ben-Gurion University of the Negev, Beer-Sheva, Israel. Email: neimano@cs.bgu.ac.il. Supported by the Israeli Science Foundation grant number (523/12) and by the European Unions Seventh Framework Programme (FP7/2007-2013) under grant agreement $n^\circ 303809$. Most of the research on this paper was carried out while the author was a student at the Hebrew University and was supported in part by a grant from the Israeli Science Foundation (195/02).

[1] This paper is a full version based on the conference papers [ABN07b, ABN09].

choosing a close neighbor as an associate is larger than that of choosing a far away neighbor. Specifically, the probability of choosing a neighbor is inversely proportional to its distance from the source. Liben-Nowell *et. al.* [LNK+05] consider a related model where the probability of choosing the k-th nearest neighbor is chosen proportional to $\propto \frac{1}{k^\alpha}$ for some parameter $\alpha > 1$, they validate this model experimentally. A person would have more interaction with his close associates than with far away ones. In the context of using metric space embedding in "small world" networks it is natural to require that the embedding of a close neighbor would be better than that of a far away neighbor.

Kleinberg, Slivkins, and Wexler [KSW04] study network embedding as a means to provide distance estimation of the Internet latency without need to measure all distances. They note a discrepancy between theory and practice: while known theoretical embedding results guarantee very weak bounds, practical network coordinates perform quite well. In order to overcome this gap, the authors suggest to study *embeddings with slack*, where the distortion bounds are provided only for distant neighbors but not of close by ones. Strong results have been obtained in this model and its generalization [ABN06, ABC+05]. However, for certain applications, one might claim that preserving only distances to far-away neighbors defeats the purpose. For example, an Internet application that is induced by a social structure might interact mostly amongst local neighbors and so on. Our study on local embedding can be viewed as addressing the same question of [KSW04] when indeed preserving local distances is more important than preserving far away distances.

A related notion of *nearest neighbor preserving embedding* was studied by [IN07], where the goal is to maintain that the (approximate) nearest neighbor in the embedding was an approximate nearest neighbor in the original metric. However, this notion is focused only on the first nearest neighbor and moreover does not provide a distortion guarantee, and therefore only useful in Nearest Neighbor type applications.

Our main results show that embeddings that preserve distances between k nearest neighbors can have distortion bounded by a function of k alone, essentially replacing $O(\log n)$ with $O(\log k)$ in known theorems. Moreover, this is also true for the dimension of the host space if there is no guarantee whatsoever on the far pairs (which is still useful in applications such as local distance labeling). We study a stronger version of local embedding, where we demand that the far points will not collapse too much.

In the context of data compression, our results can be viewed as a new type of dimension reduction technique. Typically, dimension reduction causes a uniform error over all points. The celebrated Johnson Lindenstrauss dimension reduction Lemma [JL84] states that high dimensional data set X in ℓ_2 can be faithfully mapped into $O(\log |X|)$ dimensions. Our techniques allow to map metric spaces into *constant* dimensional Euclidean space which preserves distances between all nearby neighbor points, i.e. the local structure of the space, with constant distortion. In the conference version of this work, we asked if there exists a local version of dimension reduction, in which the distortion may be arbitrarily small and the dimension depends solely on k (potentially $O(\log k)$). Schechtman and Shraibman [SS09] showed that this is impossible for $k \geq 2$: Whenever the distortion is smaller than $3/2$, the dimension must be at least $\Omega(\log n)$. We do show local dimension reduction for ultrametrics in any ℓ_p space. In a recent work by Bartal, Recht and Schulman [BRS11], it is shown that under certain relaxation of the demand from the local embedding, a dimension reduction into dimension $O(\log k)$ exists, in particular this holds under a certain weak growth rate condition on the metric space.

In large scale systems it is often the case that one wants to maintain a compact data structure known as a *Distance Oracle* [Pel00]. More demanding tasks are *name independent compact routing schemes* where the name of the node is independent of its location [AP90] and *mobile user schemes* which are competitive distributed protocols for routing when the target may be mobile. In all these settings it is desirable to obtain improved results for close by neighbors.

1.1 Local Embeddings

We now formally define new notions of *local distortion*. Fix a metric space (X, d). For any point x let $<_x$ be a linear order relation on the points in $X \setminus \{x\}$ such that for any $u, v \in X \setminus \{x\}$ if $d(u, x) < d(v, x)$ then $u <_x v$. For any $k \in \mathbb{N}$ let $N_k(x)$ be the set of first k elements of $X \setminus \{x\}$ according to $<_x$, i.e., $N_k(x)$ is the set of k nearest neighbors of x. For any $x \in X$ let $\bar{N}_k(x) = \{y \mid y \in N_k(x) \wedge x \in N_k(y)\}$. Let $r_k(x)$ be minimal radius such that $N_k(x) \subseteq B(x, r_k(x))$.

Definition 1. *Let (X, d_X) be a metric space on n points, (Y, d_Y) a target metric space and $k \in \mathbb{N}$, let $f : X \to Y$ be* *an embedding.*

- f *is* non-expansive *if for any* $u, v \in X$, $d_Y(f(u), f(v)) \leq d_X(u, v)$.
- f *is an embedding with* k-local distortion α *if* f *is non-expansive and for any* $u, v \in X$ *such that* $v \in N_k(u)$,

$$d_Y(f(u), f(v)) \geq \frac{d_X(u, v)}{\alpha}.$$

- f *is an embedding with* strong k-local distortion α *if* f *is non-expansive and for any* $u, v \in X$,

$$d_Y(f(u), f(v)) \geq \frac{\min\{d_X(u, v), r_k(u)\}}{\alpha}$$

- f *is an embedding with (strong)* scaling *local distortion* α*, for a non-decreasing function* $\alpha : \mathbb{N} \to \mathbb{R}_+$*, if* f *has (strong)* k*-local distortion* $\alpha(k)$*, for all* $k \in \mathbb{N}$ *simultaneously.*
- *Given a distribution* $\mathcal{D}$ *on maps* $f : X \to Y$*, we say that* $\mathcal{D}$ *has probabilistic (strong)* $\{k, \textit{scaling}\}$*-local distortion if the appropriate lower bound holds for all embedding in the support of* $\mathcal{D}$ *and the appropriate upper bound holds in expectation over* $\mathcal{D}$.
- *Given a set* $S \subseteq X$*, we say that the embedding* f *has (strong)* $\{k, \textit{scaling}\}$*-local distortion with respect to* S *if the appropriate upper and lower bounds hold for all pairs* (u, v) *such that* $u \in S$.

We also study a related notion of *proximity distortion*.

Definition 2. *Let* (X, d_X) *be a metric space on* n *points with* $\min_{x,y \in X}\{d_X(x, y)\} \geq 1$*, let* (Y, d_Y) *be a target metric space, let* $t \geq 1$*, let* $f : X \to Y$ *be an embedding.*

- f *is an embedding with* t-proximity distortion α *if for any* $u, v \in X$ *such that* $d(u, v) \leq t$,

$$d_X(u, v) \geq d_Y(f(u), f(v)) \geq \frac{d_X(u, v)}{\alpha}.$$

- f *is an embedding with* scaling proximity distortion α*, for non-decreasing function* $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$*, if it has* t*-proximity distortion* $\alpha(t)$*, for all* t *simultaneously.*

1.2 Overview of Results

We begin by providing some basic results in this model. Theorem 1 shows that any metric space can be embedded into a single tree (in fact, an ultrametric) with scaling local distortion k. In Theorem 2 it is shown that strong k-local embeddings with distortion $O(\log k)$ are possible for any k using $O(\log n \cdot \log k)$ dimensions. We remark that a similar result is implicit in [MN04]. In Theorem 3 we give an embedding with strong scaling local distortion of $\tilde{O}(\log k)$ and dimension $O(\log^2 n)$ using a variation of Bourgain's embedding method. The dimension can be improved to $O(\log n)$ as shown in Theorem 12 by using partition-based embedding techniques. It is not hard to see that logarithmic dependence of the dimension on the number of points is necessary for a *strong* local embedding[2].

Another aspect of k-local embeddings is that the dimension can be bounded in terms of k (for *non*-strong local embeddings). In this introductory section we demonstrate this phenomenon for the special case of $k = 1$: Theorem 4 shows that 1-local embeddings into ℓ_p with $\sqrt[p]{3}$ distortion requires only 3 dimensions. We conclude the introductory section with a simple yet surprising result, that local (non-strong) dimension reduction is in fact possible for an equilateral metric (even for adversarial choice of nearest neighbors). This lies in contrast to the standard setting, in which an equilateral metric is the "bad example" for dimension reduction.

[2]Consider an equilateral metric (where all distances are 1); a strong local embedding imposes constraints on all pairs, thus a standard volume argument suggests that the required dimension is $\Omega(\log_\alpha n)$ for distortion α.

After the basic results we study embeddings into ultrametrics and into distributions of ultrametrics. Theorem 5 shows a strong k-local distortion $O(\log k)$ into a distribution of ultrametrics. Its scaling counterpart, Theorem 6 obtains scaling local distortion $\tilde{O}(\log k)$, and worst case distortion $O(\log n)$. The proof of Theorem 6 is unique, all known embeddings into ultrametrics are non-contracting, we provide an embedding which is non-expanding in expectation. This requires to make subtle modifications of known ultrametric construction algorithms. We then prove in Theorem 7 that embedding a graph into a distribution of spanning trees must have large local distortion, that depends on n.

While Theorem 2 provides an embedding into ℓ_p with k-local distortion, its dimension is a function of the size of the data set. Our main result, Theorem 8, does not have the strong local guarantees, and therefore significantly improves the dependence on the dimension. It provides a new form of dimension reduction; an embeddings that require only $O(\log^2 k)$ dimensions and optimal $O(\log k)$ distortion. Some of the novel techniques introduced to prove this theorem are:

- *Bounded cardinality* probabilistic partitions, which are a variation on the standard diameter bounded partitions.
- Assignment of carefully chosen *vector colors* to the clusters of the partitions, instead of the usual $\{0, 1\}$ values, which are used to prevent dependencies.

We combine these with an application of the Lovász local lemma and other embedding tools. To obtain the optimal $O(\log k)$ dimension, we have Theorem 9, which requires that the metric space obeys a weak form of growth bound (formally defined later). Our result shows that the k-local structure of the space can be embedded in its natural dimension which is independent of the size of the original space.

We then turn to dimension reduction with arbitrarily good precision. As mentioned above, [SS09] showed this is impossible in general. Theorem 10 presents a positive result for the family of ultrametrics. In fact, our dimension reduction can be done in any ℓ_p space. As a by product of the techniques used to prove Theorem 10, we are able to show in Theorem 11 a standard dimension reduction for the class of doubling ultrametrics. The idea of having Theorem 11 was influenced by a recent result of Gottlieb and Krauthgamer [GK11], who have obtained a result of a similar flavor.

Using embeddings based on partitions, Theorem 13 provides better local scaling distortion for decomposable metrics (defined formally below, these metrics include doubling and planar metrics). In Section 9 we also provide stronger guarantees for the Metric Ramsey Partitions which depend on the local neighborhood of a node, which we later use for application to proximity problems.

Another natural property one may desire is to have embeddings whose distortion depends on the *distance* between points and not on the *cardinality* of the closer neighbors. For example, in a social network it may be desirable to obtain good distortion to all neighbors of distance t away, as a function of t. In the context of "small world" networks, [Kle00] studied a distribution that depends on the distance with exactly this type of local behavior. In Section 10 we study embeddings with *proximity distortion* – in which the distortion bound of a pair x, y is a function of $d(x, y)$. Theorem 14 is our main result using this notion. We show that embeddings into ℓ_p with scaling proximity distortion $\tilde{O}(\log t)$ are given for decomposable metrics.

In Section 11 we discuss some applications of our local embeddings. We show that in systems using a "small world" distribution, our local embeddings provide constant average distortion. We also discuss the application of our probabilistic embedding into ultrametrics to online problems with local structure of the request sequence. Finally, we discuss how our techniques can be used to provide better distance oracles and proximity ranking data structures. For example, we provide distance oracles with linear storage, constant query time and scaling local stretch $\tilde{O}(\log k)$ (that is the stretch for the kth nearest neighbor).

2 Preliminaries

We assume throughout the paper that if (X, d) is a metric space then for all $x \neq y \in X$, $d(x, y) \geq 1$. A ball around $x \in X$ with radius $r \geq 0$ is defined as $B(x, r) = \{z \in X \mid d(x, z) \leq r\}$, the open ball is $B^\circ(x, r) = \{z \in X \mid d(x, z) < r\}$. The diameter is defined as $\mathrm{diam}(X) = \max_{x,y \in X}\{d(x, y)\}$. For sets $A, B, C \subseteq X$ we define that $A \bowtie (B, C)$ if $A \cap B \neq \emptyset$ and $A \cap C \neq \emptyset$.

An ultrametric (X, d) is a metric space satisfying a strong form of the triangle inequality, that is for all $x, y, z \in X$, $d(x, z) \le \max\{d(x, y), d(y, z)\}$. The following definition is known to be equivalent to the above definition for finite spaces.

Definition 3. *An* ultrametric *is a metric space whose elements are the leaves of a rooted tree T. Each vertex $u \in T$ is associated with a label $\Delta(u) \ge 0$ such that $\Delta(u) = 0$ iff u is a leaf of T. It is required that if a u is a child of a v then $\Delta(u) \le \Delta(v)$. The distance between two leaves $x, y \in T$ is defined as $\Delta(\mathrm{lca}(x, y))$, where $\mathrm{lca}(x, y)$ is the least common ancestor of x and y in T.*

Some of our results apply to a restricted family of metric spaces with bounded growth rate. We use the following:

Definition 4 (Growth bound)**.** *Let (X, d) be a metric space and $\chi \ge 1$ a fixed real constant.*

- *X has a χ* growth bound *if $|B(u, 2r)| \le 2^\chi |B(u, r)|$ for all $u, r > 0$.*
- *X has a χ* weak growth bound *if $|B(u, 2r)| \le |B(u, r)|^\chi$ for all $u, r > 0$ such that $|B(u, r)| > 1$.*

Note that the weak growth bound is an extremely weak property that even constant-degree expanders satisfy.

In many of our scaling results we shall use the following family Ξ of functions: A function $\vartheta : \mathbb{R}_+ \to \mathbb{R}_+$ is in Ξ if it is a monotone non-decreasing function satisfying

$$\sum_{i=1}^{\infty} \frac{1}{\vartheta(i)} = 1 . \tag{1}$$

For example if we define $\log^{(0)} n = n$, and for any $i > 0$ define recursively $\log^{(i)} n = \log(\log^{(i-1)} n)$, then we can take for any constants $\theta > 0$, $t \in \mathbb{N}$ the function $\vartheta(n) = \hat{c} \prod_{j=0}^{t-1} \log^{(j)}(n) \cdot \left(\log^{(t)}(n)\right)^{1+\theta}$, for sufficiently small constant $\hat{c} > 0$, and it will satisfy the conditions.

We will use the following concentration bounds,

Lemma 1 (Chernoff)**.** *Let X_i be $\{0, 1\}$ independent random variables for $i = 1, \ldots, d$, each with $\Pr[X_i = 1] = p_i$. Let $X = \sum_{i=1}^{d} X_i$ and $\mu = \mathbb{E}[X]$, then for any $0 < \delta < 1$,*

$$\Pr[X \le (1 - \delta)\mu] \le e^{-\delta^2 \mu/2} .$$

Lemma 2 (Hoeffding)**.** *Let Z_i be independent random variables for $i = 1, \ldots, d$, let $\mathbb{E}[Z_i] = \mu_i$ and $0 \le Z_i \le M_i$. Let $Z = \sum_{i=1}^{d} Z_i$, $\mu = \sum_{i=1}^{d} \mu_i$ and $M = \sum_{i=1}^{d} M_i^2$. Then for any $\eta > 0$*

$$\Pr[|Z - \mu| \ge \eta] \le 2e^{-2\eta^2/M}.$$

A powerful tool we will often use is the following Lemma known as Lovász Local Lemma.

Lemma 3 (Local Lemma)**.** *Let $\mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_n$ be events in some probability space. Let $G(V, E)$ be a graph on n vertices with degree at most d, each vertex corresponding to an event. Assume that for any $i = 1, \ldots, n$*

$$\Pr\left[\mathcal{A}_i \mid \bigwedge_{j \in Q} \neg \mathcal{A}_j\right] \le p$$

for all $Q \subseteq \{j : (\mathcal{A}_i, \mathcal{A}_j) \notin E\}$. If $ep(d + 1) \le 1$, then

$$\Pr\left[\bigwedge_{i=1}^{n} \neg \mathcal{A}_i\right] > 0$$

3 Local Probabilistic Partitions

Several of our results use probabilistic partitions [Bar96]. In this section we review some definitions and results concerning these tools, extending the notions of [ABN06].

Definition 5. *The local growth rate of $x \in X$ at radius $r > 0$ for a given scale $\gamma > 0$ is defined as*

$$\rho(x, r, \gamma) = |B(x, r\gamma)|/|B(x, r/\gamma)|.$$

Given a subspace $Z \subseteq X$, the minimum local growth rate of Z at radius $r > 0$ and scale $\gamma > 0$ is defined as

$$\rho(Z, r, \gamma) = \min_{x \in Z} \rho(x, r, \gamma).$$

The minimum local growth rate of $x \in X$ at radius $r > 0$ and scale $\gamma > 0$ is defined as

$$\bar{\rho}(x, r, \gamma) = \rho(B(x, r), r, \gamma).$$

Claim 4. *Let $x, y \in X$, let $\gamma > 0$ and let r be such that $2(1 + 1/\gamma)r < d(x, y) \leq (\gamma - 2 - 1/\gamma)r$, then*

$$\max\{\bar{\rho}(x, r, \gamma), \bar{\rho}(y, r, \gamma)\} \geq 2.$$

Proof. Let $B_x = B(x, r(1 + 1/\gamma))$, $B_y = B(y, r(1 + 1/\gamma))$, and assume w.l.o.g that $|B_x| \leq |B_y|$. As $r(1 + 1/\gamma) < d(x, y)/2$ we have $B_x \cap B_y = \emptyset$. Note that for any $x' \in B(x, r)$, $B(x', r/\gamma) \subseteq B_x$, and similarly for any $y' \in B(y, r)$, $B(y', r/\gamma) \subseteq B_y$. On the other hand $B(x', r\gamma) \supseteq B_x \cup B_y$, since for any $y' \in B_y$, $d(x', y') \leq d(x', x) + d(x, y) + d(y, y') \leq r + r(\gamma - 1/\gamma - 2) + r(1 + 1/\gamma) = r\gamma$. We conclude that

$$\rho(x', r, \gamma) = |B(x', r\gamma)|/|B(x', r/\gamma)| \geq (|B_x| + |B_y|)/|B_x| \geq 2.$$

$\square$

Definition 6 (Partition)**.** *A partition P of X is a collection of pairwise disjoint sets $\mathcal{C}(P) = \{C_1, C_2, \ldots, C_t\}$ for some integer t, such that $X = \cup_j C_j$. The sets $C_j \subseteq X$ are called clusters. For $x \in X$ denote by $P(x)$ the cluster containing x. Given $\Delta > 0$, a partition is* Δ-bounded *if for all $j \in [t]$, $\mathrm{diam}(C_j) \leq \Delta$. For $Z \subseteq X$ we denote by $P[Z]$ the restriction of P to points in Z.*

Definition 7 (Probabilistic Partition)**.** *A* probabilistic partition *$\hat{\mathcal{P}}$ of a metric space (X, d) is a distribution over a set $\mathcal{P}$ of partitions of X. Given $\Delta > 0$, $\hat{\mathcal{P}}$ is Δ-bounded if each $P \in \mathcal{P}$ is Δ-bounded. Let* $\mathrm{supp}(\hat{\mathcal{P}}) \subseteq \mathcal{P}$ *be the set of partitions with non-zero probability under $\hat{\mathcal{P}}$.*

Definition 8 (Uniform Function)**.** *Given a partition P of a metric space (X, d), a function f defined on X is called* uniform *with respect to P if for any $x, y \in X$ such that $P(x) = P(y)$ we have $f(x) = f(y)$.*

Let $\hat{\mathcal{P}}$ be a probabilistic partition. A collection of functions defined on X, $f = \{f_P | P \in \mathcal{P}\}$ is uniform *with respect to $\mathcal{P}$ if for every $P \in \mathcal{P}$, f_P is uniform with respect to P.*

Definition 9 (Uniformly Padded Local PP)**.** *Given $\Delta > 0$ and $0 < \delta \leq 1$, let $\hat{\mathcal{P}}$ be a Δ-bounded probabilistic partition of (X, d). Given a collection of functions $\eta = \{\eta_P : X \to [0, 1] | P \in \mathcal{P}\}$, we say that $\hat{\mathcal{P}}$ is (η, δ)-padded if for any $x \in X$,*

$$\Pr[B(x, \eta_P(x)\Delta) \subseteq P(x)] \geq \delta\ .$$

*We say that $\hat{\mathcal{P}}$ is (η, δ)-*locally *padded if the event $B(x, \eta_P(x)\Delta) \subseteq P(x)$ occurs with probability at least δ regardless of the structure of the partition outside $B(x, 2\Delta)$. Formally, for all $x \in X$, for all $C \subseteq X \setminus B(x, 2\Delta)$ and all partitions P' of C,*

$$\Pr[B(x, \eta_P(x)\Delta) \subseteq P(x) \mid P[C] = P'] \geq \delta$$

*We say that $\hat{\mathcal{P}}$ is (η, δ)-*uniformly *(locally) padded if η is uniform with respect to $\mathcal{P}$.*

Definition 10. *Let (X, d) be a finite metric space. Let $\tau \in (0, 2^{-7}]$. We say that X admits a (local) τ-decomposition if for every $0 < \Delta \le \mathrm{diam}(X)$ there exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of X such that for all $\delta \le 1$ satisfying $\log(1/\delta) \le 2^6/\tau$, $\hat{\mathcal{P}}$ is $(\tau \cdot \log(1/\delta), \delta)$-(locally) padded.*

Several known families of decomposable metrics include λ-doubling metrics[3], that admit a local $\Omega(\log^{-1}\lambda)$-decomposition [GKL03, ABN08a], and metrics induced by the shortest path on K_r-minor excluded graphs, that admit a $\Omega(r^{-2})$-decomposition [KPR93, FT03]. Observe that any metric admits a local τ-decomposition for sufficiently small τ (for instance if $\tau < \frac{\min_{x \neq y \in X} d(x,y)}{\mathrm{diam}(X)}$).

The following Lemma with local properties is proven in [ABN08a], we give the proof in Appendix A for completeness.

Lemma 5. *Let (X, d) be a finite metric space. Assume X admits a (local) τ-decomposition. Let $0 < \Delta \le \mathrm{diam}(X)$, let $\hat{\delta} \in (0, 1/2]$ satisfying $\ln(1/\hat{\delta}) \le 2^6\tau^{-1}$, and let $\gamma \ge 16$. There exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of (X, d) and a collection of uniform functions $\{\xi_P : X \to \{0, 1\} \mid P \in \mathcal{P}\}$ and $\{\eta_P : X \to (0, 1/\ln(1/\hat{\delta})] \mid P \in \mathcal{P}\}$ such that for any $\hat{\delta} \le \delta \le 1$, the probabilistic partition $\hat{\mathcal{P}}$ is $(\eta \cdot \ln(1/\delta), \delta)$-uniformly padded probabilistic partition; and the following conditions hold for any $P \in \mathcal{P}$ and any $x \in X$:*

- *$\eta_P(x) \ge \tau/2$.*
- *If $\xi_P(x) = 1$ then: $2^{-7}/\ln\rho(x, 2\Delta, \gamma) \le \eta_P(x) \le 2^{-7}/\ln(1/\hat{\delta})$.*
- *If $\xi_P(x) = 0$ then: $\eta_P(x) = 2^{-7}/\ln(1/\hat{\delta})$ and $\bar{\rho}(x, 2\Delta, \gamma) < 1/\hat{\delta}$.*

Furthermore, if X admits a local τ-decomposition then $\hat{\mathcal{P}}$ is local.

3.1 Bounded Cardinality Probabilistic Partitions

In this section we describe new type of probabilistic partitions, where instead of the usual notion of bounded diameter partitions, we require a bound on the *cardinality* of the clusters. Similar partitions (without the local property) were independently shown by [CMM10].

Definition 11. *Fix an integer $k \ge 2$. Let $2/k \le \delta \le 1$. A distribution on partitions $\hat{\mathcal{P}}$ of a metric space (X, d) is k-bounded and locally padded with parameter δ if*

1. *For any $P \in \mathrm{supp}(\hat{\mathcal{P}})$ and $x \in X$, $\mathrm{diam}(P(x)) \le r_k(x)/8$.*
2. *Denote by $\mathcal{L}(x)$ the event that $B(x, 2^{-9}r_k(x)\log(1/\delta)/\log k) \subseteq P(x)$. For any $Z \subset X \setminus B(x, r_k(x)/4)$ and any partition P' of Z,*
$$\Pr[\neg\mathcal{L}(x) \mid P[Z] = P'] \le 1 - \delta$$

The first property bounds the number of points in each cluster by k (the property is actually slightly stronger, in order to enable an application of the Lovász Local Lemma. The constant 8 is somewhat arbitrary). The second property states that the probabilistic partition is locally padded with probability at least δ, where the locality is with respect to the points in $B(x, r_k(x)/4)$, as opposed to $B(x, 2\Delta)$ used in Δ-bounded partitions.

Lemma 6. *For any metric space (X, d) on n points, any integer $2 \le k \le n$ and any $2/k \le \delta \le 1$, there exists a k-bounded and locally padded probabilistic partition with parameter δ.*

Create the partition P of X into clusters by generating a sequence of clusters: $C_1, C_2, \ldots C_s$, for some fixed $s \in [n]$. Notice that we are generating a distribution over partitions and therefore the generated clusters are random variables. First we deterministically assign centers $v_1, v_2, \ldots, v_s$ by the following iterative process: Let $W_1 = X$ and $j = 1$.

[3]A metric space (X, d) is λ-doubling if for any $x \in X$ and $r > 0$, $B(x, 2r)$ can be covered by λ balls of radius r.

1. Let $v_j \in W_j$ be the point maximizing $r_k(x)$ over all $x \in W_j$.
2. Let $W_{j+1} = W_j \setminus B(v_j, r_k(v_j)/64)$.
3. Set $j = j + 1$. If $W_j \neq \emptyset$ return to 1.

Now the algorithm for the partition is as follows: Let $Z_1 = X$. For $j = 1, 2, 3 \dots s$:

- Apply the decomposition of Lemma 49 on the set Z_j with the center v_j and the parameters $\lambda = k$ and $\Delta = r_k(v_j)/16$, to obtain a partition of Z_j to $(S, Z_j \setminus S)$. Recall that $S = B(v_j, r) \cap Z_j$ for some $r \in [\Delta/4, \Delta/2]$.
- Define $C_j = S$ and $Z_{j+1} = Z_j \setminus C_j$.

Observe that some clusters may be empty, it is not necessarily the case that $v_j \in C_j$, and every cluster contains at most k points.

Fix $x \in X$ and let C be the cluster that contains x and has center v, then we have the following

$$r_k(v)/2 \leq r_k(x) \leq r_k(v) \ . \tag{2}$$

To see the left hand side of (2), note that since $d(x, v) \leq r_k(v)/32$, it follows that $|B(x, r_k(v)/2)| \leq |B(v, (1/2 + 1/32)r_k(v))| \leq k$, hence $r_k(x) \geq r_k(v)/2$. For the right hand side, we use the maximality of $r_k(v)$: Note that $W_j \subseteq C_j$ for all $1 \leq j \leq s$, so that $x \notin C_j$ for any cluster C_j formed before C implies that $x \notin W_j$ as well.

Now we are ready to show the first property of Definition 11. Fix $y \in X$ such that $r_k(x)/8 < d(x, y)$, we need to show that $y \notin P(x) = C$. Since $C \subseteq B(v, r_k(v)/32)$ we get that

$$d(v, y) \geq d(y, x) - d(x, v) > r_k(x)/8 - r_k(v)/32 \overset{(2)}{\geq} r_k(v)/16 - r_k(v)/32 = r_k(v)/32 \ ,$$

it follows that $y \notin C$.

Next we will prove the locality of the second property of the partition. Let $\eta = 2^{-9} \ln(1/\delta)/\ln k$, and for any $x \in X$ let $T_x = B(x, r_k(x)/8)$. If v is a cluster center that contains some point in $B(x, \eta \cdot r_k(v))$, then it must be that $v \in T_x$. To see this, first observe that $d(v, x) \leq r_k(v)/32 + \eta \cdot r_k(v) \leq r_k(v)/16$ and also $r_k(v) \leq d(v, x) + r_k(x) \leq r_k(v)/16 + r_k(x)$, thus $r_k(v) \leq 16r_k(x)/15$, and finally $d(x, v) \leq r_k(v)/16 \leq r_k(x)/8$. Let v be such a center. Since the choice of radius is the only randomness in the process of creating P, the event of padding for $x \in X$ is determined by the choice of radiuses for centers $v_j \in T_x$. Let $Z = X \setminus B(x, r_k(x)/4)$, then it suffices to show that any cluster C with center $v \in T_x$ satisfies $C \cap Z = \emptyset$. Fix such cluster C, and observe that for any point $y \in C$,

$$d(x, y) \leq d(x, v) + d(v, y) \leq r_k(x)/8 + r_k(v)/32 \overset{(2)}{<} r_k(x)/4 \ .$$

We conclude by proving the bound on the padding probability. Consider the distribution over the clusters $C_1, C_2, \dots C_s$ as defined above. For $1 \leq m \leq s$, define the events:

$$\begin{aligned}
\mathcal{Z}_m &= \{\forall j, 1 \leq j < m,\ B(x, \eta \cdot r_k(v_m)) \subseteq Z_{j+1}\}, \\
\mathcal{E}_m &= \{\exists j,\ m \leq j < s \text{ s.t. } B(x, \eta \cdot r_k(v_m)) \bowtie (C_j, Z_{j+1}) \wedge \mathcal{Z}_m\}.
\end{aligned}$$

Also let $T = T_x$ and $\theta = \sqrt{\delta}$. We prove the following inductive claim: For every $1 \leq m \leq s$:

$$\Pr[\mathcal{E}_m \mid \mathcal{Z}_m] \leq (1 - \theta)(1 + \theta \sum_{j \geq m, v_j \in T} k^{-1}). \tag{3}$$

Note that $\Pr[\mathcal{E}_s] = 0$ and $\Pr[\mathcal{Z}_1] = 1$. Assume the claim holds for $m + 1$ and we will prove for m. Define the events:

$$\begin{aligned}
\mathcal{F}_m &= \{B(x, \eta \cdot r_k(v_m)) \bowtie (C_m, Z_{m+1}) \wedge \mathcal{Z}_m\}, \\
\mathcal{G}_m &= \{B(x, \eta \cdot r_k(v_m)) \subseteq Z_{m+1} \wedge \mathcal{Z}_m\} = \{\mathcal{Z}_{m+1} \wedge \mathcal{Z}_m\}, \\
\bar{\mathcal{G}}_m &= \{B(x, \eta \cdot r_k(v_m)) \nsubseteq Z_{m+1} \wedge \mathcal{Z}_m\} = \{\bar{\mathcal{Z}}_{m+1} \wedge \mathcal{Z}_m\}.
\end{aligned}$$

First we bound $\Pr[\mathcal{F}_m]$. By Lemma 49, noting that $\eta \cdot r_k(v_m) = \frac{1}{16}\ln(1/\theta)/\ln k \cdot \Delta$ and that indeed $\theta = \sqrt{\delta} \geq \sqrt{2/k} \geq 2/k = 2\chi^{-1}$, we get that

$$\begin{aligned} \Pr[\mathcal{F}_m \mid \mathcal{Z}_m] &= \Pr[B(x, \eta \cdot r_k(v_m)) \bowtie (C_m, Z_{m+1}) | \mathcal{Z}_m] \\ &\leq (1-\theta)(\Pr[B(x, \eta \cdot r_k(v_m)) \nsubseteq Z_{m+1} | \mathcal{Z}_m] + \theta k^{-1}) \\ &= (1-\theta)(\Pr[\bar{\mathcal{G}}_m \mid \mathcal{Z}_m] + \theta k^{-1}) . \end{aligned} \tag{4}$$

Using the induction hypothesis we prove the inductive claim, recalling that $\Pr[\mathcal{F}_m] = 0$ if $v_m \notin T$ (note that $v_m \in T$ is determined by a deterministic process),

$$\begin{aligned} \Pr[\mathcal{E}_m \mid \mathcal{Z}_m] &\leq \Pr[\mathcal{F}_m \mid \mathcal{Z}_m] + \Pr[\mathcal{G}_m \mid \mathcal{Z}_m] \Pr[\mathcal{E}_{m+1} \mid \mathcal{Z}_m \wedge \mathcal{G}_m] \\ &\leq (1-\theta)(\Pr[\bar{\mathcal{G}}_m \mid \mathcal{Z}_m] + \theta k^{-1})\mathbf{1}_{\{v_m \in T\}} + \Pr[\mathcal{G}_m \mid \mathcal{Z}_m] \cdot (1-\theta)(1 + \theta \sum_{j \geq m+1, v_j \in T} k^{-1}) \\ &\leq (1-\theta)(1 + \theta \sum_{j \geq m, v_j \in T} k^{-1}) , \end{aligned}$$

where the second inequality follows from (4) and the induction hypothesis. Note that for any $x \in X$, $|T_x| \leq k$, so that $\sum_{j \geq 1, v_j \in T_x} k^{-1} \leq 1$. We conclude from the claim (3) for $m = 1$, observing that conditioned on event $\mathcal{Z}_m$, $x \in Z_m$ so by (2) $r_k(x) \leq r_k(v_m)$. Finally,

$$\begin{aligned} \Pr[B(x, \eta \cdot r_k(x)) \nsubseteq P(x)] &\leq \Pr[\exists j,\, 1 \leq j < s \text{ s.t. } B(x, \eta \cdot r_k(v_j)) \bowtie (C_j, Z_{j+1}) \mid \mathcal{Z}_j] \\ &= \Pr[\mathcal{E}_1 \mid \mathcal{Z}_1] \\ &\leq (1-\theta)(1 + \theta \cdot \sum_{j \geq 1, v_j \in T} k^{-1}) \\ &\leq (1-\theta)(1+\theta) = 1 - \delta. \end{aligned}$$

.

4 Basic Results

4.1 Embedding into an Ultrametric with Scaling Local Distortion

The following theorem is a strengthening of the known embeddings of metrics into an ultrametric [Bar96, BLMN05, HPM06]:

Theorem 1. *For any finite metric space (X, d) on n points there exists an embedding into an ultrametric with strong scaling local distortion k.*

Proof. Let T be any minimum spanning tree of the metric (X, d) (viewed as a complete graph with vertex set X and edge weights that correspond to distances d). Let P_{xy} be the unique path connecting x and y in T, and define $\rho(x, y)$ as the maximal weight of an edge in P_{xy}. We claim that (X, ρ) is indeed an ultrametric, that is for any $x, y, z \in X$,

$$\rho(x, y) \leq \max\{\rho(x, z), \rho(y, z)\}. \tag{5}$$

Indeed, consider the maximal edge $e \in P_{xy}$, then one of the paths P_{xz}, P_{yz} must contain this edge, therefore (5) holds.

Next we show that $\rho(x, y) \leq d(x, y)$. If the edge $(x, y) \in T$ then there is equality, otherwise, consider the cycle $P_{xy} \cup (x, y)$. A known property of minimum spanning trees states that any edge e in an MST is not longer than the maximal edge in a cycle containing e, therefore the maximal edge length in P_{xy} is bounded by $d(x, y)$. It remains to show that if $y \in N_k(x)$, then $\rho(x, y) \geq d(x, y)/k$. This will be proven by showing that the path P_{xy} must contain an edge of weight at least $d(x,y)/k$. Assume by contradiction this is not so, then the triangle inequality implies that there are at least k points at distance strictly smaller than $d(x,y)$ from x (the first k points in P_{xy} after x), which is a contradiction to the fact that $y \in N_k(x)$. We conclude the the maximal edge in P_{xy} is of weight at least $d(x,y)/k$.

To see that we actually obtain strong scaling local distortion, if $y \notin N_k(x)$ then still $\rho(x,y) \geq r_k(x)/k$. Otherwise, the same first k points of P_{xy} are of distance strictly smaller than $r_k(x)$ to x, which is contradiction to the definition of $r_k(x)$. □

4.2 Embedding into ℓ_p with k-Local Distortion

Theorem 2. *For any metric space (X,d) on n points, there exists an embedding $f : X \to \ell_p$ with* strong *k-local distortion $O(\frac{\log k}{p})$ and dimension $O(2^p \log n \log k)$.*

Remark 1. *In Section 6 we show how to obtain dimension which is independent of n, using more sophisticated techniques.*

Proof. The embedding is basically Bourgain's embedding [Bou85] with improvement for large p by Matoušek [Mat90], however we use only the first $(\log k)/p$ densities.

Let $s = 2^p$, $t = \lceil \log_s k \rceil$, and $q = c \cdot s \log n$ for a constant c to be determined later. Choose subsets $A_{ij} \subseteq X$ for every $i \in [t]$, $j \in [q]$, such that each point is included in A_{ij} independently with probability $\frac{1}{s^i}$. We now define the embedding $f : X \to \mathbb{R}^{tq}$ by defining for each $i \in [t]$, $j \in [q]$ a function $f_{ij} : X \to \mathbb{R}$ by $f_{ij}(u) = d(u, A_{ij})$, and

$$f(u) = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{q} f_{ij}(u) \; .$$

Fix some pair $u, v \in X$. Let $L = \min\{d(u,v), r_k(u)\}$. For $i \in [t] \cup \{0\}$ define $\rho_i = \max\{r_{s^i}(u), r_{s^i}(v)\}$. Let m be the minimal positive integer such that $\rho_m + \rho_{m-1} \geq L/4$, and observe that $\rho_t \geq r_k(u)$, thus $m \leq t$. If $\rho_m > L/2$ redefine $\rho_m = L/2$. We will prove that with high probability, for all pairs and all $i \in [m]$ there are at least $q/(24s)$ values of j such that,

$$|d(u, A_{ij}) - d(v, A_{ij})| \geq \rho_i - \rho_{i-1} \; , \tag{6}$$

Before proving (6), let us see how to conclude the proof of the theorem. In order to bound the expansion of the embedding we note that the triangle inequality suggests that $|d(u, A_{ij}) - d(v, A_{ij})| \leq d(u,v)$ for all i, j, thus

$$\|f(u) - f(v)\|_p^p = \sum_{i=1}^{t} \sum_{j=1}^{q} |d(u, A_{ij}) - d(v, A_{ij})|^p \leq tq \cdot d(u,v)^p \; .$$

Now we turn to the contraction. First observe that $\rho_0 = 0$, and since the ρ_i are monotonically increasing, $\rho_m \geq L/8$. For each $i \in [m]$ let $Q_i \subseteq [q]$ be the subset of indices j for which (6) holds, then $|Q_i| \geq q/(24s)$ for all i, we conclude that

$$\begin{aligned}
\|f(u)-f(v)\|_p^p &= \sum_{i=1}^{t}\sum_{j=1}^{q}|d(u,A_{ij})-d(v,A_{ij})|^p \\
&\geq \sum_{i=1}^{m}\sum_{j\in Q_i}|d(u,A_{ij})-d(v,A_{ij})|^p \\
&\overset{(6)}{\geq} \sum_{i=1}^{m}\sum_{j\in Q_i}(\rho_i-\rho_{i-1})^p \\
&\geq \frac{q}{24s\cdot m^{p-1}}\left(\sum_{i=1}^{m}(\rho_i-\rho_{i-1})\right)^p \\
&\geq \frac{q}{24s\cdot t^{p-1}}\rho_m^p \\
&\geq \frac{q}{24s\cdot t^{p-1}}(L/8)^p\ .
\end{aligned}$$

Recall that $L=\min\{d(u,v),r_k(u)\}$ is the lower bound quantity for strong k-local distortion. The distortion is bounded by $8(tq\cdot\frac{24s\cdot t^{p-1}}{q})^{1/p}=O(t)=O((\log k)/p)$. It remains to prove (6) for sufficiently many values of j. Fix some $i\in[m]$. Assume that $\rho_i=r_{s^i}(v)$ (the case where $\rho_i=r_{s^i}(u)$ is symmetric), and define events for all $j\in[q]$

$$\begin{aligned}
\mathcal{E}_j &= d(A_{ij},u)\leq\rho_{i-1} \\
\mathcal{F}_j &= d(A_{ij},v)\geq\rho_i\ .
\end{aligned}$$

It is easy to check that if both events hold then (6) holds for this j. We will use that $e^{x/2}\leq 1+x\leq e^x$ for $0\leq x\leq 1/2$. Since $|B(u,\rho_{i-1})|\geq s^{i-1}$,

$$\Pr[\mathcal{E}_j]=1-(1-s^{-i})^{|B(u,\rho_{i-1})|}\geq 1-(1-s^{-i})^{s^{i-1}}\geq 1-e^{-s^{-i}\cdot s^{i-1}}=1-e^{-s^{-1}}\geq 1/(2s)\ ,$$

and since $|B^\circ(v,\rho_i)|<s^i$ (observe that this is true for ρ_m as well, even if we redefined it),

$$\Pr[\mathcal{F}_j]=(1-s^{-i})^{|B^\circ(v,\rho_i)|}\geq(1-s^{-i})^{s^i}\geq 1/(2e)\ .$$

These events depend on inclusion in A_{ij} of points that belong to disjoint balls, and thus are independent. We conclude that

$$\Pr[|d(u,A_{ij})-d(v,A_{ij})|\geq\rho_i-\rho_{i-1}]\geq\Pr[\mathcal{E}_j\wedge\mathcal{F}_j]\geq 1/(12s)\ .$$

Let X_{ij} be indicator random variable for the event that both $\mathcal{E}_j$ and $\mathcal{F}_j$ hold, and let $X_i=\sum_{j=1}^{q}X_{ij}$. Note that $\mathbb{E}[X_i]\geq q/(12s)=c/12\cdot\log n$, and by Chernoff bound

$$\Pr[X_i<\mathbb{E}[X_i]/2]\leq e^{-\mathbb{E}[X_i]/4}\leq e^{-(c\log n)/48}=n^{-4}\ ,$$

where the last equation is by choosing c as a large enough constant. By the union bound over all values of i and all pairs u,v, with probability at least $1-1/n$, (6) holds for at least $\mathbb{E}[X_i]/2\geq q/(24s)$ values of j for all pairs and all i. □

4.3 Embedding into ℓ_p with Scaling Local Distortion

Theorem 3. *For any finite metric space (X,d) on n points and $\vartheta\in\Xi$ there exists an embedding into ℓ_p with strong scaling local distortion $O\left(\left(\frac{\log k}{p}\right)^{1-\frac{1}{p}}\left(\vartheta\left(\frac{\log k}{p}\right)\right)^{\frac{1}{p}}\right)$, worse case distortion $O((\log n)/p)$ and dimension $O(2^p\log^2 n)$.*

Proof. The embedding is similar to Bourgain's embedding described in the previous section, the main novelty is the use of a function $\vartheta \in \Xi$ to scale down each coordinate as described below.

Let $s = 2^p$,$t = \log_s n$, $q = c \cdot s \log n$ for a constant c to be determined later. Choose subsets $A_{ij} \subseteq X$ for every $i \in [t]$, $j \in [q]$, such that each point is included in A_{ij} independently with probability $\frac{1}{s^i}$. We now define the embedding $f : X \to \mathbb{R}^{tq}$ by defining for each $i \in [t]$, $j \in [q]$ a function $f_{ij} : X \to \mathbb{R}$ by $f_{ij}(u) = \frac{d(u,A_{ij})}{\vartheta(i)^{1/p}}$, and

$$f(u) = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{q} f_{ij}(u) \; .$$

Fix $u, v \in X$ and some integer $k \in [n]$. Let $L = \min\{d(u, v), r_k(u)\}$. For $i \in [t] \cup \{0\}$ define $\rho_i = \max\{r_{s^i}(u), r_{s^i}(v)\}$. Let m be the minimal positive integer such that $\rho_m + \rho_{m-1} \geq L/4$, if $\rho_m > L/2$ redefine $\rho_m = L/2$. Observe that $r_{s^{m-1}}(u) \leq r_k(u)$ thus $m \leq 1 + \log_s k$. The same proof as the previous section suggests that with high probability, for all pairs and all $i \in [m]$ there are at least $q/(24s)$ values of j such that,

$$|d(u, A_{ij}) - d(v, A_{ij})| \geq \rho_i - \rho_{i-1} \; . \tag{7}$$

To bound the expansion we have that

$$\begin{aligned} \|f(u) - f(v)\|_p^p &= \sum_{i=1}^{t} \sum_{j=1}^{q} \left(\frac{|d(u, A_{ij}) - d(v, A_{ij})|}{\vartheta(i)^{1/p}} \right)^p \\ &\leq q \cdot d(u, v)^p \sum_{i=1}^{t} \frac{1}{\vartheta(i)} \\ &\leq q \cdot d(u, v)^p \; . \end{aligned} \tag{8}$$

Now we turn to the contraction. For each $i \in [m]$ let $Q_i \subseteq [q]$ be the subset of indices j for which (6) holds, then $|Q_i| \geq q/(24s)$ for all i, we conclude that

$$\begin{aligned} \|f(u) - f(v)\|_p^p &= \sum_{i=1}^{t} \sum_{j=1}^{q} \left(\frac{|d(u, A_{ij}) - d(v, A_{ij})|}{\vartheta(i)^{1/p}} \right)^p \\ &\geq \sum_{i=1}^{m} \sum_{j \in Q_i} \left(\frac{|d(u, A_{ij}) - d(v, A_{ij})|}{\vartheta(i)^{1/p}} \right)^p \\ &\geq \frac{q}{24s} \sum_{i=1}^{m} \frac{1}{\vartheta(i)} (\rho_i - \rho_{i-1})^p \\ &\geq \frac{q}{24s \cdot \vartheta(m) \cdot m^{p-1}} \left(\sum_{i=1}^{m} (\rho_i - \rho_{i-1}) \right)^p \\ &\geq \frac{q}{26s \cdot \vartheta(\log_s k) \cdot (\log_s k)^{p-1}} \rho_m^p \\ &\geq \frac{q}{26s \cdot \vartheta(\log_s k) \cdot (\log_s k)^{p-1}} (L/8)^p \; . \end{aligned} \tag{9}$$

The distortion is bounded by $8 \left(q \cdot \frac{26s \cdot \vartheta(\log_s k) \cdot (\log_s k)^{p-1}}{q} \right)^{1/p} = O\left(\left(\frac{\log k}{p} \right)^{1-\frac{1}{p}} \left(\vartheta \left(\frac{\log k}{p} \right) \right)^{\frac{1}{p}} \right)$.

To see the worse case distortion of $O(\log_s n)$, let $\bar{\vartheta}(i) = \min\{\vartheta(i), \log_s n\}$ and use $\bar{\vartheta}^{-1/p}$ as the scaling factor in the embedding. This has the effect that in (9) we get a lower bound of $\frac{q}{26s \cdot \bar{\vartheta}(\log_s k) \cdot (\log_s k)^{p-1}} (L/8)^p$, which is always lower bounded by $\frac{q}{O(s \log_s^p n)} (L/8)^p$.

It remains to show an upper bound of $2q$ on the expansion. Let i' be the largest integer such that $\vartheta(i') \le \log_s n$, then

$$\sum_{i=1}^{t} \frac{1}{\bar{\vartheta}(i)} \le \sum_{i=1}^{i'} \frac{1}{\vartheta(i)} + \sum_{i=i'+1}^{t} \frac{1}{\log_s n} < 2\,,$$

plugging this into the last inequality of (8) concludes the proof. ☐

Note that for any $\epsilon > 0$ there exists $\hat{c}$ such that $\vartheta(k) = \hat{c} \cdot k(\log k)^{1+\epsilon}$ and $\vartheta \in \Xi$, hence

Corollary 7. *For any finite metric space (X,d) on n points and any constant $\epsilon > 0$ there exists an embedding into ℓ_p with scaling local distortion*

$$O\left(\frac{\log k}{p}(\log\log k)^{\frac{1+\epsilon}{p}}\right)$$

4.4 Lower Dimension for 1-Local Distortion

Theorem 4. *For any finite metric space X there exists an embedding into 3 dimensional ℓ_p with 1-local distortion $\sqrt[p]{3}$.*

Proof. Let $G = (V,E)$ be an unweighed graph with vertices corresponding to the points of X, and a pair $(u,v) \in E$, iff $v \in N_1(u)$ of u. Since each node has outdegree one, each connected component in G (viewed as undirected graph) has at most one cycle. Fix some component H and, let r_H be an arbitrary node of H, and if there is an odd cycle, such that the two farthest points from r_H on it, call them u, v are such that $v \in N_1(u)$ (or $u \in N_1(v)$) set $w_H = u$ (or $w_H = v$). Otherwise let w_H be arbitrary point.

Define 2 sets A_1, A_2 as follows: for any connected component H in G, insert into A_1 all the vertices in even distance from r_H, and into A_2 all the vertices in odd distance from r_H. Define the embedding into $\mathbb{R}^3$ as

$$f(u) = (d(u,A_1), d(u,A_2), g(u))$$

where $g(u)$ is $d(u, N_1(u))$ if $u = w_H$ and 0 otherwise.

Fix any u, v, to bound the expansion observe that no coordinate can have value larger than $d(u,v)$. This is by the triangle inequality for the first two coordinates, for the third, it is non-zero only if one of u, v, say u, is w_H, but even then its value $d(u, N_1(u)) \le d(u,v)$. This suggests expansion at most $\sqrt[p]{3}$. To bound contraction, assume $v \in N_1(u)$, and then u, v are connected by an edge in G, and lie in some component H. If their distance to r_H has different parity, then w.l.o.g $v \in A_1$ and $d(u,A_1) = d(u,v)$ and $d(v,A_1) = 0$. The only case where their distance to r_H has the same parity is that both of them are the farthest points from r_H in an odd cycle, but in such a case u is w_H and we get contribution from the third coordinate (if v was chosen as w_H, then also $u \in N_1(v)$ and again we get contribution of $d(u,v)$). ☐

Remark 2. *For any metric space (X,d) there is an embedding into the line which is an isometry on nearest neighbors*[4].

Proof. Let $G = (V,E)$ be the weighted graph containing only edges between points $x, y \in X$ such that $y \in N_1(x)$, the weight of the edge is simply $d(x,y)$. Similarly to Theorem 4 the graph G contains no cycles. Choose an arbitrary root r_C for every connected component C of G. Define the embedding $f : X \to \mathbb{R}^+$ as $f(x) = d(x, r_C)$ where C is the connected component containing x.

If $x, y \in X$ are such that $y \in N_1(x)$ then they are in the same connected component C, which is a tree, therefore $|d(x,r_C) - d(y,r_C)| = d(x,y)$. ☐

[4] Unlike the previous embedding, this embedding may have arbitrary expansion.

4.5 Local Dimension Reduction for the Equilateral Metric

The "usual suspect" for high dimensionality in ℓ_2 is the equilateral metric. Alon [Alo03] used this metric to provide the best known lower bound for dimension reduction, in particular, he showed that an n point equilateral requires dimension at least $\Omega(\log n/(\log(1/\epsilon)\cdot\epsilon^2))$, for $1+\epsilon$ distortion. However, this is *not* the case for local embedding, in the sense that the dimension required does not depend on n.

To embed an equilateral metric, first consider the neighborhood graph $G=(X,E)$, where $(u,v)\in E$ iff $v\in N_k(u)$ (note that we allow adversarial choice of neighbors). By Lemma 12 there exists a proper coloring of G with $2k+1$ colors. Using the standard dimension reduction of [JL84] we can embed the $m=2k+1$ color classes into $O((\log m)/\epsilon^2)$ dimensional ℓ_2 space with k-local distortion $1+\epsilon$. This is because for any point $u\in X$, all the points in $N_k(x)$ have different color than the color of x, so the distance between them is maintained up to the $1+\epsilon$ factor loss of the dimension reduction.

5 Probabilistic Local Embedding into Ultrametrics

Probabilistic embedding of metrics into ultrametrics [Bar96] has many applications in online and approximation algorithms. The basic theorem states that every metric space probabilistically embeds into a distribution over ultrametrics with $O(\log n)$ expected distortion [Bar98, FRT03, Bar04]. Here we extend this result to local embeddings. We will use the following lemma implicitly proven in [CKR01, FRT03].

Lemma 8. *Given a finite metric space (X,d) and $0<\Delta\le\operatorname{diam}(X)$, there exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of (X,d) such that for any $x\in X$ and any $0\le\eta\le 1/8$,*

$$\Pr[B(x,\eta\Delta)\nsubseteq P(x)]\le 16\eta\log\rho(x,\Delta/4,4)\ .$$

A proof of this lemma can be found in [MN06] and also in [ABN06] (with different constants).

5.1 Probabilistic Embedding into Trees with k-Local Distortion

Theorem 5. *For any finite metric space (X,d) on n points there exists a probabilistic embedding into a distribution of ultrametrics with strong k-local distortion $O(\log k)$.*

To prove this theorem we follow the constructions of [Bar96, FRT03, Bar04], and create am ultrametric from a probabilistic decomposition. A careful analysis of these construction reveals that the distortion of a pair x,y is bounded by $O(\log S)$ where S is the size of the largest cluster containing x,y in the partition. This is trivially $O(\log n)$. In order to obtain improved distortion for k nearest neighbors, we delete clusters that contain more than k points. Now the main difficulty is to create a meaningful tree from these partial partitions, and proving the required distortion bound.

Let $\Delta_0=\operatorname{diam}(X)$. For each $i>0$ define $\Delta_i=\frac{\Delta_0}{2^i}$, and create a Δ_i-bounded probabilistic partition $\hat{\mathcal{Q}}_i$ given by Lemma 8. Fix for each $i>0$ some $Q_i\in\mathcal{Q}_i$, and for every cluster $C\in Q_i$ define an arbitrary center $\upsilon(C)\in C$. Since we are required to give sufficient lower bound only on k-nearest neighbors, we shall eliminate large clusters. Formally, a cluster $C\in Q_i$ is called *large* if $|B(\upsilon(C),3\Delta_i)|>k$. Define P_i' by removing all the large clusters from Q_i, that is, $P_i'=\{C\in Q_i\mid |B(\upsilon(C),3\Delta_i)|\le k\}$. Define U_i to be the union of all the points in the large clusters of Q_i, formally $U_i=\{x\in C\mid C\in Q_i,|B(\upsilon(C),3\Delta_i)|>k\}$. In order to make this collection of sets *laminar*, we perform the following iterative process: for each integer $i>0$ define $P_i=\{C\cap D\mid C\in P_i',D\in P_{i-1}\cup\{U_{i-1}\}\}$ (where $P_0=\{X\}$, $U_0=\emptyset$). If x is not covered by any cluster of P_i we write $P_i(x)=\perp$.

The next step is to build an ultrametric T from this (partial) hierarchical partition. First observe that if $C\in Q_i$ is not large, then any cluster $D\in Q_{i+1}$ such that $C\cap D\neq\emptyset$, is not large as well. This holds because if $u\in C\cap D$ then $d(\upsilon(D),\upsilon(C))\le d(\upsilon(D),u)+d(u,\upsilon(C))\le\Delta_{i+1}+\Delta_i$, thus $|B(\upsilon(D),3\Delta_{i+1})|\le|B(\upsilon(C),4\Delta_{i+1}+\Delta_i)|=|B(\upsilon(C),3\Delta_i)|\le k$. Now for each $C\in P_i$ define a tree node labeled Δ_i, its children are the clusters of P_{i+1} that are contained in C (by the observation above every point in C is contained in some cluster of P_{i+1}), and if there exists a cluster in P_{i-1} that contain C, then its node will be the parent of the node corresponding to C. We got a collection of trees $T_1, \dots, T_l$ with labels $a_1 \le \dots \le a_l$, combine these into a single tree by adding $l-1$ additional nodes $u_2, \dots, u_l$, such that u_j has as children the roots of T_{j-1}, T_j, and is labeled a_j. This concludes the description of T, which is rooted at u_l. It is not hard to check that the labels are weakly monotone, in a sense that a parent has label at least as large as any of its children. Recall that the distance between two leaves is defined as the label of their least common ancestor.

Observation 9. *Let $x, y \in X$, and i be the maximal integer such that $P_i(x) = P_i(y)$, then $d_T(x, y) \le \Delta_i$.*

Proof. Since at least one of the cluster $P_{i+1}(x), P_{i+1}(y) \neq \perp$, we may assume w.l.o.g that $P_{i+1}(x) \neq \perp$. If it is the case that $P_i(x) \neq \perp$ as well, then the tree node corresponding to $P_i(x)$ is the least common ancestor of x, y, and is labeled Δ_i. Otherwise, x, y will fall into different trees among $T_1, \dots, T_l$, say $x \in T_a$ and $y \in T_b$, then the root of T_a is labeled Δ_{i+1} and the root of T_b is labeled by at most Δ_{i+1}, thus we may assume $a \ge b$. Since we combine trees by taking the smaller labeled trees first, the least common ancestor of x, y, which is the node u_a, is labeled by Δ_{i+1}. □

Fix some $x, y \in X$, and let $L = \min\{d(x, y), r_k(x)\}$. We begin by proving a bound on the contraction of the embedding. Let i be the smallest integer such that $4\Delta_i \le L$. Let $v = v(Q_i(x))$, then

$$B(v, 3\Delta_i) \subseteq B(x, 4\Delta_i) \subseteq B(x, L) \subseteq B(x, r_k(x)),$$

which implies that $|B(v, 3\Delta_i)| \le k$, and $Q_i(x)$ was not a large cluster. Therefore $P_i(x) \neq \perp$ and is labeled by Δ_i, furthermore, as the partition Q_i is Δ_i-bounded and $d(x, y) \ge L$, it follows that $P_i(x) \neq P_i(y)$. The monotonicity of labels suggests that

$$d_T(x, y) \ge \Delta_i \ge L/8 \ .$$

We now turn to bounding the expected expansion when sampling a partition according to the distribution $\hat{\mathcal{Q}}$. Observe that if $P_i(x) \neq \perp$ then it must be that $|B(x, \Delta_i)| \le |B(v(Q_i(x)), 3\Delta_i)| \le k$. Let $\delta_i(x) = 1$ if $|B(x, \Delta_i)| \le k$ and $\delta_i(x) = 0$ otherwise. Let $\mathcal{E}_i$ be the event that i is the minimal such that $P_i(x) \neq P_i(y)$.

Claim 10. *$\mathcal{E}_i$ implies that $Q_i(x) \neq Q_i(y)$ and $\delta_i(x) + \delta_i(y) \ge 1$.*

Proof. Assume by contradiction that $Q_i(x) = Q_i(y)$, then if for all $i' < i$, $Q_{i'}(x) = Q_{i'}(y)$ then $P_i(x) = P_i(y)$. Otherwise consider the maximal $i' < i$ such that $Q_{i'}(x) \neq Q_{i'}(y)$, there are two cases: if both clusters $Q_{i'}(x), Q_{i'}(y)$ are large, then $Q_{i'}(x), Q_{i'}(y) \subseteq U_{i'}$, and when we create the clusters of $P_{i'+1}, P_{i'+2}, \dots, P_i$, x, y will not be separated in any of these (partial) partitions (using the maximality of i'), thus we arrive to $P_i(x) = P_i(y)$, which contradicts $\mathcal{E}_i$. Otherwise at least one of the clusters, say $Q_{i'}(x)$, is small and will be included in $P'_{i'}$, hence also $P_{i'}(x) \neq P_{i'}(y)$ which contradicts the minimality of i.

Assume now that $\delta_i(x) = \delta_i(y) = 0$, then we remove the large clusters $Q_i(x)$, $Q_i(y)$ when creating P'_i, thus $P_i(x) = P_i(y) = \perp$. □

Using Lemma 8 with parameter $\eta = d(x, y)/\Delta_i$ we get that for all i satisfying $\Delta_i > 8d(x, y)$,

$$\Pr[Q_i(x) \neq Q_i(y)] \le \frac{16d(x, y) \cdot \log \rho(x, \Delta_i/4, 4)}{\Delta_i} \tag{10}$$
$$\Pr[Q_i(x) \neq Q_i(y)] \le \frac{16d(x, y) \cdot \log \rho(y, \Delta_i/4, 4)}{\Delta_i} \ .$$

Using Observation 9 and Claim 10 we bound the expected distance by

$$\begin{aligned} \mathbb{E}[d_T(x, y)] &\le \sum_{i>0} \Pr[\mathcal{E}_i] \cdot \Delta_{i-1} \\ &\le \sum_{i>0} \Pr[Q_i(x) \neq Q_i(y)](\delta_i(x) + \delta_i(y)) \cdot \Delta_{i-1} \\ &\le 2\sum_{i>0} \Pr[Q_i(x) \neq Q_i(y)]\delta_i(x) \cdot \Delta_i + 2\sum_{i>0} \Pr[Q_i(x) \neq Q_i(y)]\delta_i(y) \cdot \Delta_i \ , \end{aligned} \tag{11}$$

We focus on the first summation, let m be the maximal such that $\Delta_m > 8d(x,y)$, then

$$
\begin{aligned}
2\sum_{i>0}\Pr[Q_i(x)\neq Q_i(y)]\delta_i(x)\cdot\Delta_i &\leq 2\sum_{i=1}^{m}\Pr[Q_i(x)\neq Q_i(y)]\delta_i(x)\cdot\Delta_i + 2\sum_{i>m}\Delta_i \\
&\stackrel{(10)}{\leq} 32\sum_{i=1}^{m} d(x,y)\cdot\log\rho(x,\Delta_i/4,4)\cdot\delta_i(x) + 2\Delta_m \\
&\leq 32d(x,y)\sum_{i>0}\log\left(\frac{|B(x,\Delta_i)|}{|B(x,\Delta_{i+4})|}\right)\delta_i(x) + 16d(x,y) \\
&\leq 128d(x,y)\log k + 16d(x,y)\ ,
\end{aligned}
$$

where the last inequality hold because the summation is telescopic (observe that $\delta_i(x)$ is weakly monotone increasing with respect to i), and as $\delta_i(x) = 0$ when $|B(x,\Delta_i)| > k$, the largest element is bounded by $\log k$. Since we can bound the second summation of (11) in the exact same manner, we get that

$$\mathbb{E}[d_T(x,y)] \leq O(\log k)\cdot d(x,y)\ ,$$

which concludes the proof of Theorem 5.

5.2 Probabilistic Embedding into Trees with Scaling Local Distortion

Theorem 6. *For any finite metric space (X,d) on n points and $\vartheta\in\Xi$ there exists a probabilistic embedding into a distribution of ultra-metrics with strong scaling local distortion $O(\vartheta(\log k))$, and worst case distortion $O(\log n)$.*

The construction for the scaling local distortion is more involved than the one of the previous section, as here we are required to handle all values of k simultaneously. We use an idea similar to the one used in the proof of Theorem 3, and scale down the labels by a factor proportional to the logarithm of their size, and apply a function $\vartheta\in\Xi$ in order to guarantee convergence. The main obstacle is that after scaling down, the labels are no longer monotone, which means that creating the tree in the standard manner [Bar96, FRT03, Bar04] will not give a legal tree. We therefore apply a certain iterative process, that handles any non-monotonicity by "beaming up" clusters from lower levels to replace the violating clusters, thus increasing their label. We now turn to the formal proof.

Let $\Delta_0 = \mathrm{diam}(X)$ and for any integer $i > 0$ let $\Delta_i = \frac{\Delta_0}{4^i}$. For all $i > 0$ create a Δ_i-bounded probabilistic partition $\hat{\mathcal{P}}_i$ as in Lemma 8, define for each $P_i\in\mathcal{P}_i$ and any cluster $C\in P_i$ an arbitrary center $\upsilon(C)\in C$ and define

$$b(C) = \log\left(|B(\upsilon(C),4\Delta_i)|\right).$$

Fix some collection of partitions $P = \{P_i\in\mathcal{P}_i \mid i>0\}$, and let the label of a cluster $C\in P_i$ be $\alpha(C) = \frac{\Delta_i}{\vartheta(b(C))}$.

Claim 11. *For all $i>0$, if $C\in P_i$, $D\in P_{i-1}$ and $C\cap D\neq\emptyset$ then $\alpha(D)\leq 4\alpha(C)$.*

Proof. Since $d(\upsilon(C),\upsilon(D)) \leq \Delta_{i-1}+\Delta_i \leq 2\Delta_{i-1}$, we get $B(\upsilon(C),4\Delta_i)\subseteq B(\upsilon(D),4\Delta_i+2\Delta_{i-1})\subseteq B(\upsilon(D),4\Delta_{i-1})$ which suggests that $b(C)\leq b(D)$ hence $\alpha(C) = \frac{\Delta_i}{\vartheta(b(C))} = \frac{\Delta_{i-1}}{4\vartheta(b(C))} \geq \alpha(D)/4$. □

Note that a cluster $C\in P_i$ may have a label smaller than a cluster $D\in P_j$ for $j>i$ and $C\cap D\neq\emptyset$, hence creating a laminar family from the partition in the usual manner will not maintain the weak monotonicity property of labels. To overcome this hurdle we recursively define a sequence of hierarchical partitions $\mathcal{Q}^{(1)},\dots,\mathcal{Q}^{(\log\Delta_0)}$ where for each i, $\mathcal{Q}^{(i)}$ is a sequence of i partitions for scales Δ_1 to Δ_i. Initially $\mathcal{Q}^{(1)} = \{Q_1^{(1)} = P_1\}$. Given a hierarchical partition $\mathcal{Q}^{(i-1)} = \{Q_1^{(i-1)},\dots,Q_{i-1}^{(i-1)}\}$ and P_i we define a hierarchical partition $\mathcal{Q}^i = \{Q_1^{(i)},\dots,Q_i^{(i)}\}$ in the following manner.

1. **"Beam up" phase**: For any $C \in P_i$ and $j < i$ let $R_j(C) = \{D \in Q_j^{(i-1)} \mid D \cap C \neq \emptyset \wedge \alpha(D) < \alpha(C)\}$, let $s_j(C) = C \cap \bigcup_{D \in R_j(C)} D$. Intuitively, We want to "beam up" each $s_j(C)$ to be a cluster in Q_j. Formally, for any $j < i$, let $Q_j^{(i)} = \{D \setminus \bigcup_{\{C \mid D \in R_j(C)\}} C \mid D \in Q_j^{(i-1)}\} \cup \{s_j(C) \mid C \in P_i\}$. The labels are naturally maintained: each cluster $D \setminus \bigcup_{\{C \mid D \in R_j(C)\}} C$ gets label $\alpha(D)$ and each cluster $s_j(C)$ gets label $\alpha(C)$.
2. **"Laminarization" phase**: Let $Q_i^{(i)} = \{C \cap D \mid C \in P_i, D \in Q_{i-1}^{(i)}\}$. Each cluster $C \cap D$ gets label $\alpha(C)$.

For each set of partitions $P \in \mathcal{P}$ denote $Q = \mathcal{Q}^{(\log \Delta_0)} = \{Q_1, \ldots, Q_{\log \Delta_0}\}$. Note that the "laminarization" phase guarantees that Q is indeed hierarchical. Construct a labeled tree T from Q and its labels in the natural manner, where each node corresponding to a cluster in Q_i, has as children the nodes corresponding to clusters in Q_{i+1} that it contains, and as a parent the node corresponding to the cluster in Q_{i-1} that contains it. Note that T indeed represents an ultrametric, since the "beam up" phase guarantees that if $C \in Q_i, D \in Q_{i-1}$ such that $C \subseteq D$ then $\alpha(D) \geq \alpha(C)$.

Fix $x, y \in X$, and let $L = \min\{d(x,y), r_k(x)\}$. We begin by showing a bound on the contraction of the embedding. Let i be the smallest integer such that $5\Delta_i \leq L$, since P_i is Δ_i-bounded, x, y are separated in the partition P_i. Let $v = v(P_i(x))$, then

$$B(v, 4\Delta_i) \subseteq B(x, 5\Delta_i) \subseteq B(x, L) \subseteq B(x, r_k(x)),$$

which implies that $b(P_i(x)) \leq \log k$, hence $\alpha(P_i(x)) \geq \frac{\Delta_i}{\vartheta(\log k)} \geq \Omega\left(\frac{L}{\vartheta(\log k)}\right)$. It remains to show that $\alpha(Q_i(x)) \geq \alpha(P_i(x))$. This holds since if in the "beam up" phase some cluster replaced the part of $P_i(x)$ that contained x it must have had a larger label than $\alpha(P_i(x))$, and its radius is only smaller than the radius of $P_i(x)$ therefore $Q_i(x) \neq Q_i(y)$. This concludes the bound on the contraction.

We now turn to prove a bound on the expected expansion of the embedding. Let $\alpha_i = \max\{\alpha(P_i(x)), \alpha(P_i(y))\}$. Define the events

$$\mathcal{C}_i = \{P_i(x) \neq P_i(y)\},$$

$$\mathcal{M}_i = \left\{\mathcal{C}_i \wedge \bigwedge_{j>i} \alpha_i \geq \alpha_j\right\}.$$

Given a hierarchical partition Q, the distance between x, y in the tree T created from Q is the label of the cluster $Q_h(x)$, where h is the maximal such that $Q_h(x) = Q_h(y)$. Observe that if i is the minimal such that event $\mathcal{M}_i$ holds, then it is the case that $Q_i(x) \neq Q_i(y)$, because it cannot be that both $P_i(x)$, $P_i(y)$ are replaced in the "beam up" phase. We also claim that $d_T(x,y) \leq 2\alpha_i$. To prove this, fix the minimal j such that $s_j(P_i(x)) \neq \emptyset$ or $s_j(P_i(y)) \neq \emptyset$ (assume for now that $s_j(P_i(x)) \neq \emptyset$), then the "beam up" phase will make $s_j(P_i(x))$ a cluster in Q_j. This suggests that $\alpha(P_i(x)) \geq \alpha(P_j(x))$, and since we did not replace the $j-1$ level, also $\alpha_{j-1} \geq \alpha_i$, and by the minimality of i, it must be that $P_{j-1}(x) = P_{j-1}(y)$, thus also $Q_{j-1}(x) = Q_{j-1}(y)$. We conclude by Claim 11 that

$$d_T(x,y) = \alpha(Q_{j-1}(x)) \leq 4\alpha(P_j(x)) \leq 4\alpha(P_i(x)) = \frac{4\Delta_i}{\varphi(b(P_i(x)))} \leq \frac{4\Delta_i}{\varphi(\log|B(x,\Delta_i)|)}.$$

If it is the case that $s_j(P_i(y))$ was the one to "beam up", we get a similar bound, and conclude that

$$d_T(x,y) \leq \frac{4\Delta_i}{\varphi(\log|B(x,\Delta_i)|)} + \frac{4\Delta_i}{\varphi(\log|B(y,\Delta_i)|)}. \tag{12}$$

Let l be the maximal integer such that $\Delta_l \geq 8d(x,y)$. We also use Lemma 8 with parameter $\eta = d(x,y)/\Delta_i$ to bound

$$\Pr[\mathcal{C}_i] \leq 16 \frac{d(x,y) \log \rho(x, \Delta_i/4, 4)}{\Delta_i}, \tag{13}$$

and similar bound for y. Finally we bound the expected expansion,

$$\begin{aligned} \mathbb{E}_T[d_T(x,y)] &\leq \sum_{i>0} \Pr[\mathcal{M}_i] \cdot \mathbb{E}_T[d_T(x,y) \mid \mathcal{M}_i] & (14) \\ &\overset{(12)}{\leq} \sum_{i=1}^{l} \Pr[\mathcal{C}_i] \cdot \frac{4\Delta_i}{\varphi(\log|B(x,\Delta_i)|)} + \sum_{i=1}^{l} \Pr[\mathcal{C}_i] \cdot \frac{4\Delta_i}{\varphi(\log|B(y,\Delta_i)|)} + \sum_{i>l} 4\Delta_i, \end{aligned}$$

focus on the first summand of (14), then

$$\begin{aligned}
\sum_{i=1}^{l} \Pr[\mathcal{C}_i] \cdot \frac{4\Delta_i}{\varphi(\log|B(x,\Delta_i)|)} &\overset{(13)}{\leq} \sum_{i=1}^{l} 16 \frac{d(x,y)\log\rho(x,\Delta_i/4,4)}{\Delta_i} \cdot \frac{4\Delta_i}{\varphi(\log|B(x,\Delta_i)|)} \\
&\leq 64d(x,y)\sum_{i=1}^{l}(\log|B(x,\Delta_i)| - \log|B(x,\Delta_{i+4})|) \cdot \frac{1}{\varphi(\log|B(x,\Delta_i)|)} \\
&\leq 64d(x,y)\sum_{i=1}^{l}\sum_{j=\log|B(x,\Delta_{i+4})|}^{\log|B(x,\Delta_i)|}\frac{1}{\varphi(j)} \\
&\leq 256d(x,y)\sum_{j>0}\frac{1}{\varphi(j)} \\
&= O(d(x,y)) ,
\end{aligned} \tag{15}$$

The second summation of (14) is bounded is the same manner, and for the third one, $\sum_{i>l} 4\Delta_i \leq 4\Delta_l \leq O(d(x,y))$.

This conclude the bound on the expansion, it remains to show that the worst case contraction can be bounded by $O(\log n)$. Let $\bar{\vartheta}(k) = \min\{\vartheta(k), \log n\}$, and use $\bar{\vartheta}$ instead of ϑ when defining the labels. Notice that labels can only increase, hence the lower bound remains true, and furthermore for any pair x, y we have $d_T(x,y) \geq \frac{d(x,y)}{\log n}$. It remains to show that the expected expansion remains a universal constant. Let t be the largest integer such that $\vartheta(\log|B(x,\Delta_t)|) > \log n$. Then divide the term in (15) into two summations, up to t and after, to bounds the first,

$$\sum_{i=1}^{t}(\log|B(x,\Delta_i)| - \log|B(x,\Delta_{i+4})|) \cdot \frac{1}{\bar{\vartheta}(\log|B(x,\Delta_t)|)} \leq \frac{4}{\log n}(\log|B(x,\Delta_1)| - \log|B(x,\Delta_{t+4})|) \leq 4,$$

and as in (15),

$$\sum_{i=t+1}^{l}(\log|B(x,\Delta_i)| - \log|B(x,\Delta_{i+4})|) \cdot \frac{1}{\bar{\vartheta}(\log|B(x,\Delta_t)|)} \leq \sum_{i=t+1}^{l}\sum_{j=b_{i+4}}^{b_i}\frac{4}{\vartheta(j)} = O(1) .$$

This concludes the proof of Theorem 6.

5.3 Lower Bound for Spanning Trees

An important variant in embedding into trees occurs in a graph setting, when we seek an embedding into a spanning tree of the graph. Probabilistic embedding into spanning trees has been studied in [AKPW95, EEST05, ABN08b]. In [ABN07a] embeddings into a single spanning tree and into a distribution on spanning trees, with constant average distortion are shown. However, local embedding into a single spanning tree can incur distortion $n-1$ even for $k=1$ (take the cycle graph, finding a spanning tree is done by removing some edge, which will incur the distortion for an adversarial choice of nearest neighbors). In what follows we show that unlike the general trees setting, probabilistic embedding into a distribution of *spanning* trees cannot overcome the $\Omega(\log n)$ lower bound even for $k=1$.

Theorem 7. *There exists a graph G on n vertices such that any embedding of its shortest-path metric into a distribution of spanning trees of G, will incur 1-local distortion of $\Omega(\log n)$.*

Proof. Let (X,d) be an n point metric space induced by a graph G on n vertices and $2n$ edges, whose girth (length of shortest cycle) is $c\log n + 1$, for some universal constant c. Such graphs are known to exist [Bol78]. Note that every edge in G is part of a cycle of length at least $c\log n$, therefore removing that edge will increase the distance between its end points to $c\log n$. Also note that every spanning tree of G is obtained by removing at least half of the edges, hence in any distribution over spanning trees there exists some edge (u,v) with $\Pr_T[(u,v) \notin T] \geq 1/2$, thus $\mathbb{E}[d_T(u,v)] \geq (c/2)\log n$. ◻

6 Embedding into ℓ_p with k-Local Distortion and Low Dimension

In this section we strengthen the results of Section 4.2 by providing an embedding with both k-local distortion *and dimension* depending on k and not on n, the number of points. Our main result is an embedding with optimal k-local distortion of $O(\log k)$, and dimension $O(\log^2 k)$ for arbitrary metrics. Later we will show a tight $O(\log k)$ bound on the dimension as well, for metrics satisfying a weak growth bound.

6.1 Main Result

Theorem 8. *For any n point metric space (X, d), a parameter $k \le n$, and an integer p satisfying $p \le \ln k/2$ there exists an embedding into ℓ_p with k-local distortion $O((\log k)/p)$ and dimension $O(e^p \log^2 k)$.*

Let $s = e^p$, $D = cs \ln^2 k$ for some universal constant c to be determined later. The proof of this theorem will require a composition of two functions $f : X \to \mathbb{R}^D$ and $g : X \to \mathbb{R}^D$ with the following properties:

1. The functions f, g are non-expansive, *i.e.* for all $u, v \in X$
$$\|f(u) - f(v)\|_p \le d(u, v) \ , \ \|g(u) - g(v)\|_p \le d(u, v)$$
2. For any pair $u, v \in X$ such that $v \in N_k(u)$ and $d(u, v) < r_k(u)/8$,
$$\|f(u) - f(v)\|_p > Cp \cdot d(u, v)/\log k$$
for a universal constant C.
3. For any pair $u, v \in X$ such that $v \in N_k(u)$ and $r_k(u)/8 \le d(u, v) \le r_k(u)$,
$$\|g(u) - g(v)\|_p > C'p \cdot d(u, v)/\log k$$
for a universal constant C'.

The embedding is defined as $f \oplus g$, and it follows directly that the dimension is $O(e^p \log^2 k)$ and from the properties that the k-local distortion is $O((\log k)/p)$.

A main tool that will assist us in obtaining the low dimension is Lovász Local Lemma. Note that the embedding of Section 4.2 only uses logarithmic number of dimensions in order to obtain high success probability for all pairs. If we could argue that the event that a certain k nearest neighbor pair has bounded contraction depends only on $\mathrm{poly}(k)$ other k nearest neighbor pairs, we could have used the local lemma, and its algorithmic versions [Bec91, Mos09, MT10] to construct an embedding which is good for all k nearest neighbor pairs using $\mathrm{poly}(\log k)$ dimensions. Unfortunately, this is not the case, as it could be that for some $u \in X$ there are many points whose distance to u is slightly larger than $r_k(u)$, and these points have some k-nearest neighbor inside $B(u, r_k(u))$, so there will be many dependencies. In Section 6.2 we show that assuming a weak growth bound will solve this issue, and we may obtain optimal dimension of $O(\log k)$ with optimal distortion $O(\log k)$. For the general setting, we can only use the local lemma for pairs u, v that are "close" with respect to $r_k(u)$, as defined in the second property. Thus the map f is very similar to the map of Section 4.2, with fewer dimension and analysis that uses the Local Lemma. The main difficulty is construction the map g, which provides bounded contraction for pairs that are "far-away" with respect to $r_k(u)$. From a high level, the map g is based on padded probabilistic partitions, similarly to maps of [Rao99, KLMN04, ABN06], however there are several subtle differences whose combination yields the desired result. We highlight two of the new ideas here:

1. In order to define the map g, we use a new type of probabilistic partition, where clusters are bounded not by their diameter but by the number of points they contain. Since we need to apply the Local Lemma, the padding probability must depend only on local events. A related partitioning notion was suggested by Charikar, Makarychev and Makarychev in [CMM10], however their partition algorithm was based on the probabilistic partitions of [CKR01, FRT03], which are inherently non-local and hence cannot be used for our application. The construction of our bounded cardinality probabilistic partition uses the truncated exponential distribution approach of [Bar96, ABN06]. The proof requires some technical modifications to adapt to the bounded cardinality case (see Definition 11).

2. A common use of probabilistic partitions for embeddings is to randomly color each cluster by 0 or 1, or uniformly in $[0,1]$ (see [Rao99, KLMN04, ABN06]). This typically means that the distortion of a pair depends on the color event of the clusters of both vertices. Even in the local setting it could be that some nodes participates in many pairs (for example the center node in a star metric), then this may create dependencies among many pairs and hence prohibit the use of the Local Lemma. To overcome this issue without any growth bound assumptions, we deterministically color each cluster into a $\bar{D} = \Theta(\log k)$ dimensional vector in such a way that if v is among the k nearest neighbors of u and u, v belong to different clusters C, C' then the hamming distance between the colors of C and C' is at least $\bar{D}/8$. This allows to define the success event for the pair u, v for the map g only as a function of the probabilistic partition around u *independent* of the events around v.

6.1.1 The "Small" Distance Map

First we describe the map f, it is essentially the same map f produced in Section 4.2, with the difference of choosing $q = c \cdot s \log k$. In order for the same analysis to go through (observe that the distortion does not depend on q), it suffices to verify that for any $u, v \in X$, (6) holds for at least $q/(24s)$ values of $j \in [q]$.

Fix some $u \in X$, $v \in N_k(u)$ with $d(u,v) \le r_k(u)/8$ and $i \in [m]$ (m was the minimal integer such that $\rho_m + \rho_{m-1} \ge d(u,v)/4$). Recall that for some $j \in [q]$, events $\mathcal{E}_j$ and $\mathcal{F}_j$ suggest that (6) will hold. Observe that these events depend on the inclusion in A_{ij} of some points in $B(u, d(u,v)) \cup B(v, d(u,v)) \subseteq B(u, 2d(u,v))$ (because $\rho_i \le d(u,v)$ for all $i \le m$). Let $\mathcal{A}_i(u,v)$ be the event that there are at least $q/(24s)$ values of $j \in [q]$ for which both events $\mathcal{E}_j$ and $\mathcal{F}_j$ hold, let $\mathcal{A}(u,v)$ be the event that $\mathcal{A}_i(u,v)$ holds for all $i \in [m]$, and let $\mathcal{A}$ be the event that $\mathcal{A}(u,v)$ hold for all $u \in X$ and $v \in N_k(u)$. We will show by the Lovász Local Lemma that $\Pr[\mathcal{A}] > 0$, and by its algorithmic versions we can find a good map f in polynomial time.

Define a dependency graph whose vertices are the events $\mathcal{A}(u,v)$ for all $u, v \in X$ such that $d(u,v) \le r_k(u)/8$, and two events $\mathcal{A}(u,v)$, $\mathcal{A}(u',v')$ are connected by an edge iff $u' \in \bar{N}_k(u)$ (recall that $\bar{N}_k(u) = \{x \in N_k(u) \mid u \in N_k(x)\}$, so this is a symmetric relation). The degree of the graph is at most k^2, because there are at most k points $u' \in \bar{N}_k(u)$, and each u' has at most k points v' satisfying $d(u',v') \le r_k(u')/8$. Assume that events $\mathcal{A}(u,v)$ and $\mathcal{A}(u',v')$ are not connected by an edge, we will show that the balls $B(u, 2d(u,v))$ and $B(u', 2d(u',v'))$ are disjoint, thus the events are indeed independent. To see this, first we note that $r_k(u) \le 2d(u,u')$ and also $r_k(u') \le 2d(u,u')$. Because if, for instance, the former does not hold, then $d(u,u') < r_k(u)/2 \le (d(u,u') + r_k(u'))/2$, so $u' \in N_k(u)$ and also $d(u,u') < r_k(u')$ thus $u \in N_k(u')$, which contradicts the fact that $u' \notin \bar{N}_k(u)$. A symmetric argument shows that the latter must hold as well. We conclude that

$$2\left(d(u,v) + d(u',v')\right) \le \frac{1}{4}\left(r_k(u) + r_k(u')\right) \le \frac{1}{4}\left(2d(u,u') + 2d(u,u')\right) = d(u,u') \ .$$

Let X_{ij} be indicator random variable for the event that both $\mathcal{E}_j$ and $\mathcal{F}_j$ hold, and let $X_i = \sum_{j=1}^q X_{ij}$. By the calculation done in Section 4.2 we have that $\mathbb{E}[X_i] \ge q/(12s) = c/12 \cdot \log k$, and by Chernoff bound

$$\Pr[\neg \mathcal{A}_i(u,v)] = \Pr[X_i < \mathbb{E}[X_i]/2] \le e^{-\mathbb{E}[X_i]/4} \le e^{-(c \log k)/48} = k^{-4} \ ,$$

where the last equation is by choosing c as a large enough constant. By the union bound (recall that $m \le t < k$)

$$\Pr[\neg \mathcal{A}(u,v)] = \Pr[\exists i \in [m], \neg \mathcal{A}_i(u,v)] \le t \cdot k^{-4} \le k^{-3} \ .$$

Now by Lemma 3 on the dependency graph we defined, with degree $d = k^2$ and $p = k^{-3}$, there is some positive probability that all the good events $\mathcal{A}(u,v)$ hold simultaneously.

6.1.2 The "Large" Distances Embedding

We now detail the map g, that has bounded contraction for pairs such that $r_k(u)/8 \le d(u,v) \le r_k(u)$.

Recall that $s = e^p$, and let $\delta = 1/s$, observe that $\delta = e^{-p} \ge 2/k$. Let $D' = \hat{D} \cdot \bar{D}$ where $\hat{D} = cs \cdot \ln k$ for a constant c to be determined later, and $\bar{D} = 16 \log k$. Let $\hat{\mathcal{P}}$ be a k-bounded locally padded probabilistic partition with parameter δ, as in Lemma 6. For each $t \in [\hat{D}]$ fix some $P = P^{(t)} \in \mathcal{P}$ (the particular choice of P will be detailed later in Lemma 15). Define a directed graph $G = (V, A)$, which will be the k-neighborhood graph between the clusters of the partition P. Let the vertex set V be the clusters of P. Draw a directed edge (C, C') between clusters C and C' iff there exists points $x \in C$, $y \in C'$ such that $y \in N_k(x)$. As every cluster contains at most k points, the out-degree of G is at most k^2.

We use the following property of directed graphs with bounded out-degree.

Lemma 12. *Any directed graph $G = (V, A)$ with maximal out-degree k can be properly colored*[5] *using $2k+1$ colors.*

Proof. The proof is by induction. Assume we can color with $2k + 1$ colors any graph on less that $|V|$ vertices, whose out-degree is bounded by k. Since $|A| \le k \cdot |V|$, there must be a vertex $x \in V$ of total degree[6] at most $2k$. Create a graph H by removing x and all the edges touching x from G, and note that the degree of every vertex in H is still trivially bounded by k. Using the induction hypothesis, properly color H with $2k + 1$ colors. Now we add x back to the graph, since it has at most $2k$ edges touching it, we can color it with a color none of its neighbors has. □

We also use a set S of vectors in $\{-1, 1\}^{O(\log k)}$ such that any two points in S are "far" from each other.

Lemma 13. *For any integer $\bar{D} > 1$ and $\Omega(1/\bar{D}) < \delta \le 1/2$ there exists a set $S \subseteq \{-1, 1\}^{\bar{D}}$, $|S| \ge 2^{\bar{D}(1-H(\delta))/2}$ (H is the entropy function), such that for any $u, v \in S$, the Hamming distance between u and v is at least $\delta\bar{D}$.*

There is a randomized procedure to produce such a set S, which is a classical result in error correcting codes [Gil52]. In particular, fixing $\delta = 1/8$ and recalling that $\bar{D} = 16 \log k$ we get a set S of $2k^2 + 1$ vectors in $\{-1, 1\}^{\bar{D}}$ such that the Hamming distance between each two vectors is at least $\bar{D}/8$. Using Lemma 12 we can properly color G with $m = 2k^2 + 1$ colors, and define $\sigma = \sigma^{(t)} : V \to S$, such that if $(C, C') \in A$ then $\sigma(C) \ne \sigma(C')$, by giving each color class of V a distinct vector in S. For any $t \in [\hat{D}]$ define $g^{(t)} : X \to \mathbb{R}^{\bar{D}}$ by

$$g^{(t)}(u) = \bar{D}^{-1/p} \cdot d(u, X \setminus P^{(t)}(u)) \cdot \sigma(P^{(t)}(u)).$$

The embedding $g : X \to \mathbb{R}^{D'}$ is the normalized concatenation of the $g^{(t)}$s,

$$g(u) = \hat{D}^{-1/p} \bigoplus_{t=1}^{\hat{D}} g^{(t)}(u)$$

Observe that for any cluster $C \in P$, $\sigma(C)$ is a $\bar{D} = O(\log k)$ dimensional vector hence $g^{(t)}$ is a mapping into $\bar{D}$ dimensions and g is a mapping into $D' = \hat{D} \cdot \bar{D} = O(e^p \log^2 k)$ dimensions.

Lemma 14. *For any $u, v \in X$, $\|g^{(t)}(u) - g^{(t)}(v)\|_p \le O(d(u, v))$.*

Proof. Fix any $P = P^{(t)}$. We distinguish between two cases

Case 1: $P(u) = P(v)$. Denote by $(a_1, \ldots a_{\bar{D}}) = \sigma(P(u)) = \sigma(P(v))$, then as $\big|d(u, X \setminus P(u)) - d(v, X \setminus P(u))\big| \le d(u, v)$,

$$\|g^{(t)}(u) - g^{(t)}(v)\|_p^p = \bar{D}^{-1}\big|d(u, X \setminus P(u)) - d(v, X \setminus P(u))\big|^p \sum_{i=1}^{\bar{D}} |a_i|^p \le d(u, v)^p.$$

Case 2: $P(u) \ne P(v)$, then $d(u, X \setminus P(u)) \le d(u, v)$ and also $d(v, X \setminus P(v)) \le d(u, v)$, hence

$$\|g^{(t)}(u) - g^{(t)}(v)\|_p^p \le \|g^{(t)}(u)\|_p^p + \|g^{(t)}(v)\|_p^p \le 2\bar{D}^{-1}\big|d(u, v)\big|^p \sum_{i=1}^{\bar{D}} 1^p \le (2d(u, v))^p.$$

[5] Properly colored means that the end points of every directed edge are colored by different colors.

[6] The total degree of a vertex in a directed graph is the number of edges touching the vertex.

$\square$

Lemma 15. *There exist partitions $P^{(t)} \in \mathrm{supp}(\hat{\mathcal{P}})$ for each $t \in [\hat{D}]$, such that for any $u, v \in X$ with $v \in N_k(u)$ and $r_k(u)/8 < d(u,v) \le r_k(u)$,*

$$\|g(u) - g(v)\|_p \ge \Omega(p \cdot d(u,v)/\log k) \ .$$

Proof. Fix any $t \in [\hat{D}]$ and let $P = P^{(t)}$. From the first property of Definition 11, $v \notin P(u)$. Since $v \in N_k(u)$ by the proper coloring $\sigma(P(u)) \ne \sigma(P(v))$. Let $I_{uv} = I_{uv}(t) \subseteq [\bar{D}]$ be the subset of at least $\bar{D}/8$ coordinates such that for any $i \in I_{uv}$ we have $\sigma(P(u))_i \ne \sigma(P(v))_i$, and note that for any two positive numbers a, b we have that $|a \cdot \sigma(P(u))_i - b \cdot \sigma(P(v))_i| = a + b$. By the second property of Definition 11 with probability $1/s$ we have that u is padded, if it holds then $d(u, X \setminus P(u)) \ge 2^{-9} r_k(u) \cdot \log s/\log k \ge 2^{-9} p \cdot d(u,v)/\log k$. Given the padding event,

$$\begin{aligned}
\|g^{(t)}(u) - g^{(t)}(v)\|_p^p &\ge \bar{D}^{-1} \sum_{i \in I_{uv}} (d(u, X \setminus P(v)) + d(v, X \setminus P(v)))^p \\
&\ge |I_{uv}|/\bar{D} \cdot d(u, X \setminus P(u))^p \\
&\ge (1/8) \cdot (2^{-9} p \cdot d(u,v)/\log k)^p.
\end{aligned}$$

Let $\mathcal{Z}_t$ be an indicator for the event that u is padded in $P^{(t)}$. Note that this event is sufficient for obtaining bounded contraction in the t-th coordinate. Define $\mathcal{E}_{u,v}$ as an indicator random variable for the event that exists a subset $T \subseteq [\hat{D}]$ of size at least $\hat{D}/(2s)$ such that for all $t \in T$, $\mathcal{Z}_t$ holds. Note that if $\mathcal{E}_{u,v}$ holds then

$$\|g(u) - g(y)\|_p^p \ge \hat{D}^{-1} \sum_{t \in T} \|g^{(t)}(u) - g^{(t)}(v)\|_p^p \ge \Omega((1/s) \cdot (p \cdot d(u,v)/\log k)^p).$$

As required, so it remains to show that there exists some choice of randomness such that all events $\mathcal{E}_{u,v}$, for pairs $u, v \in X$ such that $r_k(u)/8 < d(u,v) \le r_k(u)$, hold simultaneously.

Let $\mathcal{Z} = \sum_{t \in [\hat{D}]} \mathcal{Z}_t$, then $\mathbb{E}[\mathcal{Z}] \ge \hat{D}/s$. In order for $\mathcal{E}_{u,v}$ to hold, we need that $\mathcal{Z} \ge \hat{D}/(2s)$. Using Chernoff bound,

$$\Pr[\mathcal{Z} \le \hat{D}/(2s)] = \Pr \mathcal{Z} \le \mathbb{E}[\mathcal{Z}]/2] \le e^{-\hat{D}/(8s)} \le 1/(4k^2) \ ,$$

where the last inequality holds for $c \ge 16 \ln 4$. Define a dependency graph whose vertices are events $\mathcal{E}_{u,v}$, and draw an edge $(\mathcal{E}_{u,v}, \mathcal{E}_{x,y})$ iff $x \in \bar{N}_k(u)$ (note that this is a symmetric definition). The degree of the graph is at most k^2. Recall that the second property of Definition 11 states that the padding probability of any $x \in X$ is essentially determined by the partition restricted to $B(x, r_k(x)/4)$. If $\mathcal{E}_{u,v}$ is not connected by an edge in the dependency graph to $\mathcal{E}_{x,y}$, then it must be that $B(u, r_k(u)/4) \cap B(x, r_k(x)/4) = \emptyset$. Assume by contradiction that this is not so, then $d(u,x) \le r_k(u)/4 + r_k(x)/4$, we also know that $r_k(x) \le d(x,u) + r_k(u) \le 5r_k(u)/4 + r_k(x)/4$ so $r_k(x) \le 2r_k(u)$, and then $d(u,x) \le r_k(u)/4 + r_k(x)/4 \le r_k(u)/4 + 2r_k(u)/4 < r_k(u)$. An identical calculation shows that $d(u,x) < r_k(x)$, and this contradicts the fact that $x \notin \bar{N}_k(u)$. We conclude that we may condition on events that are not connected by an edge to $\mathcal{E}_{u,v}$, and the bound on the padding probability remains. By applying Lemma 3 with probability $p = 1/(4k^2)$ and degree $d \le k^2$ we conclude that there is a choice of randomness for which all events $\mathcal{E}_{u,v}$ hold simultaneously. $\square$

6.2 Optimal Dimension for Weak Growth Bounded Metrics

Theorem 9. *For any finite metric space (X, d) on n points with a χ-weak growth bound there exists an embedding into ℓ_p with k-local distortion $O(\mathrm{poly}(\chi) \cdot \log k)$ and dimension $O(\mathrm{poly}(\chi) \cdot \log k)$.*

The proof will follow the lines of the "small" distance map of Section 6.1.1, with two main differences. We will use embedding that are based on locally padded probabilistic partitions, similar in spirit to those used in [ABN06], instead of Bourgain type embedding, to achieve a better dimension of $O(\log k)$. This creates some complications: Bourgain type embeddings are inherently defined with respect to densities, so we could simply eliminate densities higher than k. However, in the partition based embedding, we need to carefully delete contribution from dense clusters, and argue that it does not harm the analysis. In order to apply the Lovász Local Lemma, we will use the growth bound to argue that there are few dependencies between pairs of k-nearest neighbors.

Let $D = c \ln k$, for some constant c to be defined later. We will define an embedding $f : X \to \mathbb{R}^D$ with k-local distortion $O(\log k)$. We define f by defining for each $1 \le t \le D$, a function $f^{(t)} : X \to \mathbb{R}_+$, and letting $f = D^{-1/p} \bigoplus_{1 \le t \le D} f^{(t)}$. Fix t, $1 \le t \le D$. In what follows we define $f^{(t)}$. Let $\Delta_0 = \mathrm{diam}(X)$ and for every integer $i > 0$ let $\Delta_i = \Delta_0/4^i$. We construct for all $i > 0$ a Δ_i-bounded η_i-uniformly locally padded probabilistic partition $\hat{\mathcal{P}}_i$ as in Lemma 5 with parameter $\gamma = 64$. Let ξ_i be as defined in the lemma. Denote by Ω the probability space of all possible embeddings f. Now for every $i > 0$ fix a partition $P_i \in \mathcal{P}_i$. We define the embedding by defining the coordinates for each $x \in X$. For every cluster $C \in P_i$ fix an arbitrary center $v(C) \in C$. Let $\ell(x)$ be the minimal integer such that $|B(v(P_{\ell(x)}(x)), 3\gamma\Delta_{\ell(x)})| \le k^{\chi^4}$.

We now define $\bar{\xi}$ in the following manner :

$$\bar{\xi}_i(x) = \begin{cases} 0 & i < \ell(x) \\ \xi_i(x) & \text{otherwise} \end{cases}$$

Define for $x \in X$, $0 < i \in I$, $\phi_i^{(t)} : X \to \mathbb{R}_+$, by $\phi_i^{(t)}(x) = \bar{\xi}_i(x)/\eta_i(x)$. Lemma 5 and the definition of $\ell(x)$ ensures that $\bar{\xi}_i$ and η_i are uniform functions with respect to $\hat{\mathcal{P}}_i$ so we have:

Claim 16. *For any $x, y \in X$ and $i > 0$, if $P_i(x) = P_i(y)$ then $\phi_i^{(t)}(x) = \phi_i^{(t)}(y)$.*

Claim 17. *For any $x \in X$, $t \in [D]$,*

$$\sum_{i>0} \phi_i^{(t)}(x) \le 2^{10}\chi^4 \log k$$

Proof. By the second property of Lemma 5 and the definition of $\bar{\xi}_i$, if $\bar{\xi}_i = 1$ then in particular it must be $\xi_i(x) = 1$ and so $1/\eta_i(x) \le 2^7 \log \rho(x, 2\Delta_i, \gamma)$. Now,

$$\begin{aligned} \sum_{i>0} \phi_i^{(t)}(x) &= \sum_{i>0|\bar{\xi}_i(x)=1} \eta_i^{-1}(x) \\ &\le \sum_{i \ge \ell(x)} 2^7 \log \left(\frac{|B(x, 2\gamma\Delta_i)|}{|B(x, 2\Delta_i/\gamma)} \right) \\ &\le 2^9 \cdot \chi^4 \log k. \end{aligned}$$

The last in-equation holds since $d(x, v(P_{\ell(x)}(x))) \le \Delta_{\ell(x)}$, and so $|B(x, 2\gamma\Delta_{\ell(x)})| \le |B(v(P_{\ell(x)}(x)), 3\gamma\Delta_{\ell(x)})| \le k^{\chi^4}$, so the summation telescopes to $4\chi^4 \log k$. □

For each $0 < i$ we define a function $f_i^{(t)} : X \to \mathbb{R}_+$ and for $x \in X$, let $f^{(t)}(x) = \sum_{i>0} f_i^{(t)}(x)$. Let $\{\sigma_i^{(t)}(C) | C \in P_i, 0 < i\}$ be i.i.d symmetric $\{0, 1\}$-valued Bernoulli random variables. The embedding is defined as follows: for each $x \in X$:

- For each $0 < i$, let $f_i^{(t)}(x) = \sigma_i^{(t)}(P_i(x)) \cdot \min\{\phi_i^{(t)}(x) \cdot d(x, X \setminus P_i(x)), \Delta_i\}$.

The following claim was essentially proved in [ABN06], we include the proof for completeness

Claim 18. *For any $x, y \in X$ and $0 < i$: $f_i^{(t)}(x) - f_i^{(t)}(y) \le \min\{\phi_i^{(t)}(x) \cdot d(x, y), \Delta_i\}$.*

Proof. First note that for any set $A \subset X$ and positive real r, by the triangle inequality

$$\min\{d(x, A), r\} - \min\{d(y, A), r\} \le \min\{d(x, y), r\} \; . \tag{16}$$

If it is the case that $P_i(x) = P_i(y)$, then using the uniformity of $\phi_i^{(t)}$ and (16),

$$\begin{aligned} f_i^{(t)}(x) - f_i^{(t)}(y) &= \sigma_i^{(t)}(P_i(x)) \cdot \left(\min\{\phi_i^{(t)}(x) \cdot d(x, X \setminus P_i(x)), \Delta_i\} - \min\{\phi_i^{(t)}(x) \cdot d(y, X \setminus P_i(x)), \Delta_i\} \right) \\ &\le \min\{\phi_i^{(t)}(x) \cdot d(x, y), \Delta_i\} \; . \end{aligned}$$

Otherwise, if $P_i(x) \neq P_i(y)$ then $d(x, X \setminus P_i(x)) \leq d(x,y)$, and then

$$f_i^{(t)}(x) - f_i^{(t)}(y) \leq f_i^{(t)}(x) \leq \min\{\phi_i^{(t)}(x) \cdot d(x,y), \Delta_i\} .$$

$\square$

Lemma 19. *There exists a universal constant $C_1 > 0$ such that for any $x, y \in X$:*

$$|f^{(t)}(x) - f^{(t)}(y)| \leq C_1 \chi^4 \log k \cdot d(x,y).$$

Proof. From Claims 17 and 18 we get

$$\begin{aligned} f^{(t)}(x) - f^{(t)}(y) &= \sum_{i>0} f_i^{(t)}(x) - f_i^{(t)}(y) \\ &\leq \sum_{i>0} \phi_i^{(t)}(x) \cdot d(x,y) \\ &\leq 2^9 \chi^4 \log k \cdot d(x,y) \end{aligned}$$

$\square$

6.2.1 Lower Bound Analysis

For every pair of points x, y such that $y \in N_k(x)$, we define a critical scale $i = i_{xy}$, which is the unique integer satisfying $6\Delta_i \leq d(x,y) < 6\Delta_{i-1}$. If N is the set of all such x, y pairs, we partition it according to the critical scales, that is, $N_i = \{(x,y) \mid i_{xy} = i \wedge y \in N_k(x)\}$, and let $N = \bigcup_{i>0} N_i$. Fix some x, y and $i = i_{xy}$, define the event $\mathcal{Z}_t(x,y)$ as

$$\left(|f_i^{(t)}(x) - f_i^{(t)}(y)| \geq \Delta_i \wedge \left| \sum_{j<i} f_j^{(t)}(x) - f_j^{(t)}(y) \right| \leq \frac{\Delta_i}{2} \right) \bigvee$$
$$\left(f_i^{(t)}(x) = f_i^{(t)}(y) = 0 \wedge \left| \sum_{j<i} f_j^{(t)}(x) - f_j^{(t)}(y) \right| > \frac{\Delta_i}{2} \right) .$$

Also define a function $g_i : N_i \to 2^D$ as follows

$$g_i(x,y) = \{t \in [D] \mid \mathcal{Z}_t(x,y)\} ,$$

and let $\mathcal{Z}(x,y)$ be the event that $|g_i(x,y)| \geq D/16$. Finally let $\mathcal{Z} = \bigcap_{(x,y)\in N} \mathcal{Z}(x,y)$, then we would like to show the following

Lemma 20.

$$\Pr[\mathcal{Z}] > 0.$$

Before proving this Lemma, let us see that it is sufficient to bound the contraction of the embedding.

Claim 21. *For any integer $i > 0$, $(x,y) \in N_i$ and $t \in g_i(x,y)$,*

$$|f^{(t)}(x) - f^{(t)}(y)| > d(x,y)/2^9 .$$

Proof. By Claim 18 and since Δ_j is a geometric series

$$|\sum_{j>i} f_j^{(t)}(x) - f_j^{(t)}(y)| \leq \sum_{j>i} \Delta_j = \Delta_i/3 .$$

Since $\mathcal{Z}_t(x,y)$ holds $|\sum_{j\le i} f_j^{(t)}(x) - f_j^{(t)}(y)| \ge \Delta_i/2$, so

$$|\sum_{j>0} f_j^{(t)}(x) - f_j^{(t)}(y)| \ge |\sum_{j\le i} f_j^{(t)}(x) - f_j^{(t)}(y)| - |\sum_{j>i} f_j^{(t)}(x) - f_j^{(t)}(y)| \ge \Delta_i/6 \ge d(x,y)/2^8 \, .$$

□

Lemma 22. *There exists a universal constant C_2 and a choice of embedding f such that for any $x, y \in N$:*

$$|f(x) - f(y)| \ge C_2 d(x,y).$$

Proof. Using Lemma 20, let f be an embedding such that event $\mathcal{Z}$ holds. Consider any integer $i > 0$, $(x,y) \in N_i$ and $t \in g_i(x,y)$. Using Claim 21 and the fact that $|g_i(x,y)| \ge D/16$,

$$\begin{aligned}
\|f(x) - f(y)\|_p^p &= \frac{1}{D}\sum_{t\in D} |f^{(t)}(x) - f^{(t)}(y)| \\
&\ge \frac{1}{D}\sum_{t\in g_i(x,y)} |f^{(t)}(x) - f^{(t)}(y)| \\
&\ge (1/16)\frac{d(x,y)}{2^8}
\end{aligned}$$

□

Proof of Lemma 20. We shall make use the following variation of the Local Lemma, in which the bad events have rating, and events may only depend on other events with equal or larger rating.

Lemma 23 (Local Lemma)**.** *Let $\mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_n$ be events in some probability space. Let $G(V, A)$ be a directed graph on n vertices with out-degree at most d, each vertex corresponding to an event. Let $c : V \to [m]$ be a rating function of events, such that if $(\mathcal{A}_i, \mathcal{A}_j) \in A$ then $c(\mathcal{A}_i) \le c(\mathcal{A}_j)$. Assume that for any $i = 1, \ldots, n$*

$$\Pr\left[\mathcal{A}_i \mid \bigwedge_{j\in Q} \neg\mathcal{A}_j\right] \le p$$

for all $Q \subseteq \{j : (\mathcal{A}_i, \mathcal{A}_j) \notin A \wedge c(\mathcal{A}_i) \ge c(\mathcal{A}_j)\}$. If $ep(d+1) \le 1$, then

$$\Pr\left[\bigwedge_{i=1}^{n} \neg\mathcal{A}_i\right] > 0$$

Proof. We iteratively apply the Lovász Local Lemma on every rating level $k \in [m]$, and prove the property by induction on k. For $k \in [m]$ denote by $V_k \subseteq V$ all the events with rating k, and by $G_k = (V_k, E_k)$ the induced subgraph on V_k. The base of the induction $k = 1$, by the assumption for all $\mathcal{A}_i \in V_1$,

$$\Pr\left[\mathcal{A}_i \mid \bigwedge_{j\in Q} \neg\mathcal{A}_j\right] \le p,$$

for any Q satisfying $Q \subseteq \{j : (\mathcal{A}_i, \mathcal{A}_j) \notin E_1 \wedge c(\mathcal{A}_j) = 1\}$. This means that by the usual local lemma on the graph G_1 there is a choice of randomness for which all the bad events in V_1 do not occur.

Fix some $k \in [m]$ and assume all events in $V_1, \ldots V_{k-1}$ do not hold. Note that by definition an event in V_k depends only on events of rating k or higher, so given that events in $V_1, \ldots V_{k-1}$ are fixed to not happen, for all $\mathcal{A}_i \in V_k$ by the assumption

$$\Pr\left[\mathcal{A}_i \mid \bigwedge_{j\in Q} \neg\mathcal{A}_j\right] \le p,$$

for any Q satisfying $Q \subseteq \{j : (\mathcal{A}_i, \mathcal{A}_j) \notin E_k \wedge c(\mathcal{A}_j) = k\} \cup \{j : \mathcal{A}_j \in V_1 \cup \cdots \cup V_{k-1}\}$. So once again by the usual local lemma on G_k there is non-zero probability that all the events in V_k do not occur. $\square$

Define a directed graph $G = (V, A)$ where $V = \{\mathcal{Z}(x,y) \mid (x,y) \in N_i\}$, define the ranking function as $c(\mathcal{Z}(x,y)) = i_{xy}$ and

$$(\mathcal{Z}(x,y), \mathcal{Z}(x',y')) \in A \Leftrightarrow d(x,x') \leq 6r_k(x) \wedge c(\mathcal{Z}(x,y)) \leq c(\mathcal{Z}(x',y')),$$

note that the definition is not symmetric, G is a directed graph, and that the rating matches the requirements of Lemma 23.

Claim 24. *The out-degree of G is bounded by k^{χ^4}.*

Proof. Fix any vertex $\mathcal{Z}(x,y) \in V$. By the weak growth bound condition $|B(x, 6r_k(x))| \leq |B^\circ(x, r_k(x))|^{\chi^3} \leq k^{\chi^3}$, and for each $x' \in B(x, 6r_k(x))$ there are at most k possible values of y' such that $(x', y') \in N$, hence the out degree is bounded by k^{χ^4}. $\square$

The following claim will establish that pairs in N_i that are not connected by an edge in G, are far apart with respect to Δ_i.

Claim 25. *Let $(x,y) \in N_i$. If $(x',y') \in N_i$ such that $(\mathcal{Z}(x,y), \mathcal{Z}(x',y')) \notin A$ then $d(\{x,y\}, \{x',y'\}) > 4\Delta_i$.*

Proof. By definition of N_i both $d(x,y), d(x',y') < 6\Delta_{i-1}$. By definition of G, and since $d(x,y) \leq r_k(x)$, we have that $d(x,x') > 6r_k(x) \geq 6d(x,y)$, and thus $d(x',y) \geq d(x,x') - d(x,y) \geq 5d(x,y)$. We also have that $d(y',y) \geq d(x',y) - d(y',x') \geq 5d(x,y) - 6\Delta_{i-1} \geq 30\Delta_i - 6\Delta_{i-1} > 4\Delta_i$. Similarly, $d(x,y') \geq d(x,x') - d(x',y') \geq 6d(x,y) - 6\Delta_{i-1} > 4\Delta_i$. $\square$

We now show a claim about event $\mathcal{Z}_t(x,y)$ which is essential for using the local Lemma.

Claim 26. *For all $(x,y) \in N_i$,*

$$\Pr\left[\neg\mathcal{Z}_t(x,y) \mid \bigwedge_{(x',y') \in Q} \mathcal{Z}(x',y')\right] \leq 7/8$$

for all $Q \subseteq \{(x',y') \in N_{i'} : (\mathcal{Z}(x,y), \mathcal{Z}(x',y')) \notin A \wedge c(\mathcal{Z}(x,y)) \geq c(\mathcal{Z}(x',y'))\}$.

Proof. Fix $(x,y) \in N_i$, $t \in [D]$. We will show that event $\mathcal{Z}(x,y)$ can hold with probability at least $1/8$, and that this will depend only on the padding event for x or y, and on the choice of σ_i for $P_i(x)$, $P_i(y)$, so if we show that any outcome for events in $\mathcal{Z}_(x',y')$ for $x', y' \in Q$ cannot affect the padding and σ, we will be done. For $i' < i$ and any $x', y' \in N_{i'}$ the event $\mathcal{Z}(x',y')$ depend only on the random choices for the first i' levels of the partition, so the padding in scale i and choice of σ_i will be independent of these events. Otherwise $i' = i$. For any $x', y' \in N_i$ such that $(\mathcal{Z}(x,y), \mathcal{Z}(x',y')) \notin A$, by Claim 25 $d(\{x,y\}, \{x',y'\}) > 4\Delta_i$. This suggests that x, y and x', y' fall into different clusters in scale i, hence the choice of σ_i is independent for each. By the locality of our partition, the padding event for $x \in X$ in scale Δ_i is essentially determined by the partition restricted to $B(x, 2\Delta_i)$, so we have that the padding probability in scale i for x, y remains bounded even given any outcome of the padding for x', y'.

Even though the event $\mathcal{Z}(x,y)$ does depend on scales $j < i$, we will show that there is probability at least $1/8$ for it to hold given any outcome for scales $j < i$. Since $(2 + 1/16)2\Delta_i < 6\Delta_i \leq d(x,y) \leq 6\Delta_{i-1} \leq (14 - 1/16)2\Delta_i$, we get from Claim 4 that $\max\{\bar{\rho}(x, 2\Delta_i, \gamma), \bar{\rho}(y, 2\Delta_i, \gamma)\} \geq 2$. We will show that $|B(\upsilon(P_i(y)), 3\gamma\Delta_i)| \leq k^{\chi^4}$. As $d(x, \upsilon(P_i(y))) \leq d(x,y) + \Delta_i \leq 2d(x,y)$, due the weak growth bound assumption,

$$|B(\upsilon(P_i(y)), 3\gamma\Delta_i)| \leq |B(x, 48\Delta_i + 2d(x,y))| \leq |B(x, 9d(x,y))| \leq |B(x, 9r_k(x))| \leq |B^\circ(x, r_k(x))|^{\chi^4} \leq k^{\chi^4}.$$

The same argument holds for $|B(\upsilon(P_i(x)), 3\gamma\Delta_i)|$. This suggests that $\bar{\xi}_i(y) = \xi_i(y)$ and $\bar{\xi}_i(x) = \xi_i(x)$. W.l.o.g assume that $\bar{\rho}(x, 2\Delta_i, \gamma) \geq 2$, which implies that $\xi_i(x) = 1$, hence $\phi_i^{(t)}(x) = 1/\eta_i(x)$. If it is the case that

$$\left| \sum_{j<i} f_j^{(t)}(x) - f_j^{(t)}(y) \right| \leq \frac{\Delta_i}{2},$$

then it is enough that the following will hold

- $B(x, \eta_i(x)\Delta_i) \subseteq P_i(x)$,
- $\sigma_i^{(t)}(P_i(x)) = 1$.
- $\sigma_i^{(t)}(P_i(y)) = 0$.

By Lemma 5, the definition of σ and the fact that $P_i(x) \neq P_i(y)$, the probability of each of these events is independently at least $1/2$. If all these events occur then $|f_i^{(t)}(x) - f_i^{(t)}(y)| \geq \min\{\eta_i^{-1}(x) \cdot d(x, X \setminus P_i(x)), \Delta_i\} \geq \Delta_i$.

If on the other hand

$$\left| \sum_{j<i} f_j^{(t)}(x) - f_j^{(t)}(y) \right| > \frac{\Delta_i}{2},$$

then we just need

- $\sigma_i^{(t)}(P_i(x)) = \sigma_i^{(t)}(P_i(y)) = 0$,

which again holds with probability $1/4$. In any case with probability at least $1/8$ event $\mathcal{Z}_t(x, y)$ will hold. $\square$

Notice that events $\mathcal{Z}_t(x, y)$ are independent of events $\mathcal{Z}_t(x, y)$ for $t \neq t'$. For any $x, y \in N_i$ let $\mathcal{B} = \sum_{t \in D} \mathcal{Z}_t(x, y)$, then $\mathbb{E}[\mathcal{B}] \geq D/8$, and by Chernoff bound we get that

$$\Pr\left[\mathcal{B} < D/16\right] \leq \Pr\left[\mathcal{B} < \mathbb{E}[\mathcal{B}]/2\right] \leq e^{-\mathbb{E}[\mathcal{B}]/8} = e^{-D/64} \leq k^{-\chi^5},$$

where we choose $c = 64\chi^5$.

Let $p = k^{-\chi^5}$. To conclude, using Claim 26

$$\Pr\left[\neg\mathcal{Z}(x, y) \mid \bigwedge_{(x', y') \in Q} \mathcal{Z}(x', y') \right] \leq p$$

for all $Q \subseteq \{(x', y') \in N_{i'} \mid (\mathcal{Z}(x, y), \mathcal{Z}(x', y')) \notin A \wedge c(\mathcal{Z}(x, y)) \geq c(\mathcal{Z}(x', y'))\}$. Note that $ep(k^{\chi^4} + 1) \leq 1$, hence by Lemma 23

$$\Pr[\mathcal{Z}] > 0$$

$\square$

7 Local Dimension Reduction for Ultrametrics

Even though local dimension reduction is impossible in general, we show that it is possible for the class of ultrametrics. The local dimension reduction can be done in any ℓ_p space, in contrast to the Johnson-Lindenstrauss [JL84] (non-local) dimension reduction that can be done in ℓ_2, is impossible in ℓ_1, ℓ_∞ and unknown for other ℓ_p spaces. The proof has three main steps: first embed the ultrametric to a $\theta - e$HST (exact Hierarchically Separated Tree, see definition below) with distortion θ, then embed the HST with k-local distortion 1 into a bounded degree HST, where the bound is polynomial in k. Finally we extend the general framework of [BM04, BLMN04], who showed dimension reduction for ultrametrics, and give lower dimension for bounded degree HSTs. We show that such HSTs can be embedded, preserving all distances up to $1+\epsilon$, into a dimension that is logarithmic in the degree of the HST. Recall,

- **Ultrametric:** An ultrametric (X, d) is a metric space satisfying a strong form of the triangle inequality, for all $x, y, z \in X$, $d(x, z) \le \max\{d(x, y), d(y, z)\}$.
- **HST:** For $\theta \ge 1$, a $\theta - e\mathrm{HST}$ is a finite metric space defined on the branches of a rooted infinite tree, having a finite number of branches. For branches x, y denote by $\mathrm{lca}(x, y)$ the least common ancestor of x and y in the tree, *i.e.* , the deepest node in $x \cap y$, and by $\mathrm{dlca}(x, y)$ its depth. The distance between branches is defined as $d(x, y) = \theta^{-\mathrm{dlca}(x,y)}$. Denote by x_i the i-th node in the branch x.

Theorem 10. *Let (X, d) be an ultrametric, then for any $p \ge 1$, $0 < \epsilon \le 1$ and $k \le |X|$ there is an embedding of X into ℓ_p with k-local distortion $1+\epsilon$ and dimension $O(\epsilon^{-3} \log k)$.* [7]

To prove this theorem we first introduce the following lemmata. The first lemma is a variant of a lemma of [Bar96] (a proof is given in [BLMN05]):

Lemma 27. *For any $\theta > 1$, any ultrametric embeds in a $\theta - e\mathrm{HST}$ with distortion θ.*

Lemma 28. *For all $\theta > 1$, any $\theta - e\mathrm{HST}$ T' can be embedded into a $\theta - e\mathrm{HST}$ T, where every internal node in the tree representation of T has degree at most $2k^2 + 1$, with k-local distortion 1.*

Proof. For a $\theta - e\mathrm{HST}$ T' let $r(T')$ denote the root of T' and for a node $u \in T'$ let $c(u)$ the set of all children of u. The intuition behind the construction of T from T' is by defining a neighborhood graph on the children of the root, and unite those children which are not connected by an edge in this graph, thus obtaining a small number of children. Once we have few children continue recursively on each of them. Formally, perform the following recursive process on T' creating T:

1. Let $r = r(T')$. Define a neighborhood graph with vertices $c(r) = \{v_1, \ldots, v_\ell\}$ by adding a directed edge (v_i, v_j) iff one of the branches x in the subtree rooted at v_i has $y \in N_k(x)$ where y is a branch in the subtree rooted in v_j. It can be seen that only children with at most k branches have out-going edges, hence the out-degree of this graph is bounded by k^2.
2. Using Lemma 12 properly color the graph with $m = 2k^2 + 1$ colors. For any $1 \le i \le m$ let $v_{i_1}, \ldots, v_{i_s}$ be the children colored by color i, replace them by a single node r_i and set $c(r_i) = \bigcup_{j=1}^{s} c(v_{i_j})$.
3. For each $1 \le i \le m$ continue recursively on the subtree rooted at r_i.

Let x, y be two branches such that $y \in N_k(x)$. Then for any level i of the recursive process, let v be the current root - if it is the case that $v \notin x$ or $v \notin y$ then the unions done to the children of v cannot affect $d(x, y)$. Otherwise, let v_j, v_ℓ be the children of v which lie on branches x, y respectively (it could be that $j = \ell$). Since the graph we define on the children of v contains the (directed) edge (v_j, v_ℓ), the vertices v_j, v_ℓ will be colored by distinct colors and will not be united. It follows that the distance between x, y will never change in the process.

Note that the construction of the tree in the deeper recursion levels is done with respect to the *original* set $N_k(x)$, which guarantees that distances between k-nearest neighbors are preserved. □

Lemma 29. *Let $0 < \epsilon \le 1/2$, $m \in \mathbb{N}$ and $\theta = e^\epsilon$. Let T be a $\theta - e\mathrm{HST}$ with branches H, such that the out-degree of every node in T is at most m. Then T can be embedded into ℓ_p with distortion $1+\epsilon$ and dimension $\bar{D} = O((\log m) \cdot (\ln(2 \min\{p, 1/\epsilon\})/\epsilon)^2 \cdot \max\{1, 1/(\epsilon p)\})$.*

[7] The actual bound on the dimension is in fact $O((\log k) \cdot (\ln(2 \min\{p, 1/\epsilon\})/\epsilon)^2 \cdot \max\{1, 1/(\epsilon p)\})$, which is $O(\epsilon^{-3} \log k \cdot p^{-1} \log p)$ for $p = O(1/\epsilon)$, and $O(\epsilon^{-2} \log(1/\epsilon) \log k)$ for $p = \Omega(1/\epsilon)$.

Proof. Let $\theta = e^{\epsilon}$, $\alpha = \ln(2\min\{p, 1/\epsilon\})$, $d = \alpha/\epsilon$ (note that $\theta^d = \min\{2p, 2/\epsilon\}$). Let q be a parameter to be determined later, let $D = cq \cdot \max\{1, 1/(\epsilon p)\}$ for some constant c to be determined later, and set $\bar{D} = D \cdot d$. Let $(e_i)_{i\in\{0,\dots,d-1\}}$ be the standard orthonormal basis of $\mathbb{R}^d$, and $(e_i)_{i\in\mathbb{N}}$ its extension to a periodic sequence modulo d. For each node $a \in T$ let $b_a \in \{0,1\}$ be a random symmetric i.i.d bit. Define for all $t \in [D]$, $i > 0$ $f_i^{(t)} : H \to \mathbb{R}^d$ as

$$f_i^{(t)}(x) = \theta^{-i} b_{x_i} e_i ,$$

and define $f^{(t)} : H \to \mathbb{R}^d$ as $f^{(t)}(x) = \sum_{i=1}^{\infty} f_i^{(t)}(x)$. Finally define $f : H \to \mathbb{R}^{\bar{D}}$ by

$$f(x) = \bigoplus_{t=1}^{D} f^{(t)}(x) .$$

Fix some $x, y \in H$ and $t \in [D]$. For any $j \in \{0,\dots,d-1\}$ let $Z_j^{(t)} = |(f^{(t)}(x) - f^{(t)}(y))_j|^p$. Let $i_j = \min\{i > \mathrm{dlca}(x,y) \mid i = j \mod (d)\}$ and let $I_j = \{i \geq i_j \mid i = j \mod d\}$. Note that since $x_i = y_i$ for any $i \leq \mathrm{dlca}(x,y)$, we have that $Z_j^{(t)} = |\sum_{i\in I_j}(f_i^{(t)}(x) - f_i^{(t)}(y))_i|^p$. Then we have the following

$$0 \leq Z_j^{(t)} \leq \left(\sum_{i\in I_j} \theta^{-i}\right)^p = \left(\theta^{-i_j}\sum_{i=0}^{\infty}\theta^{-id}\right)^p = \left(\frac{\theta^{-i_j}}{1-\theta^{-d}}\right)^p \tag{17}$$

Claim 30. *For any $j \in \{0,\dots,d-1\}$ and $t \in [D]$, $\Pr[Z_j^{(t)} \geq \theta^{-i_j p}] \geq 1/16$. Moreover, this bound depends only on random variables b_{x_i}, b_{y_i} where $i \in \{i_j, i_j + d\}$.*

Proof. There is probability of $1/16$ that the random bits $b_{x_{i_j}} = 1$, $b_{y_{i_j}} = 0$, $b_{x_{i_j+d}} = 1$ and $b_{y_{i_j+d}} = 0$. In such a case $(f_{i_j}^{(t)}(x) - f_{i_j}^{(t)}(y))_{i_j} = \theta^{-i_j}$ and $(f_{i_j+d}^{(t)}(x) - f_{i_j+d}^{(t)}(y))_{i_j+d} = \theta^{-i_j-d}$. Note that $\frac{\theta^{-d}}{1-\theta^{-d}} = \frac{1}{e^{\alpha}-1} \leq 1$ (because $\alpha \geq \ln 2$), which suggests that

$$|\sum_{i\in I_j\setminus\{i_j,i_j+d\}}(f_i^{(t)}(x) - f_i^{(t)}(y))_i| \leq \theta^{-i_j-2d}\sum_{i=0}^{\infty}\theta^{-di} = \frac{\theta^{-i_j-2d}}{1-\theta^{-d}} \leq \theta^{-i_j-d}.$$

It follows that with probability at least $1/16$

$$\begin{aligned} Z_j^{(t)} &= \left|\sum_{i\in I_j}(f_i^{(t)}(x) - f_i^{(t)}(y))_i\right|^p \\ &\geq \left|\theta^{-i_j} + \theta^{-i_j-d} - |\sum_{i\in I_j\setminus\{i_j,i_j+d\}}(f_i^{(t)}(x) - f_i^{(t)}(y))_i|\right|^p \\ &\geq \theta^{-i_j p} \end{aligned}$$

□

For any $1 \leq t \leq D$ let $Z^{(t)}(x,y) = Z^{(t)} = \sum_{j=0}^{d-1} Z_j^{(t)} = \|f^{(t)}(x) - f^{(t)}(y)\|_p^p$. Also let $Z(x,y) = Z =$

$\sum_{t=1}^{D} Z^{(t)}(x,y)$ and $\mu = \mathbb{E}[Z]$. Note that $\mu/d(x,y)^p$ is a constant independent of $d(x,y)$, because we can write

$$\begin{aligned}
\mu &= \sum_{t=1}^{D}\sum_{j=0}^{d-1}\mathbb{E}[Z_j^{(t)}] \\
&= \sum_{t=1}^{D}\sum_{j=0}^{d-1}\mathbb{E}\left[\left|\sum_{i\in I_j}(f_i^{(t)}(x)-f_i^{(t)}(y))_i\right|^p\right] \\
&= \sum_{t=1}^{D}\sum_{j=0}^{d-1}\mathbb{E}\left[\left|\sum_{i\in I_j}\theta^{-i}(b_{x_i}-b_{y_i})\right|^p\right] \\
&= \sum_{t=1}^{D}\sum_{j=0}^{d-1}\mathbb{E}\left[\left|\theta^{-\mathrm{dlca}(x,y)-1}\sum_{i=0}^{\infty}\theta^{-id-j}(b_{x_{\mathrm{dlca}(x,y)+id+j}}-b_{y_{\mathrm{dlca}(x,y)+id+j}})\right|^p\right] \\
&= d(x,y)^p\theta^{-p}\cdot\sum_{t=1}^{D}\sum_{j=0}^{d-1}\mathbb{E}\left[\left|\sum_{i=0}^{\infty}\theta^{-id-j}(b_{x_i}-b_{y_i})\right|^p\right]
\end{aligned}$$

The last equality holds as for calculating the expectation over the random bits it does not matter from which level of the tree they are taken - they are all independent with the same distribution. We conclude that we can scale the embedding f by this constant. It follows that it is enough to prove that there exists a choice of the random bits such that $|Z^{1/p}-\mu^{1/p}|<\epsilon\mu^{1/p}$ for all pairs, then the embedding will have distortion $1+O(\epsilon)$.

Now the analysis is different for various values of p, first we prove for the case that $p\le 4/\epsilon$, note that in this case $D\ge cq/(4\epsilon p)$. Let $M_j=\left(\frac{\theta^{-i_j}}{1-\theta^{-d}}\right)^p\le e\cdot\theta^{-i_jp}$, using that $(1-\theta^{-d})^p\ge(1-1/(2p))^p\ge e^{-p/p}=1/e$, and note that (17) suggests that $0\le Z_j^{(t)}\le M_j$. Let $M=\sum_{t=1}^{D}\sum_{j=1}^{d}M_j^2$, and we have the following bound:

$$\begin{aligned}
M &\le e^2D\sum_{j=0}^{d-1}\theta^{-2i_jp} \qquad (18)\\
&= \frac{e^2D\cdot\theta^{-2(\mathrm{dlca}(x,y)+1)p}\cdot(1-\theta^{-2pd})}{1-\theta^{-2p}} \\
&\le \frac{e^2D\cdot d(x,y)^{2p}}{e^{2\epsilon p}-1} \\
&\le \frac{4D\cdot d(x,y)^{2p}}{\epsilon p}\,,
\end{aligned}$$

where we used in the second inequality that $\frac{\theta^{-2p}}{1-\theta^{-2p}}=\frac{1}{e^{2\epsilon p}-1}$ and that $1-\theta^{-2pd}\le 1$, and in the third inequality that $e^{2\epsilon p}-1\ge 2\epsilon p$.

By linearity of expectation and by Claim 30 it follows that

$$\begin{aligned}
\mu &\ge \frac{D}{16}\sum_{j=0}^{d-1}\theta^{-i_jp} \\
&= \frac{D\cdot\theta^{-p(\mathrm{dlca}(x,y)+1)}}{16}\sum_{i=0}^{d-1}\theta^{-ip} \\
&= \frac{D\cdot\theta^{-p}\cdot d(x,y)^p}{16}\cdot\frac{1-\theta^{-dp}}{1-\theta^{-p}} \\
&\ge \frac{D\cdot d(x,y)^p}{2000\epsilon p},
\end{aligned}$$

where in the last inequality we used that $1-\theta^{-dp} \geq 1-e^{-p} \geq 1-1/e \geq 1/2$, that $1-\theta^{-p} = 1-e^{-\epsilon p} \leq \epsilon p$ and that $\theta^{-p} \geq 1/e^4$.

Now we want to show that with high enough probability it will be the case that $|Z-\mu| < \epsilon p \mu$, and since $\epsilon p \leq e^{\epsilon p}-1 \leq (1+2\epsilon)^p - 1$ (using that $e^{\epsilon} \leq 1+2\epsilon$ for $0<\epsilon \leq 1/2$) it will imply that $(1-2\epsilon)\mu^{1/p} < Z^{1/p} < (1+2\epsilon)\mu^{1/p}$ which gives the desired distortion $1+O(\epsilon)$. By Hoeffding's inequality (Lemma 2) with $\eta = \epsilon p \mu \geq \frac{D \cdot d(x,y)^p}{2000}$ and using (18),

$$\Pr[|Z-\mu| \geq \epsilon p \mu] = \Pr[|Z-\mu| \geq \eta] \leq 2e^{-2\eta^2/M} \leq 2e^{-\epsilon p D/8000000} \leq e^{-q-1},$$

for a large enough constant c. Denote by $P = e^{-q-1}$.

Consider the case $p > 4/\epsilon$, where $D = cq$. First observe that $1-\theta^{-d} = 1-\epsilon/2$ and $\theta^{-p} \leq (1-\epsilon+\epsilon^2/2)^p \leq (1-3\epsilon/4)^p$ (the last inequality holds since $\epsilon \leq 1/2$) hence $\frac{\theta^{-p}}{(1-\theta^{-d})^p} \leq \frac{(1-3\epsilon/4)^p}{(1-\epsilon/2)^p} \leq (1-\epsilon/4)^p \leq 1/e$. Also note that $1-\theta^{-p} \geq 1-e^{-4}$. Assume w.l.o.g that i_0 is the minimal among $i_0, i_1, \ldots, i_{d-1}$, then by (17) for any $0 \leq \ell \leq d-1$,

$$\begin{aligned} \sum_{j=\ell}^{d-1} Z_j^{(t)} &\leq \frac{1}{(1-\theta^{-d})^p} \sum_{j=\ell}^{d} \theta^{-i_j p} \\ &= \frac{\theta^{-(i_\ell-1)p} \cdot \theta^{-p}}{(1-\theta^{-d})^p} \cdot \frac{1-\theta^{-dp}}{1-\theta^{-p}} \\ &\leq \theta^{-(i_\ell-1)p} \cdot \frac{1/e}{1-1/e^4} \\ &\leq \theta^{-(i_\ell-1)p}/2, \end{aligned} \tag{19}$$

In particular for $\ell = 0$ we have that for any $1 \leq t \leq D$

$$Z^{(t)} \leq \theta^{-(i_0-1)p}/2 = \theta^{-\mathrm{dlca}(x,y)\cdot p}/2 = d(x,y)^p/2,$$

hence with probability 1

$$Z = \sum_{t=1}^{D} Z^{(t)} \leq D \cdot d(x,y)^p \ .$$

By Claim 30 with probability at least $1/16$ we have that $Z_0^{(t)} \geq \theta^{-i_0 p}$, and (19) suggests that $\sum_{j=1}^{d-1} Z_j^{(t)} \leq \theta^{-i_0 p}/2$. It follows that with probability $1/16$

$$Z^{(t)} \geq \theta^{-i_0 p} - |\sum_{j=1}^{d} Z_j^{(t)}| \geq \theta^{-i_0 p}/2 \geq (1-\epsilon)^p \cdot d(x,y)^p/2 \ . \tag{20}$$

Let K_t be an indicator random variable for the event that $Z^{(t)} \geq \theta^{-i_0 p}/2$, and $K = \sum_{t=1}^{D} K_t$. Note that $\mathbb{E}[K] \geq D/16$, denote $\mathbb{E}[K] = \gamma D/16$ for some $\gamma \geq 1$, then by Chernoff bound

$$\Pr[K < D/32] \leq \Pr[\mathbb{E}[K] - K > \gamma D/32] \leq e^{-\gamma D/8} \leq e^{-D/8} \leq e^{-q-1} = P$$

for c large enough. So with probability at least $1-P$ we have that at least $1/32$ fraction of the $Z^{(t)}$ are lower bounded as in (20), hence

$$Z = \sum_{t=1}^{D} Z^{(t)} \geq D/32 \cdot (1-\epsilon)^p \cdot d(x,y)^p/2 = D(1-\epsilon)^p \cdot d(x,y)^p/64 \ ,$$

as required, since the distortion is $\left(\frac{64}{(1-\epsilon)^p}\right)^{1/p} = 1+O(\epsilon)$.

Finally, in order to use the local lemma, we need to argue that the number of dependencies is small relative to the success probability. Define equivalence relation on unordered pairs of branches such that $\{x,y\} \sim \{x',y'\}$ iff

$x_{a+2d} = x'_{a+2d}$, $y_{a+2d} = y'_{a+2d}$ where $a = \mathrm{dlca}(x, y) + 1$. Denote by $[x, y]$ be the equivalence class of $\sim$ that contains the pair x, y. Claim 30 implies that for all the pairs in $[x, y]$ the success event (in either of the above cases) for each one of the pairs in $[x, y]$ is defined by exactly the same random variables - the first $2d$ nodes on the branches x, y just after $\mathrm{lca}(x, y)$. Moreover, all the pairs require exactly the same events to occur in order to succeed. Let $Y_{[x,y]}$ be an indicator variable for the event that the random choices for nodes in the first $2d$ levels after $\mathrm{lca}(x, y)$ are such that $|Z(x, y) - \mathbb{E}[Z(x, y)]| < (1 + \epsilon)^p \mathbb{E}[Z(x, y)]$. Our "bad" events will be $\neg Y_{[x,y]}$, note that $\Pr[\neg Y_{[x,y]}] \le P$.

Let $Q = Q([x, y])$ be the number of equivalence classes $[x', y']$ such that $\mathrm{lca}(x, y)$ and $\mathrm{lca}(x', y')$ are on the same branch in T and their tree distance is at most $2d$. Note that if one of these two conditions does not hold the random variables governing the success of $Y_{[x,y]}$ are completely different from those governing $Y_{[x',y']}$, hence those events are independent. We set $q = \ln Q$, hence $P = 1/(eQ)$ and we have that $eQP = 1$, and the Lovász Local Lemma implies that there is some probability that all good events $Y_{[x,y]}$ occur simultaneously for all equivalence classes $[x, y]$.

It remains to give a bound on Q in order to show that D is small enough, as promised in the lemma statement. Since the out-degree of T is bounded by m, for each $u \in T$ there are at most m^{2d} descendants of u at tree distance $2d$, and hence at most $(m^{2d})^2$ equivalence classes $[x, y]$ for which the node u is $\mathrm{lca}(x, y)$. In addition there are at most $2d \cdot m^{2d} + 2d \le m^{4d}$ other possible nodes $u' \in T$ at tree distance at most $2d$ from u such that both u and u' are on the same branch of T. It follows that $Q \le m^{8d} \le e^{9 \ln m \cdot \alpha/\epsilon}$, hence $q \le 9 \ln m \cdot \alpha/\epsilon$, which gives the desired bound on D.

□

Proof of Theorem 10. Fix any $p \ge 1$ and $0 < \epsilon \le 1$. Let $\hat{\epsilon} = \epsilon/4$ and $\theta = e^{\hat{\epsilon}}$. Using Lemma 27 embed the ultrametric (X, d) into a $\theta - e$HST T' with distortion θ, and using Lemma 28 embed T' into a $\theta - e$HST T such that the degree of any internal node of T is at most $m = 2k^2 + 1$, with k-local distortion 1. Finally using Lemma 29 with parameter $\hat{\epsilon} < 1/2$, embed T into $\bar{D}$ dimensional ℓ_p space with distortion $1 + \hat{\epsilon}$, where $\bar{D} = O((\log m) \cdot (\ln(2 \min\{p, 1/\epsilon\})/\epsilon)^2 \cdot \max\{1, 1/(\epsilon p)\}) = O((\log k)/\epsilon^3)$. The total k-local distortion is at most $\theta(1 + \hat{\epsilon}) \le 1 + \epsilon$.

□

7.1 Low Dimensional Embedding for Doubling Ultrametrics

As a bi-product of our work, the techniques used to prove Theorem 10 can be used to give near optimal embedding of Ultrametrics in their intrinsic dimension.

Recall that a metric space is called λ-doubling if for any $x \in X$, $r > 0$, the ball $B(x, 2r)$ can be covered by λ balls of radius r.

Theorem 11. *For any $p \ge 1$ and $0 < \epsilon \le 1$, every λ-doubling ultrametric (X, d) on n points embeds into L_p^D with distortion $1 + \epsilon$ where $D = O((\log \lambda)/\epsilon^2)$.* [8]

Proof. The proof is the same as that of Theorem 10 except that the bound on Q, the number of dependencies between equivalence classes, in the end of the argument can be improved, which immediately imply an improvement in D. For a λ-doubling HST T, and a node $u \in T$, we claim that the number of descendants $2d$ levels below u is at most $\lambda^{3\alpha}$. This is because all such descendants correspond to a metric distance which is smaller than that of u by a factor of $\theta^{2d} = e^{2\alpha}$, letting r be the distance of u from the root, the ball of radius θ^{-r} around any branch containing u will contain all the other branches that contain u, however any ball of radius θ^{-r-2d} include at most only the branches that contain v, where v a certain descendant of u at distance $2d$. Thus applying the doubling condition $\log(e^{2\alpha})$ times we get that the number of descendants at distance $2d$ cannot be more than $\lambda^{\log(e^{2\alpha})} \le \lambda^{3\alpha}$.

We conclude that for each u there are at most $(\lambda^{3\alpha})^2$ equivalence classes of pairs whose lca is u. Fixing some $[x, y]$, we need to bound the number of possible nodes u within tree distance $2d$ from $\mathrm{lca}(x, y)$ such that u and $\mathrm{lca}(x, y)$ are on the same branch. Let v be the node $2d$ levels above $\mathrm{lca}(x, y)$, by the same argument as above there will be at most $\lambda^{6\alpha}$ descendants of v at tree distance $4d$. Note that only a node that has out-degree at least 2 can actually

[8] The actual bound on the dimension is in fact $O((\log \lambda) \cdot (\ln(2 \min\{p, 1/\epsilon\}))^2 \cdot 1/\epsilon \cdot \max\{1, 1/(\epsilon p)\})$, which is $O(\epsilon^{-2} \log \lambda \cdot p^{-1} \log^2 p)$ for $p = O(1/\epsilon)$, and $O(\epsilon^{-1} \log^2(1/\epsilon) \log \lambda)$ for $p = \Omega(1/\epsilon)$.

be a lca, so the number of such possible lca nodes u is at most $2\lambda^{6\alpha}$. We conclude that the number of dependencies $Q \le (\lambda^{3\alpha})^2 \cdot 2\lambda^{6\alpha} \le \lambda^{8\alpha}$, and hence $D = c \log \lambda \cdot \ln(2\min\{p, 1/\epsilon\}) \cdot \max\{1, 1/(\epsilon p)\}$, and obtain a dimension $\bar{D} = d \cdot D = c \log \lambda \cdot (\ln(2\min\{p, 1/\epsilon\}))^2 \cdot \max\{1, 1/(\epsilon p)\}/\epsilon$. $\square$

8 Partition Based Embeddings into ℓ_p

8.1 Scaling with Low Dimension

The following theorem is similar to Theorem 3, it obtains better dimension and slightly worse distortion.

Theorem 12. *For any finite metric space (X, d) on n points and any $\vartheta \in \Xi$ there exists an embedding into ℓ_p with strong scaling local distortion $O(\frac{\vartheta(\log k)}{p})$, worse case distortion $O(\frac{\log n}{p})$ and dimension $O(2^p \log n)$.*

Let $s = e^p$. Let $D = c \cdot e^p \ln n$ for a constant c to be determined later. Let $\Delta_0 = \operatorname{diam}(X)$, $I = [\lceil \log \Delta_0 \rceil]$, and for all $i \in I$, $\Delta_i = \Delta_0/2^i$. We will define an embedding $f : X \to \ell_p^D$, by defining for each $1 \le t \le D$, functions $f^{(t)}, \psi^{(t)}, \mu^{(t)} : X \to \mathbb{R}^+$ and letting $f^{(t)} = \psi^{(t)} + \mu^{(t)}$ and $f = D^{-1/p} \bigoplus_{1 \le t \le D} f^{(t)}$.

Fix t, $1 \le t \le D$. In what follows we define $\psi^{(t)}$. We construct for each $i \in I$ a uniformly $(\eta_i \cdot p, 1/s)$ locally padded probabilistic partition $\hat{\mathcal{P}}_i$ as in Lemma 5, and let ξ be as defined in the lemma. Now fix a partition $P_i \in \mathcal{P}_i$ for every $i \in I$, and fix an arbitrary center $v(C)$ for each $C \in P_i$. Define the "scaling down" factor by

$$\bar{\xi}_i(x) = \frac{\xi_i(x)}{\vartheta(\log |B(v(P_i(x)), 33\Delta_i)|)} ,$$

observe that $\bar{\xi}$ is uniform as well. We define the embedding by defining the coordinates for each $x \in X$. Define for $x \in X$, $0 < i \in I$, $\phi_i^{(t)} : X \to \mathbb{R}^+$, by $\phi_i^{(t)}(x) = \bar{\xi}_i(x)/\eta_i(x)$. As in Section 6.2 we have Claim 16.

Claim 31. *For any $1 \le t \le D$ and $x \in X$, $\sum_{i \in I} \phi_i^{(t)}(x) \le O(1)$.*

Proof. Note that $d(x, v(P_i(x))) \le \Delta_i$, then

$$\begin{aligned}
\sum_{i \in I} \phi_i^{(t)}(x) &= \sum_{i \in I | \xi_i(x) = 1} \frac{\eta_i(x)^{-1}}{\vartheta(\log |B(v(P_i(x)), 33\Delta_i)|)} \\
&\le 2^7 \sum_{i \in I} \frac{\rho(x, 2\Delta_i, 16)}{\vartheta(\log |B(x, 32\Delta_i)|)} \\
&\le 2^7 \sum_{i \in I} \frac{\log |B(x, 32\Delta_i)| - \log |B(x, \Delta_i/8)|}{\vartheta(\log |B(x, 32\Delta_i)|)} \\
&\le 2^7 \sum_{i \in I} \int_{\log |B(x, \Delta_i/8)|}^{\log |B(x, 32\Delta_i)|} \frac{dz}{\vartheta(z)} \\
&\le 2^{11} \sum_{j=1}^{\infty} \frac{1}{\vartheta(j)} \\
&= O(1) .
\end{aligned}$$

$\square$

For each $0 < i \in I$ we define a function $\psi_i^{(t)} : X \to \mathbb{R}^+$ and for $x \in X$, let $\psi^{(t)}(x) = \sum_{i \in I} \psi_i^{(t)}(x)$.

Let $\{\sigma_i^{(t)}(C) | C \in P_i, 0 < i \in I\}$ be i.i.d symmetric $\{0,1\}$-valued Bernoulli random variables. The embedding is defined as follows: for each $x \in X$:

- For each $0 < i \in I$, let $\psi_i^{(t)}(x) = \sigma_i^{(t)}(P_i(x)) \cdot \phi_i^{(t)}(x) \cdot d(x, X \setminus P_i(x))$.

The proof of the following claim is essentially the same as the proof of Claim 18.

Claim 32. *For any $x, y \in X$ and $0 < i$: $f_i^{(t)}(x) - f_i^{(t)}(y) \le \phi_i^{(t)}(x) \cdot d(x,y)$.*

Next, we define the function $\mu^{(t)}$, based on the embedding technique of Bourgain [Bou85] and its generalization by Matoušek [Mat97]. Let $T' = \lceil \log_s n \rceil$ and $K = \{k \in \mathbb{N} | 1 \le k \le T'\}$. For each $k \in K$ define a randomly chosen subset $A_k^{(t)} \subseteq X$, with each point of X included in $A_k^{(t)}$ independently with probability s^{-k}. Define $\mu_k^{(t)} : X \to \mathbb{R}^+$ and for $x \in X$ and $k \in K$ let $\mu_k^{(t)} = \frac{d(x, A_k^{(t)})}{\vartheta(k)}$ and define $\mu^{(t)}(x) = \sum_{k \in K} \mu_k^{(t)}(x)$.

Lemma 33. *For any $1 \le t \le D$ and $x, y \in X$:*

$$|f^{(t)}(x) - f^{(t)}(y)| \le O(d(x,y)) \; .$$

Proof. From Claim 18 and using Claim 31 we get

$$\begin{aligned} \psi^{(t)}(x) - \psi^{(t)}(y) &= \sum_{0<i\in I} (\psi_i^{(t)}(x) - \psi_i^{(t)}(y)) \\ &\le \sum_{0<i\in I} \phi_i^{(t)}(x) \cdot d(x,y) \\ &\le O(d(x,y)) \; . \end{aligned}$$

Also

$$\begin{aligned} \mu^{(t)}(x) - \mu^{(t)}(y) &= \sum_{0<k\in K} \frac{\mu_k^{(t)}(x) - \mu_k^{(t)}(y)}{\vartheta(k)} \\ &= \sum_{0<k\in K} \frac{d(x, A_k^{(t)}) - d(y, A_k^{(t)})}{\vartheta(k)} \\ &\le d(x,y) \sum_{0<k\in K} \frac{1}{\vartheta(k)} \\ &\le d(x,y) \; . \end{aligned}$$

It follows that

$$|f^{(t)}(x) - f^{(t)}(y)| = |\psi^{(t)}(x) + \mu^{(t)}(x) - \psi^{(t)}(y) - \mu^{(t)}(y)| \le O(d(x,y)) \; .$$

□

Lemma 34. *There exists a universal constant $C > 0$ such that for any $x, y \in X$ and any $k \in [n]$, with probability at least $e^{-4p}/4$:*

$$|f^{(t)}(x) - f^{(t)}(y)| \ge C \cdot \min\{d(x,y), r_k(x)\}/\vartheta(\log k) \; .$$

Proof. Let $L = \min\{d(x,y), r_k(x)\}$, and let $0 < i \in I$ be such that $128\Delta_i < L \le 256\Delta_i$. We distinguish between the following two cases:

- **Case 1:** Either $\xi_i(x) = 1$ or $\xi_i(y) = 1$.
  Assume w.l.o.g that $\xi_i(x) = 1$. Since $B(\upsilon(P_i(x)), 33\Delta_i) \subseteq B(x, 34\Delta_i) \subseteq B(x, L)$, it follows that

$$\phi_i^{(t)}(x) = \frac{\eta_i(x)^{-1}}{\vartheta(\log |B(\upsilon(P_i(x)), 33\Delta_i)|)} \ge \frac{\eta_i(x)^{-1}}{\vartheta(\log k)} \; .$$

As $\hat{\mathcal{P}}_i$ is $(\eta \cdot p, 1/s)$-padded we have the following bound

$$\Pr[B(x, \eta_i(x)p \cdot \Delta_i) \subseteq P_i(x)] \geq 1/s.$$

Therefore with probability at least $1/s$:

$$\phi_i^{(t)}(x) \cdot d(x, X \setminus P_i(x)) \geq \frac{\eta_i(x)^{-1}}{\vartheta(\log k)} \cdot \eta_i(x) p \cdot \Delta_i \geq \frac{p \cdot \Delta_i}{\vartheta(\log k)} \, .$$

Assume that this event occurs. We distinguish between two cases:

- $|f^{(t)}(x) - f^{(t)}(y) - (\psi_i^{(t)}(x) - \psi_i^{(t)}(y))| \geq \frac{p \cdot \Delta_i}{2\vartheta(\log k)}$. In this case there is probability at least $1/4$ that $\sigma_i^{(t)}(P_i(x)) = \sigma_i^{(t)}(P_i(y)) = 0$, so that $\psi_i^{(t)}(x) = \psi_i^{(t)}(y) = 0$.
- $|f^{(t)}(x) - f^{(t)}(y) - (\psi_i^{(t)}(x) - \psi_i^{(t)}(y))| < \frac{p \cdot \Delta_i}{2\vartheta(\log k)}$.
  Since $\mathrm{diam}(P_i(x)) \leq \Delta_i < d(x, y)$ we have that $P_i(y) \neq P_i(x)$. We get that there is probability $1/4$ that $\sigma_i^{(t)}(P_i(x)) = 1$ and $\sigma_i^{(t)}(P_i(y)) = 0$ so that $\psi_i^{(t)}(x) - \psi_i^{(t)}(y) \geq \frac{p \cdot \Delta_i}{\vartheta(\log k)}$.

We conclude that with probability at least $1/4s$: $|f^{(t)}(x) - f^{(t)}(y)| \geq \frac{p \cdot \Delta_i}{2\vartheta(\log k)}$.

- **Case 2:** $\xi_i(x) = \xi_i(y) = 0$

It follows from Lemma 5 that $\max\{\bar{\rho}(x, 2\Delta_i, \gamma), \bar{\rho}(y, 2\Delta_i, \gamma)\} < s$. Let $x' \in B(x, 2\Delta_i)$ and $y' \in B(y, 2\Delta_i)$ such that $\rho(x', 2\Delta_i, \gamma) = \bar{\rho}(x, \Delta_i, \gamma)$ and $\rho(y', \Delta_i, \gamma) = \bar{\rho}(y, \Delta_i, \gamma)$. As $\gamma = 16$ and $d(x, y) > 128\Delta_i$ it follows that $B(x', 32\Delta_i)$ and $B(y', 32\Delta_i)$ are disjoint. W.l.o.g assume $|B(x', 32\Delta_i)| \geq |B(y', 32\Delta_i)|$. Let j be such that $s^{j-1} < |B(y', 32\Delta_i)| \leq s^j$, and note that by definition of $\bar{\rho}$, both $|B(x', \Delta_i/8)|, |B(y', \Delta_i/8)| \geq s^{j-2}$. Define the following events: $\mathcal{F}_x = \{B(x', \Delta_i/8) \cap A_j^{(t)} \neq \emptyset\}$, $\mathcal{F}_y = \{B(y', \Delta_i/8) \cap A_j^{(t)} \neq \emptyset\}$, and $\mathcal{E} = \{B(y', 32\Delta_i) \cap A_j^{(t)} = \emptyset\}$. As $d(x, y) > 64\Delta_i$, we have that $\mathcal{F}_x$, $\mathcal{E}$ are independent, and also $\mathcal{F}_x$ and $\mathcal{F}_y$ are independent. Now for $l \in \{x, y\}$,

$$\Pr[\mathcal{F}_l] \geq 1 - (1 - s^{-j})^{s^{j-2}} \geq 1 - e^{s^{-j} \cdot s^{j-2}} \geq 1/(2s^2) \, ,$$
$$\Pr[\mathcal{E}] \geq (1 - s^{-j})^{s^j} \geq 1/4 \, ,$$

using that $s \geq 2$. We distinguish between two cases

- $|f^{(t)}(y) - f^{(t)}(x) - (\mu_j^{(t)}(y) - \mu_j^{(t)}(x))| < 15\Delta_i/\vartheta(j)$. In this case there is probability at least $1/(8s^2)$ that both $\mathcal{E}$ and $\mathcal{F}_x$ occurred, and then $d(y, A_j^{(t)}) \geq d(y', A_j^{(t)}) - d(y, y') \geq 30\Delta_i$, and also $d(x, A_j^{(t)}) \leq d(x, x') + d(x', A_j^{(t)}) \leq 3\Delta_i$. We conclude that

$$\mu_j^{(t)} y) - \mu_j^{(t)}(x) \geq \frac{d(y, A_j^{(t)}) - d(x, A_j^{(t)})}{\vartheta(j)} \geq 27\Delta_i/\vartheta(j) \, .$$

- If it is the case that $|f^{(t)}(y) - f^{(t)}(x) - (\mu_j^{(t)}(y) - \mu_j^{(t)}(x))| \geq 15\Delta_i/\vartheta(j)$, then there is probability at least $1/(4s^4)$ that both $\mathcal{F}_x$ and $\mathcal{F}_y$ occurred, and then both $d(x, A_j^{(t)}), d(y, A_j^{(t)}) \leq 3\Delta_i$, so

$$\mu_j^{(t)}(y) - \mu_j^{(t)}(x) \leq 3\Delta_i/\vartheta(j).$$

In either case, $|f^{(t)}(y) - f^{(t)}(x)| \geq 12\Delta_i/\vartheta(j)$ happens with probability at least $1/(8s^4)$.
Observe that $s^{j-1} \leq |B(y', 32\Delta_i)| \leq |B(x, L)| \leq k$, so that $j \leq (2\log k)/p$.

It follows that with probability at least $1/(8s^4)$:

$$|f^{(t)}(x) - f^{(t)}(y)| \geq \frac{p \cdot \Delta_i}{4\vartheta(\log k} = \Omega(p \cdot d(x,y)/\vartheta(\log k)) \ .$$

□

Proof of Theorem 12. By definition

$$\|f(x) - f(y)\|_p^p = D^{-1} \sum_{1 \leq t \leq D} |f^{(t)}(x) - f^{(t)}(y)|^p.$$

Lemma 33 implies that

$$\|f(x) - f(y)\|_p \leq O(d(x,y)) \ .$$

Using Lemma 34 and applying Chernoff bounds similarly to Section 4.2 we get w.h.p for any $k \in [n]$ and $x, y \in X$, with $L = \min\{d(x,y), r_k(x)\}$:

$$\|f(x) - f(y)\|_p^p \geq \frac{1}{2^{11}} e^{-4p} \left(p \cdot L/\vartheta(\log k)\right)^p \geq \left(\Omega(p \cdot d(x,y)/\vartheta(\log k)\right)^p \ .$$

□

8.2 Decomposable Metrics

For decomposable metrics (recall Definition 10), we improve the scaling local distortion. Using a partition based embedding [KLMN04, ABN06] we get the following,

Theorem 13. *For any metric space (X,d) on n points admitting a τ-padded decomposition, for any $p \geq 1$ and $\vartheta \in \Xi$ there exists an embedding into ℓ_p with scaling local distortion $O\left(\tau^{-1+1/p}\vartheta(\log k)^{1/p}\right)$, worse case distortion $O(\tau^{-1+1/p}(\log n)^{1/p})$ and dimension $O(\log^2 n)$.*

Let $D = c \ln n$ for a constant c to be determined later. Let $D' = \lceil 32 \ln n \rceil$. We will define an embedding $f : X \to \ell_p^{D'D}$, by defining for each $t \in [D]$, an embedding $f^{(t)} : X \to \ell_p^{D'}$ and let $f = D^{-1/p} \bigoplus_{t \in [D]} f^{(t)}$. Let $\Delta_0 = \mathrm{diam}(X)$, $I = \{1 \leq i \leq \lceil \log \Delta_0 \rceil \ : \ i \in \mathbb{N}\}$, and for all $i \in I$ let $\Delta_i = \Delta_{i-1}/2$. Fix some $t \in [D]$, and in what follows we define $f^{(t)}$. For each $i \in I$, we shall create a metric (X, d_i) by contracting short pairs. More formally, consider the complete graph with vertex set X and edge weights d. Replace every weight smaller than Δ_i/n^2 by 0, and d_i is the shortest path metric on this graph. Obviously $d_i \leq d$, but observe that the maximal change in any distance is at most Δ_i/n. In what follows, balls are taken with respect to d, the original metric. Let $\hat{\mathcal{P}}_i$ be a uniformly $(\eta_i, 1/2)$ locally padded probabilistic partition $\hat{\mathcal{P}}_i$ of (X, d_i) as in Lemma 5, and let ξ be as defined in the lemma. Now fix a partition $P_i \in \mathcal{P}_i$ for every $i \in I$, and fix an arbitrary center $v(C)$ for each $C \in P_i$. Define the "scaling down" factor by

$$\bar{\xi}_i(x) = \frac{\xi_i(x)}{\vartheta(\log |B(v(P_i(x)), 33\Delta_i)|)} \ ,$$

observe that $\bar{\xi}$ is uniform as well. Define for $x \in X$, $0 < i \in I$, $\phi_i^{(t)} : X \to \mathbb{R}^+$, by $\phi_i^{(t)}(x) = \left(\bar{\xi}_i(x)/\eta_i(x)\right)^{1/p}$. As in Section 6.2 we have Claim 16.

Claim 35. *For any $1 \leq t \leq D$ and $x \in X$, $\sum_{i \in I} \phi_i^{(t)}(x)^p \leq O(1)$.*

Proof. Note that $d(x, \upsilon(P_i(x))) \le \Delta_i$, then

$$\begin{aligned}
\sum_{i\in I} \phi_i^{(t)}(x)^p &= \sum_{i\in I|\xi_i(x)=1} \frac{\eta_i(x)^{-1}}{\vartheta(\log|B(\upsilon(P_i(x)), 33\Delta_i)|)} \\
&\le 2^7 \sum_{i\in I} \frac{\rho(x, 2\Delta_i, 16)}{\vartheta(\log|B(x, 32\Delta_i)|)} \\
&\le 2^7 \sum_{i\in I} \frac{\log|B(x,32\Delta_i)| - \log|B(x,\Delta_i/8)|}{\vartheta(\log|B(x,32\Delta_i)|)} \\
&\le 2^7 \sum_{i\in I} \int_{\log|B(x,\Delta_i/8)|}^{\log|B(x,32\Delta_i)|} \frac{dz}{\vartheta(z)} \\
&\le 2^{11} \sum_{j=1}^{\infty} \frac{1}{\vartheta(j)} \\
&= O(1)\ .
\end{aligned}$$

□

For $i \in I$ let $f_i^{(t)} : X \to \mathbb{R}^+$. Let $\{\sigma_i^{(t)}(C) | C \in P_i, 0 < i \in I\}$ be i.i.d symmetric $\{0,1\}$-valued Bernoulli random variables. Let $(e_1, \ldots, e_{D'})$ be the standard orthonormal basis of $\mathbb{R}^{D'}$, and extend it to a periodic sequence $(e_1, \ldots, e_I)$ (that is, for $i \in I$, $e_i = e_{i(\mod D')}$). The embedding is defined as follows: for each $x \in X$,

- $f_i^{(t)}(x) = \sigma_i^{(t)}(P_i(x)) \cdot \phi_i^{(t)}(x) \cdot d_i(x, X \setminus P_i(x))$,
- $f^{(t)}(x) = \sum_{i\in I} f_i^{(t)}(x) \otimes e_i$.

The proof of the following claim is essentially the same as the proof of Claim 18, noting that $d_i(x,y) \le d(x,y)$.

Claim 36. *For any $x, y \in X$ and $0 < i$: $f_i^{(t)}(x) - f_i^{(t)}(y) \le \phi_i^{(t)}(x) \cdot d(x,y)$.*

Lemma 37. *For any $1 \le t \le D$ and $x, y \in X$:*

$$\|f^{(t)}(x) - f^{(t)}(y)\|_p \le O(d(x,y))\ .$$

Proof. Fix a coordinate $l \in [D']$, and let $I_l = \{i \in I \ : \ i = l(\mod D')\}$. If $i(l) = i \in I_l$ is the minimal such that x, y are not contracted in $X(i)$, then $d(x,y) \ge \Delta_i/n^2$. Observe that for all other $i' \in I_l$ with $i' < i$, $d_{i'}(x,y) = 0$, so $f_i^{(t)}(x) = f_i^{(t)}(y)$. If $i' \in I_l$ satisfies $i' = i + kD'$ for some positive integer k, then by definition of I_l, $i' \ge i + kD'$, so $\Delta_{i'} \le \Delta_i/2^{kD'} \le d(x,y)/n^{2k}$, thus the contribution of all these terms to the coordinate l is negligible - at most $d(x,y)/n$.

From Claim 36 and using Claim 35 we get

$$\begin{aligned}
\|f^{(t)}(x) - f^{(t)}(y)\|_p^p &= \sum_{l\in[D']} \left(\sum_{i\in I_l} f_i^{(t)}(x) - f_i^{(t)}(y) \right)^p \\
&\le \sum_{l\in[D']} \left(d(x,y)/n + |f_{i(l)}^{(t)}(x) - f_{i(l)}^{(t)}(y)| \right)^p \\
&\le 2 \sum_{0<i\in I} (\phi_i^{(t)}(x)^p + \phi_i^{(t)}(y)^p) \cdot d_i(x,y)^p \\
&\le O(d(x,y)^p)\ .
\end{aligned}$$

□

Lemma 38. *There exists a universal constant $C > 0$ such that for any $x, y \in X$ and $k \in \mathbb{N}$ such that $y \in N_k(x)$, with probability at least $1/8$:*

$$\|f^{(t)}(x) - f^{(t)}(y)\|_p \geq C \cdot d(x,y) \cdot \tau^{1-1/p}/\vartheta(\log k)^{1/p} .$$

Proof. Let $0 < i \in I$ be such that $128\Delta_i < d(x,y) \leq 256\Delta_i$, observe that $d_i(x,y)$ can differ from $d(x,y)$ by at most Δ_i/n, so we ignore this slight change. By Claim 4 it follows that one of $\max\{\xi_i(x), \xi_i(y)\} = 1$, so assume w.l.o.g that $\xi_i(x) = 1$. By Lemma 5, with probability at least $1/2$, x is padded, thus $d(x, X \setminus P_i(x)) \geq \eta_i(x) \cdot \Delta_i$. Also since P_i is Δ_i-bounded we get that $P_i(x) \neq P_i(y)$, and with probability $1/4$, $\sigma_i(P_i(x)) = 1$ and $\sigma_i(P_i(y)) = 0$. Also note that $|B(\upsilon(P_i(x)), 33\Delta_i)| \leq |B(x, 34\delta_i)| \leq |B(x, r_k(x)/2)| \leq k$. If all these things occur, since $B(\upsilon(P_i(x)), 33\Delta_i) \subseteq B(x, 34\Delta_i) \subseteq B(x, r_k(x)/2)$, it follows that

$$|f_i^{(t)}(x) - f_i^{(t)}(y)| = f_i^{(t)}(x) \geq \frac{\eta_i(x)^{-1/p}}{\vartheta(\log|B(\upsilon(P_i(x)), 33\Delta_i)|)^{1/p}} \cdot \eta_i(x) \cdot \Delta_i \geq \frac{d(x,y) \cdot \eta_i(x)^{1-1/p}}{\vartheta(\log k)^{1/p}} .$$

By the first property of Lemma 5, $\eta_i(x) \geq \tau/2$. There are several terms in the coordinate $l \in [D']$ such that $i \in I_l$, however we will show that these terms are zero or negligible. If $i' \in I_l$ is such that $i' < i$, then $\Delta_{i'} \geq \Delta_{i-D'} \geq n^4\Delta_i \geq n^4 d(x,y)/256 > n^2 d(x,y)$ (assuming $n > 16$), so the pair x, y will be contracted in scale i' and thus necessarily $f_{i'}^{(t)}(x) = f_{i'}^{(t)}(y)$. If it is the case that $i' > i$, then $\Delta_{i'} \leq \Delta_{i-D'} \leq \Delta_i/n^4$, so $f_{i'}(t)(x) \leq d(x, X \setminus P_{i'}(x))/\eta_{i'}(x) \leq \Delta_{i'}/(1/\log n) < \Delta_i/n^3 < d(x,y)/n^2$, and the same holds for $f_{i'}^{(t)}(y)$, so that all these terms are negligible with respect to $\frac{d(x,y)\cdot\tau^{1-1/p}}{2\vartheta(\log k)}$.

We conclude that with probability at least $1/8$,

$$\begin{aligned}\|f^{(t)}(x) - f^{(t)}(y)\|_p^p &\geq \left|\sum_{i \in I_l} f_i^{(t)}(x) - f_i^{(t)}(y)\right|^p \\ &\geq \left(\frac{d(x,y) \cdot \tau^{1-1/p}}{4\vartheta(\log k)^{1/p}}\right)^p .\end{aligned}$$

□

Proof of Theorem 13. Let $x, y \in X$, and $k \in \mathbb{N}$ such that $y \in N_k(x)$. Then by Lemma 37

$$\|f(x) - f(y)\|_p^p = \frac{1}{D} \bigoplus_{t \in [D]} \|f^{(t)}(x) - f^{(t)}(y)\|_p^p \leq O(d(x,y))^p .$$

Let Z_t be an indicator random variable for the event that $\|f^{(t)}(x) - f^{(t)}(y)\|_p^p \geq \left(\frac{d(x,y)\cdot\tau^{1-1/p}}{4\vartheta(\log k)^{1/p}}\right)^p$, and $Z = \sum_{t \in [D]} Z_t$. By Lemma 38 $\mathbb{E}[Z] \geq D/8$, and since Z_t are independent it follows by standard Chernoff bound that

$$\Pr[Z < \mathbb{E}[Z]/16] \leq e^{-D/64} \leq 1/n^2 ,$$

when c is sufficiently large constant. By a union bound on all pairs, with probability at least $1/2$, we have that for any $k \in \mathbb{N}$ and any $x, y \in X$ with $y \in N_k(x)$,

$$\|f(x) - f(y)\|_p^p = \frac{1}{D} \bigoplus_{t \in [D]} \|f^{(t)}(x) - f^{(t)}(y)\|_p^p \geq \left(\frac{d(x,y) \cdot \tau^{1-1/p}}{4\vartheta(\log k)^{1/p}}\right)^p .$$

□

9 Local Ramsey Partitions

In this section we extend the results of [MN06, Bar07] and exhibit Ramsey Partitions with improved local guarantees. These are later used to obtain improved local approximations for distance oracles and for approximate ranking.

Definition 12. *Let (X, d) be a metric space and $\vartheta \in \Xi$. Let P be a hierarchical partition*[9] *of X, let t be a parameter.*

- *A point $x \in X$ is* completely locally padded *with parameter t if $B(x, 2^i/t_i) \subseteq P_i(x)$ for $t_i = \min\{t, \vartheta(\log|B(x, 2^i)|)\}$ for all i.*
- *A point $x \in X$ is k-locally padded with parameter t if $B(x, 2^i/t) \subseteq P_i(x)$ for all $i > 0$ such that $2^i \le r_k(x)$.*

We will use a lemma that appeared in [MN06], which is a strengthening of Lemma 8. It also follows directly from the uniform padding lemma of [ABN06] (with different constants).

Lemma 39. *For any metric space (X, d), $\Delta > 0$ there exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of X such that for all $\eta \in (0, 1/8]$ and $x \in X$,*

$$\Pr[B(x, \eta\Delta) \subseteq P(x)] \ge \rho(x, \Delta/4, 4)^{-16\eta} .$$

The following Lemma extends a similar lemma of [MN06], by giving better padding parameters depending on the locality.

Lemma 40. *For any finite metric space (X, d) and parameter $t > 8$, there exists a distribution on ultrametrics such that any point $x \in X$ is completely locally padded with parameter t with probability $n^{\Omega(-1/t)}$.*

We also have the following k-local variation

Lemma 41. *For any finite metric space (X, d), $k \in \mathbb{N}$ and parameter $t > 8$, there exists a distribution on ultrametrics such that any point $x \in X$ is k-locally padded with parameter t with probability $k^{\Omega(-1/t)}$.*

Proof of Lemma 40. Create 2^i-bounded probabilistic partitions as in Lemma 39 independently for each scale $i > 0$. Let $\ell = \ell(x, t)$ be the largest integer such that $\vartheta(\log|B(x, 2^\ell)|) \le t$

$$\begin{aligned}
\Pr[\forall i, B(x, 2^i/t_i) \subseteq P(x)] &\ge \prod_{i>0} \left(\frac{|B(x, 2^i)|}{|B(x, 2^i/8)|}\right)^{-16/t_i} \\
&\ge \prod_{i=1}^{\ell} \left(\frac{|B(x, 2^i)|}{|B(x, 2^i/8)|}\right)^{-16/\vartheta(\log|B(x,2^i)|)} \cdot \prod_{i>\ell} \left(\frac{|B(x, 2^i)|}{|B(x, 2^i/8)|}\right)^{-16/t} \\
&\ge 2^{\sum_{i=1}^{\ell} \log\left(\frac{|B(x,2^i)|}{|B(x,2^i/8)|}\right)(-16/\vartheta(\log|B(x,2^i)|))} \cdot n^{-48/t}
\end{aligned}$$

We now bound the exponent of the first multiple,

$$\begin{aligned}
\sum_{i=1}^{\ell} \log\left(\frac{|B(x, 2^i)|}{|B(x, 2^i/8)|}\right)(16/\vartheta(\log|B(x, 2^i)|)) &\le \sum_{i=1}^{\ell} \sum_{j=\log|B(x,2^i/8)|}^{\log|B(x,2^i)|} 16/\vartheta(\log|B(x, 2^i)|) \\
&\le \sum_{i=1}^{\ell} \sum_{j=\log|B(x,2^i/8)|+1}^{\log|B(x,2^i)|} 16/\vartheta(j) \\
&\le 3\sum_{j>0} 16/\vartheta(j) = O(1).
\end{aligned}$$

Which gives probability $n^{\Omega(-1/t)}$ as required. ∎

[9] Recall that this is a collection of 2^i-bounded partitions P_i, for all integers i, such that if $i < j$ then P_i is a refinement of P_j.

Proof of Lemma 41. Similarly to the previous lemma, create 2^i-bounded probabilistic partitions as in Lemma 39 independently for each scale $i > 0$. For each $x \in X$ let $\ell = \max\{i \mid 2^i \le r_k(x)\}$.

$$\Pr[\forall i \in [1,\ell], B(x,2^i/t) \subseteq P(x)] \ge \prod_{i=1}^{\ell} \left(\frac{|B(x,2^i)|}{|B(x,2^i/8)|}\right)^{-16/t} \ge k^{-48/t}.$$

□

10 Embedding with Proximity Distortion

Embeddings with local distortion provide bounds on the distortion of a pair x, y as a function of how many neighbors are closer. It is natural to ask if one can provide distortion bounds that are simply a function of the distance between x and y. We call such embeddings: embeddings with proximity distortion. See Definition 2. For decomposable metrics we have the following result.

Theorem 14. *For any finite metric (X,d) on n points that admits a τ-padded decomposition and $\vartheta \in \Xi$ there exists an embedding into ℓ_p with scaling proximity distortion $O(\tau^{-1}\vartheta(\log t))$ and dimension $O(\log n)$.*

Let $D = c\log n$ for a constant c to be determined later. We will define an embedding $f : X \to \mathbb{R}^D$ with scaling proximity distortion $O(\tau^{-1}\vartheta(\log t))$. We define f by defining for each integer $s \in [D]$, a function $f^{(s)} : X \to \mathbb{R}_+$ and let $f = D^{-1/p}\bigoplus_{1\le s\le D} f^{(s)}$. Fix $s \in [D]$, and in what follows we define $f^{(s)}$. Let $I = \lceil \log \operatorname{diam}(X)\rceil$, and for every integer $i \in [I]$ construct a 2^i-bounded uniformly τ-padded probabilistic partition $\hat{\mathcal{P}}_i$ as guaranteed by Definition 10. For all $i \in [I]$ fix a partition $P_i \in \mathcal{P}_i$. Let $\{\sigma_i^{(s)}(C) | C \in P_i, i \in [I]\}$ be i.i.d symmetric $\{0,1\}$-valued Bernoulli random variables. The embedding is defined as follows: for each $x \in X$, $i \in [I]$ let $f_i^{(s)} : X \to \mathbb{R}_+$, by

$$f_i^{(s)}(x) = \sigma_i^{(s)}(P_i(x)) \cdot \frac{d(x, X \setminus P_i(x))}{\vartheta(i)},$$

and let $f^{(s)}(x) = \sum_{i\in[I]} f_i^{(s)}(x)$.

Lemma 42. *For any $x, y \in X$:*

$$|f^{(s)}(x) - f^{(s)}(y)| \le d(x,y).$$

Proof. For all $i \in [I]$, if $P_i(x) = P_i(y)$ then by the triangle inequality $f_i^{(s)}(x) - f_i^{(s)}(y) \le \frac{d(x,y)}{\vartheta(i)}$, otherwise $f_i^{(s)}(x) - f_i^{(s)}(y) \le f_i^{(s)}(x) \le \frac{d(x,y)}{\vartheta(i)}$. Similarly if we switch the roles of x, y, so we get that

$$|\sum_{i\in[I]} (f_i^{(s)}(x) - f_i^{(s)}(y))| \le \sum_{i\in[I]} |f_i^{(s)}(x) - f_i^{(s)}(y)| \le d(x,y)\sum_{i\in[I]} 1/\vartheta(i) \le d(x,y)\ .$$

□

Lemma 43. *There exists a universal constant $C_2 > 0$ such that for any $x, y \in X$ where $d(x,y) \le t$, and any $s \in [D]$, with probability at least 1/8:*

$$|f^{(s)}(x) - f^{(s)}(y)| \ge C_2 \frac{\tau \cdot d(x,y)}{\vartheta(\log t)}.$$

Proof. Let $0 < \ell \in I$ be such that $2^\ell < d(x,y) \le 2^{\ell+1}$. Note that $P_\ell(x) \ne P_\ell(y)$ and that $\ell \le \log t$. We distinguish between two cases:

Case 1: $|\sum_{i\in[I]|i\ne\ell}(f_i^{(s)}(x) - f_i^{(s)}(y))| \ge \frac{\tau\cdot d(x,y)}{4\vartheta(\log t)}$. In this case there is probability at least $1/4$ that both $\sigma_\ell^{(s)}(P_\ell(x)) = \sigma_\ell^{(s)}(P_\ell(y)) = 0$, so that $f_\ell^{(s)}(x) = f_\ell^{(s)}(y) = 0$.

Case 2: $|\sum_{i\in[I]|i\neq\ell}(f_i^{(s)}(x)-f_i^{(s)}(y))|<\frac{\tau\cdot d(x,y)}{4\vartheta(\log t)}$. Since the partition $\hat{\mathcal{P}}_\ell$ is τ-padded, $\Pr[B(x,\tau\cdot 2^\ell)\subseteq P_\ell(x)]\geq 1/2$, so there is probability $1/8$ that $\sigma_\ell^{(s)}(P_\ell(x))=1$, $\sigma_\ell^{(s)}(P_\ell(y))=0$ and $B(x,\tau\cdot 2^\ell)\subseteq P_\ell(x)$. Then follows $|f_\ell^{(s)}(x)-f_\ell^{(s)}(y)|\geq\frac{\tau\cdot 2^\ell}{\vartheta(\ell)}\geq\frac{\tau\cdot d(x,y)}{2\vartheta(\log t)}$.

In either case, with probability at least $1/8$,

$$|f^{(s)}(x)-f^{(s)}(y)|=|\sum_{i\in[I]}(f_i^{(s)}(x)-f_i^{(s)}(y))|\geq\frac{\tau\cdot d(x,y)}{4\vartheta(\log t)}.$$

□

To concludes the proof of Theorem 14 apply a Chernoff bound to obtain that at least $D/16$ coordinates $s\in[D]$ satisfy $|f^{(s)}(x)-f^{(s)}(y)|\geq\frac{\tau\cdot d(x,y)}{4\vartheta(\log t)}$ with probability $\geq 1-1/n^2$, and apply a union bound on all pairs.

10.1 Growth Bounded Metrics

For growth-bounded metrics the local distortion results can be translated into proximity distortion. Recall that a metric (X,d) is said to be χ-growth bounded if for all $x\in X$, $r>0:|B(x,2r)|\leq 2^\chi|B(x,r)|$.

Claim 44. *Let (X,d) be an χ-growth bounded metric, then there exists an embedding into ℓ_p with scaling proximity distortion $O(\vartheta(\chi\log t))$.*

Proof. By definition of growth bound if $x,y\in X$ such that $d(x,y)\leq t$ then $|B(x,t)|\leq t^\chi$ (recall that we assume $d(x,y)\geq 1$ for $x\neq y$), hence $y\in B(x,r_{t^\chi}(x))$, so there exists an embedding where the distortion of x,y is bounded by $O(\vartheta(\log(t^\chi)))=O(\vartheta(\chi\log t))$. □

All the other results translate into proximity distortion for growth-bounded metrics in a similar manner.

11 Applications

11.1 Small World Model

For many applications, the notion of distortion can be too restrictive, as it requires a bound on all pairs of points in the metric space. In some situations, we could be satisfied with a promise on the average performance guarantee of an embedding, the notions of *average distortion* and *l_q-distortion* have been extensively studied in [ABN06].

Definition 13. *Let (X,d), (Y,ρ) be metric spaces, and $f:X\to Y$ be a non-expansive embedding. The average distortion of f with respect to a distribution Π over $\binom{X}{2}$ is*

$$\mathrm{avgdist}^{(\Pi)}(f)=\sum_{x,y\in X}\Pi(x,y)\frac{d(x,y)}{\rho(f(x),f(y))}\ .$$

The distortion of average of f with respect to Π is

$$\mathrm{distavg}^{(\Pi)}(f)=\frac{\sum_{x,y\in X}\Pi(x,y)d(x,y)}{\sum_{x,y\in X}\Pi(x,y)\rho(f(x),f(y))}\ .$$

Given a metric space (X,d) and a certain distributions Π on $\binom{X}{2}$ such that local pairs are given higher probability, then our embedding techniques yields constant average distortion with respect to Π. For instance, let $\alpha>0$ and consider any of Kleinberg's "small world" distributions $\Pi(x,y\mid x)=\frac{k^{-(1+\alpha)}}{\sum_{i=1}^n i^{-(1+\alpha)}}$ [Kle00], where $d(x,y)=r_k(x)$.

Lemma 45. *Let (X,d) be a metric space, and Π a probability distribution on $\binom{(X}{2)}$ satisfying that for all integers k and $x,y \in X$ with $d(x,y) \geq r_k(x)$, the conditional probability satisfies $\Pi(x,y \mid x) \leq \frac{1}{\vartheta(k)\cdot\vartheta(\log k)}$. Then there exists an embedding f into ℓ_p with* $\mathrm{avgdist}^{(\Pi)}(f) = O(1)$.

Proof. Let $f : X \to \ell_p$ be a scaling local embedding with distortion $c \cdot \vartheta(\log k)$ for some constant c, as given by Theorem 3. Let $\Pi(x,\cdot)$ denote the marginal distribution on the first element of $\binom{X}{2}$, then

$$\begin{aligned}
\mathrm{avgdist}^{(\Pi)}(f) &= \sum_{x\in X} \Pi(x,\cdot) \sum_{y\in X} \Pi(x,y \mid x) \frac{d(x,y)}{\|f(x)-f(y)\|_p} \\
&\leq \sum_{x\in X} \Pi(x,\cdot) \sum_{k=1}^{n} \sum_{y\in X} \Pi(x,y \mid x \wedge d(x,y) = r_k(x)) c\vartheta(\log k) \\
&\leq c \sum_{k=1}^{n} \frac{\vartheta(\log k)}{\vartheta(k)\cdot\vartheta(\log k)} = c\sum_{k=1}^{n} 1/\vartheta(k) = O(1).
\end{aligned}$$

□

Lemma 46. *Let (X,d) be a metric space, and Π a probability distribution on $\binom{(X}{2)}$ satisfying that for all integers k and $x,y \in X$ with $d(x,y) \geq r_k(x)$, the conditional probability satisfies $\Pi(x,y \mid x) \leq \frac{Z}{d(x,y)\cdot\vartheta(k)\cdot\vartheta(\log k)}$, where Z is a scaling factor. Then there exists an embedding f into ℓ_p with* $\mathrm{distavg}^{(\Pi)}(f) = \frac{\mathbb{E}_{(x,y)\sim\Pi}[d(x,y)]}{\mathbb{E}_{(x,y)\sim\Pi}[\|f(x)-f(y)\|_p]} = O(1)$.

Proof. Let $f : X \to \ell_p$ be a scaling local embedding with distortion $O(\vartheta(\log k))$ for some constant c, as given by Theorem 3.

$$\begin{aligned}
\mathrm{distavg}^{(\Pi)}(f) &= \frac{\sum_{x,y\in X} \Pi(x,y)\cdot d(x,y)}{\sum_{x,y\in X} \Pi(x,y)\cdot \|f(x),f(y)\|_p} \\
&= \frac{\sum_{x\in X} \Pi(x,\cdot) \sum_{k=1}^{n} \Pi(x,y \mid x \wedge d(x,y) = r_k(x)) d(x,y)}{\sum_{x\in X} \Pi(x,\cdot)\sum_{k=1}^{n} \Pi(x,y \mid x \wedge d(x,y) = r_k(x)) d(x,y)/(c\vartheta(\log k))} \\
&\leq \frac{Z\sum_{k=1}^{n} \frac{1}{\vartheta(k)\cdot\vartheta(\log k)}}{Z\sum_{k=1}^{n} \frac{1}{c\vartheta(k)}} \leq O(1).
\end{aligned}$$

□

11.2 Online Problems

Consider any online problem defined on a metric space, which has poly-logarithmic competitive ratio algorithm based on probabilistic embedding into a distribution of ultrametrics, e.g. the metrical task system [BBBT97], file allocation [Bar96], k-server [BBMN11]. Obtaining poly-logarithmic approximation is desirable, but it may be desirable, in addition, to obtain better results if the demand sequence happens to have a local nature.

Instead of using the standard embedding of [Bar96, FRT03, Bar04] we can use the embedding given in Theorem 6 (or Theorem 5 if k is known in advance). This provides the following local strengthening to the standard competitive ratio bound: if the request sequence is such that the objective function contains only distances between pairs u,v such that v is the k-th nearest neighbor of u then the competitive ratio improves as a function of k, that is, the $O(\log n)$ overhead due to the embedding is replaced by an overhead of only $\vartheta(\log k)$ (or $O(\log k)$ using Theorem 5). Observe that in embedding of Theorem 5 the bound on the contraction holds with probability 1, which suffices for all known applications.

11.3 Local Distance Oracles

In [MN06, Bar07] it is shown how Ramsey partitions can be used to obtain efficient proximity data structures. Using the same data structure as [MN06] we can get local variations on distance oracles. We start by showing that our construction of a "local" ultrametric can be applied in the setting of Ramsey partitions to give improved local guarantees.

Claim 47. *Let (X,d) be a metric space, let $k>1$ be an integer, fix some $t>1$, let P be a hierarchical partition, and $S\subseteq X$ be the set of points that are k-locally padded in P with parameter t. Then there exists an ultrametric T that has strong k-local distortion $O(t)$ with respect to S.*

Proof. Create an ultrametric T from the hierarchical partition P as described in the proof of Theorem 5. Recall that a cluster $C\in P_i$ is labeled by 2^i, and is called *large* if $|B(\upsilon(C),3\cdot 2^i)|>k$. Large clusters are ignored in the construction of T. Fix some $x\in S$ and $y\in X$, let $L=\min\{r_k(x),d(x,y)\}$.

We begin by showing a lower bound on $d_T(x,y)$. Let i be the maximal integer such that $L>2^{i+2}$, then as in the proof of Theorem 5, $y\notin P_i(x)$ and the cluster $P_i(x)$ is not *large*, thus $d_T(x,y)\ge 2^i\ge L/8$.

To see the upper bound on $d_T(x,y)$, let i be the minimal integer such that $d(x,y)\le 2^i/t$ (if there is no such i then $d(x,y)\ge \operatorname{diam}(X)/t$, so $d_T(x,y)\le\operatorname{diam}(X)\le t\cdot d(x,y)$). Since x is k-locally padded, $B(x,2^i/t)\subseteq P_i(x)$ thus $y\in P_i(x)$. If $P_i(x)$ is a large cluster, the least common ancestor of x,y in T will only have a smaller label, thus $d_T(x,y)\le 2^i=2t\cdot 2^{i-1}/t<2t\cdot d(x,y)$. Multiplying every label by $1/2t$ concludes the proof. □

Claim 48. *Let (X,d) be a metric space , fix some $t>1$, let P be a hierarchical partition, and $S\subseteq X$ be the set of points that are completely locally padded in P with parameter t. Then there exists an ultrametric T that has strong scaling local distortion $O(\min\{t,\vartheta(\log k)\})$ with respect to S.*

Proof. Create an ultrametric T from the hierarchical partition P as described in the proof of Theorem 6. For technical reasons, we shall only use scales which are powers of 8 in the construction of T, and define $I=\{i\in\mathbb{N} : i \bmod 3=0\}$. Recall that to create the tree T, for a cluster $C\in P_i$ where $i\in I$ we assign the label $\alpha(C)=\frac{2^i}{\min\{t,\vartheta(\log|B(\upsilon(C),2^{i+2})|)\}}$ and perform the "beam-up" and "laminarization" phases to ensure monotonicity of labels. The proof is very similar to the proof of Theorem 6, we give most of the details below. Fix $x\in S$ and $y\in X$, let $L=\min\{r_k(x),d(x,y)\}$.

To see the lower bound on $d_T(x,y)$, let $i\in I$ be the maximal such that $L>2^{i+3}$, then x,y are separated in P_i, and also $|B(\upsilon(P_i(x)),2^{i+2})|\le|B(x,2^{i+3})|\le k$, so that

$$d_T(x,y)\ge\alpha(P_i(x))=\frac{2^i}{\min\{t,\vartheta(\log|B(\upsilon(P_i(x)),2^{i+2})|)\}}\ge\frac{L}{2^6\min\{t,\vartheta(\log k)\}}\ .$$

In the reminder of the proof we show that $d_T(x,y)\le O(d(x,y))$. Let $i\in I$ be the minimal integer such that $d(x,y)\le 2^i/t_i$ and for all $i>j\in I$, $\alpha(P_j(x))\le\alpha(P_i(x))$ (so that the cluster $P_i(x)$ is not replaced by any lower level cluster in the "beam-up" phase), where t_i as is defined in Definition 12. Since P is completely locally padded, we have that $B(x,2^i/t_i)\subseteq P_i(x)$, so $y\in P_i(x)$. Since $B(\upsilon(P_i(x)),2^{i+2})\supseteq B(x,2^i)$ we have that $\alpha(P_i(x))\le 2^i/t_i$, and since $t_i\ge t_{i-3}$,

$$d_T(x,y)\le\alpha(P_i(x))\le\frac{2^i}{t_i}\le\frac{8\cdot 2^{i-3}}{t_{i-3}}\ .$$

If it is the case that $d(x,y)>2^{i-3}/t_{i-3}$, then clearly $d_T(x,y)<8d(x,y)$. Otherwise, if $d(x,y)\le 2^{i-3}/t_{i-3}$, then the minimality of i suggests that there exists $i-3>j\in I$ such that $\alpha(P_j(x))>\alpha(P_{i-3}(x))$ (otherwise $i-3$ would have been chosen), and also that $d(x,y)>2^j/t_j$ (otherwise j would have been chosen). Now observe that $B(\upsilon(P_{i-3}(x)),4\cdot 2^{i-3})\subseteq B(x,2^i)$, so that $\alpha(P_{i-3}(x))\ge 2^{i-3}/t_i$, which suggests that

$$d_T(x,y)\le\frac{8\cdot 2^{i-3}}{t_i}\le 8\alpha(P_{i-3}(x))\le 8\alpha(P_j(x))\le\frac{2^j}{t_j}<8d(x,y)\ .$$

In either case we have that $d_T(x,y)\le O(d(x,y))$. □

Theorem 15. *For any finite metric space and any $t > 1$ there exists the following type of distance oracles.*

1. **Fixed** k**:** *$O(t)$ stretch for all $y \in B(x, r_k(x))$*[10]*, $O(1)$ query time, $O(n \cdot k^{1/t})$ memory.*
2. **Scaling:** *$\min\{O(\vartheta(\log k)), O(t)\}$ stretch for all $y \in B(x, r_k(x))$ and all $k \in \mathbb{N}$, $O(1)$ query time, $O(n^{1+1/t})$ memory.*
3. **Scaling:** *k stretch for all $y \in B(x, r_k(x))$ and all $k \in \mathbb{N}$, $O(1)$ query time, $O(n)$ memory.*

Proof. The first two distance oracles are based on [MN06], we refer the reader there for more details. For the first distance oracle, we follow the proof of Lemma 4.2 from [MN06] to argue that there exist ultrametrics $T_1, \ldots, T_r$ (and an efficient randomized algorithm to find them), such that $\sum_{i=1}^{r} |T_i| = O(n \cdot k^{1/t})$, with the property that for every $x \in X$ there exists $i \in [r]$ and T_i was created from a hierarchical partition in which x is k-locally padded with parameter t. For the second distance oracle, we shall use Lemma 4.2 from [MN06] directly (with parameter $p = 1$) to argue that there exist $m = O(tn^{1/t})$ ultrametrics $T_1, \ldots, T_m$ (and an efficient randomized algorithm to find them), such that $\sum_{i=1}^{m} |T_i| = O(n^{1+1/t})$, with the property that for every $x \in X$ there exists $i \in [m]$ and T_i was created from a hierarchical partition in which x is completely locally padded with parameter t.

Then the data structure for both oracles is simply this collection of ultrametrics, every point in X has pointer to an ultrametric in which it is k-locally padded (completely padded). Given x, y, the procedure to approximate $d(x, y)$, is in $O(1)$ time inspect the ultrametric T in which x is k-locally padded (completely padded) and return $d_T(x, y)$. By Claim 47 and Claim 48 the stretch is indeed $O(t)$, $\min\{O(\vartheta(\log k)), O(t)\}$ respectively, and by construction the size of this data structures is $O(n \cdot k^{1/t})$, $O(n^{1+1/t})$ respectively. By the argument appearing in [MN06] one can preprocess an ultrametric in linear space as to return the least common ancestor of any pair in $O(1)$ time, so the query time is $O(1)$ for both cases.

The third distance oracle is the ultrametric with strong scaling local distortion k, given in Theorem 1. □

11.4 Approximate Ranking

The ranking problem is defined as follows: Given a metric space (X, d) on $\{1, \ldots, n\}$ points, find for any $x \in X$ a permutation $\pi^{(x)}$ of X, such that for all $y, z \in X$: if $y = \pi^{(x)}(i)$, $z = \pi^{(x)}(j)$ and $i < j$ then $d(x, y) \le d(x, z)$. A t-approximate ranking is relaxing the last condition to $d(x, y) \le t \cdot d(x, z)$. It is shown in [MN06], and in [Bar07] via a deterministic construction, that for any $t > 1$ there exist a data structure with $O(t)$-approximate ranking which can be pre-processed in $O(tn^{2+1/t} \log n)$ time, uses $O(tn^{1+1/t})$ space, and support queries such as $\pi^{(x)}(i)$ or finding $i \in [n]$ such that $\pi^{(x)}(i) = y$ in $O(1)$ time. We show a variation on this result, in which the approximation factor scales in a relative manner to the locality of the query points. Our data structure is based on local Ramsey partitions for embedding the points with scaling local distortion into an ultrametric. We provide a theorem similar to Theorem 15 for approximate ranking.

Theorem 16. *For any finite metric space (X, d) there exist data structures for approximate ranking, such that for any $x \in X$:*

1. **Fixed** k**:** *$O(t)$ approximation for any $y \in B(x, r_k(x))$ and any $z \in X$, $O(1)$ query time, and $O(tn \cdot k^{1/t})$ memory.*
2. **Scaling:** *$\min\{O(\vartheta(\log k)), O(t)\}$ approximation for any $y \in B(x, r_k(x))$, any $z \in X$ and all $k \in [n]$, $O(1)$ query time, and $O(tn^{1+1/t})$ memory.*
3. **Scaling:** *k approximation for all x, y such that $y, z \in B(x, r_k(x))$, $O(1)$ query time, and $O(n)$ memory.*

Proof. Again the first two constructions are based on [MN06], and we refer the reader there for more details. Fix some $x \in X$, $y \in B(x, r_k(x))$ and $z \in X$. Consider the tree T in which x is k-locally padded (completely padded) with parameter t, and do the following process to define an ordering: go over the path from the leaf x to the root of T, and

[10] If $y \notin B(x, r_k(x))$, then the query will return value at least $r_k(x)/O(t)$

for every node u on this path, append all the leaves who are descendants of u and did not appear in the ordering, to the tail of the ordering (in arbitrary order between them). By an argument similar to Lemma 4.2 of [MN06] one can have a collection of $O(tn \cdot k^{1/k})$ ultrametrics (over all of X) such that any point x is k-locally padded with parameter t in at least one of them, and there exists a collection of $O(tn^{1+1/t})$ ultrametrics, such that any point x is completely padded with parameter t in at least one of them. The data structure is this collection of ultrametrics, and by the argument appearing in [MN06], one can compute in $O(1)$ time, given i find $z = \pi^{(x)}(i)$ and given z find i such that $\pi^{(x)}(i) = z$.

The approximation guarantee follows from the padding properties of the partitions. If $y \in B(x, r_k(x))$ and $z \in X$ are such that y appear before z in $\pi^{(x)}$, then $\mathrm{lca}(x, y)$ is either descendant of $\mathrm{lca}(x, z)$ in T or equal to it. The monotonicity of labels implies that $d_T(x, y) \le d_T(x, z)$. For the first data structure, by Claim 47 we have that

$$d(x, y)/O(t) \le d_T(x, y) \le d_T(x, z) \le d(x, z) ,$$

so $d(x, y) \le O(t) \cdot d(x, z)$. For the second data structure we obtain $d(x, y) \le \min\{O(\vartheta(\log k)), O(t)\} \cdot d(x, z)$ by using Claim 48.

The third data structure is once again based on the ultrametric with strong scaling local distortion k, given in Theorem 1. □

12 Open Problems

Most of the results in this paper are either tight or nearly tight. The tightness of our k-local results follows from known metric embedding lower bounds. There are several obvious questions. There is a small gap between our scaling local distortion upper bounds (such as in theorems 3,6) of $O(\vartheta(\log k))$ and the lower bound of $\Omega(\log k)$. Another interesting question is whether there is a k-local dimension reduction with constant distortion (independent of k)? Although small $1 + \epsilon$ distortion was ruled out by [SS09], it is still concievable that every finite subset of ℓ_2 have a k-local embedding into ℓ_2^d with $O(1)$ distortion, where d depends only on k, perhaps even $d = O(\log k)$.

References

[ABC+05] Ittai Abraham, Yair Bartal, T-H. Hubert Chan, Kedar Dhamdhere Dhamdhere, Anupam Gupta, Jon M. Kleinberg, Ofer Neiman, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '05, pages 83–100, Washington, DC, USA, 2005. IEEE Computer Society.

[ABN06] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 271–286, New York, NY, USA, 2006. ACM Press.

[ABN07a] Ittai Abraham, Yair Bartal, and Ofer Neiman. Embedding metrics into ultrametrics and graphs into spanning trees with constant average distortion. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '07, pages 502–511, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.

[ABN07b] Ittai Abraham, Yair Bartal, and Ofer Neiman. Local embeddings of metric spaces. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, STOC '07, pages 631–640, New York, NY, USA, 2007. ACM.

[ABN08a] Ittai Abraham, Yair Bartal, and Ofer Neiman. Embedding metric spaces in their intrinsic dimension. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '08, pages 363–372, Philadelphia, PA, USA, 2008. Society for Industrial and Applied Mathematics.

[ABN08b] Ittai Abraham, Yair Bartal, and Ofer Neiman. Nearly tight low stretch spanning trees. In *FOCS '08: Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 781–790, Washington, DC, USA, 2008. IEEE Computer Society.

[ABN09] Ittai Abraham, Yair Bartal, and Ofer Neiman. On low dimensional local embeddings. In *SODA '09: Proceedings of the Nineteenth Annual ACM -SIAM Symposium on Discrete Algorithms*, pages 875–884, Philadelphia, PA, USA, 2009. Society for Industrial and Applied Mathematics.

[AKPW95] Noga Alon, Richard M. Karp, David Peleg, and Douglas West. A graph-theoretic game and its application to the k-server problem. *SIAM J. Comput.*, 24(1):78–100, 1995.

[Alo03] Noga Alon. Problems and results in extremal combinatorics. *I, Discrete Math.*, 273:31–53, 2003.

[AP90] Baruch Awerbuch and David Peleg. Sparse partitions. In *Proceedings of the 31st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 503–513, 1990.

[Bar96] Yair Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*, pages 184–193. IEEE Comput. Soc. Press, Los Alamitos, CA, 1996.

[Bar98] Yair Bartal. On approximating arbitrary metrics by tree metrics. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 183–193, 1998.

[Bar04] Yair Bartal. Graph decomposition lemmas and their role in metric embedding methods. In *12th Annual European Symposium on Algorithms*, pages 89–97, 2004.

[Bar07] Yair Bartal. Metric ramsey decompositions and their applications, 2007. manuscript.

[BBBT97] Yair Bartal, Avrim Blum, Carl Burch, and Andrew Tomkins. A polylog(n)-competitive algorithm for metrical task systems. In *STOC '97: Proceedings of the twenty-nineth annual ACM symposium on Theory of computing*, pages 711–719. ACM Press, 1997.

[BBMN11] Nikhil Bansal, Niv Buchbinder, Aleksander Madry, and Joseph Naor. A polylogarithmic-competitive algorithm for the k-server problem. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 267 –276, oct. 2011.

[Bec91] József Beck. An algorithmic approach to the Lovàsz local lemma. i. *Random Structures and Algorithms*, 2:343–366, 1991.

[BLMN04] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. Low dimensional embeddings of ultrametrics. *Eur. J. Comb.*, 25:87–92, January 2004.

[BLMN05] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric ramsey-type phenomena. *Annals Math*, 162(2):643–709, 2005.

[BM04] Yair Bartal and Manor Mendel. Dimension reduction for ultrametrics. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 664–665. Society for Industrial and Applied Mathematics, 2004.

[BN03] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6):1373–1396, 2003.

[Bol78] Béla Bollobás. *Extremal graph theory*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1978.

[Bou85] Jean Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[BRS11] Yair Bartal, Ben Recht, and Leonard J. Schulman. Dimensionality reduction: beyond the johnson-lindenstrauss bound. In *SODA '11: Proceedings of the 24th annual ACM-SIAM Symposium on Discrete Algorithms*, 2011.

[CCRK04] Manuel Costa, Miguel Castro, Antony I. T. Rowstron, and Peter B. Key. Pic: Practical internet coordinates for distance estimation. In *24th International Conference on Distributed Computing Systems*, pages 178–187, 2004.

[CKR01] Gruia Calinescu, Howard Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, SODA '01, pages 8–16, Philadelphia, PA, USA, 2001. Society for Industrial and Applied Mathematics.

[CMM10] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Local global tradeoffs in metric embeddings. *SIAM J. Comput.*, 39:2487–2512, April 2010.

[EEST05] Michael Elkin, Yuval Emek, Daniel A. Spielman, and Shang-Hua Teng. Lower-stretch spanning trees. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 494–503, New York, NY, USA, 2005. ACM Press.

[FRT03] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *STOC '03: Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 448–455. ACM Press, 2003.

[FT03] Jittat Fakcharoenphol and Kunal Talwar. An improved decomposition theorem for graphs excluding a fixed minor. In *RANDOM-APPROX*, pages 36–46, 2003.

[Gil52] Edgar N. Gilbert. A comparison of signalling alphabets. *Bell systems technical journal*, 31:504–522, 1952.

[GK11] Lee-Ad Gottlieb and Robert Krauthgamer. A nonlinear approach to dimension reduction. In *Proceedings of the twenty second annual ACM-SIAM symposium on Discrete algorithms*, pages 888–899, 2011.

[GKL03] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *FOCS '03: Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, page 534, Washington, DC, USA, 2003. IEEE Computer Society.

[HPM06] Sariel Har-Peled and Manor Mendel. Fast construction of nets in low-dimensional metrics and their applications. *SIAM J. Comput*, 35(5):1148–1184, 2006.

[IN07] Piotr Indyk and Assaf Naor. Nearest-neighbor-preserving embeddings. *ACM Transactions on Algorithms*, 3(3), 2007.

[Ind01] Piotr Indyk. Algorithmic applications of low-distortion geometric embeddings. In *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, pages 10–33, 2001.

[JL84] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[Kle00] Jon Kleinberg. The small-world phenomenon: an algorithm perspective. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 163–170, New York, NY, USA, 2000. ACM Press.

[KLMN04] Robert Krauthgamer, James R. Lee, Manor Mendel, and Assaf Naor. Measured descent: A new embedding method for finite metrics. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 434–443. IEEE, October 2004.

[KPR93] Philip Klein, Serge A. Plotkin, and Satish Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, STOC '93, pages 682–690, New York, NY, USA, 1993. ACM.

[KSW04] Jon M. Kleinberg, Aleksandrs Slivkins, and Tom Wexler. Triangulation and embedding using small sets of beacons. In *FOCS*, pages 444–453, 2004.

[LNK+05] David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Science*, 102:11623–11628, August 2005.

[Mat90] Jiri Matoušek. Note on bi-lipschitz embeddings into low-dimensional euclidean spaces. *Comment. Math. Univ. Carolinae*, 31:589–600, 1990.

[Mat97] Jiri Matoušek. On embedding expanders into l_p spaces. *Israel J. Math.*, 102:189–197, 1997.

[MN04] Manor Mendel and Assaf Naor. Euclidean quotients of finite metric spaces. *Advances in Mathematics*, 189(2):451–494, 2004.

[MN06] Manor Mendel and Assaf Naor. Ramsey partitions and proximity data structures. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 109–118, Washington, DC, USA, 2006. IEEE Computer Society.

[Mos09] Robin A. Moser. A constructive proof of the lovász local lemma. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 343–350, New York, NY, USA, 2009. ACM.

[MT10] Robin A. Moser and Gábor Tardos. A constructive proof of the general lovász local lemma. *J. ACM*, 57(2):1–15, 2010.

[Pel00] David Peleg. *Distributed Computing: A Locality-Sensitive Approach*. SIAM Monographs on Discrete Mathematics and Applications, 2000.

[Rao99] Satish Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry*, pages 300–306, New York, 1999. ACM.

[SS09] Gideon Schechtman and Adi Shraibman. Lower bounds for local versions of dimension reductions. *Discrete Comput. Geom.*, 41:273–283, January 2009.

[XSB06] Lin Xiao, Jun Sun, and Stephen Boyd. A duality view of spectral methods for dimensionality reduction. In *ICML '06: Proceedings of the 23rd international conference on Machine learning*, pages 1041–1048, New York, NY, USA, 2006. ACM Press.

A Proof of Lemma 5

The following decomposition lemma was shown in [ABN06].

Lemma 49 (Probabilistic Decomposition). *For any metric space (X, d), a subset $Z \subseteq X$, a point $v \in X$, real parameters $\chi \geq 2, \Delta > 0$, let r be a random variable sampled from a truncated exponential density function with parameter $\lambda = 8\ln(\chi)/\Delta$*

$$f(r) = \begin{cases} \frac{\chi^2}{1-\chi^{-2}}\lambda e^{-\lambda r} & r \in [\Delta/4, \Delta/2] \\ 0 & \textit{otherwise} \end{cases}$$

If $S = B(v, r) \cap Z$ and $\bar{S} = Z \setminus S$ then for any $\theta \in [\chi^{-1}, 1)$ and any $x \in Z$:

$$\Pr\left[B(x,\eta\Delta)\bowtie(S,\bar{S})\right]\le(1-\theta)\left(\Pr\left[B(x,\eta\Delta)\nsubseteq\bar{S}\right]+\frac{2\theta}{\chi}\right).$$

where $\eta=2^{-4}\ln(1/\theta)/\ln\chi$.

We are now ready to prove Lemma 5. By sub-partition we mean a partition $\{C_i\}_i$ lacking the requirement that $\bigcup_i C_i=X$. The intuition behind the construction is that we perform the partition of [ABN06] as long as the local growth rate is small enough. Once the growth rate is large with respect to the decomposability parameter, we assign all the points who were not covered by the first partition a cluster generated by the probabilistic partition known to exists from Definition 10. This is done in two phases:

Phase 1 Define the sub-partition P_1 of X into clusters by generating a sequence of clusters: $C_1,C_2,\ldots C_s$, for some $s\in[n]$. Notice that we are generating a distribution over sub-partitions and therefore the generated clusters are random variables. First we deterministically assign centers $v_1,v_2,\ldots,v_s$ and parameters $\chi_1,\chi_2,\ldots,\chi_s$. Let $W_1=X$ and $j=1$. Conduct the following iterative process:

1. Let $v_j\in W_j$ be the point minimizing $\hat{\chi}_j=\rho(x,2\Delta,\gamma)$ over all $x\in W_j$.
2. If $2^6\ln(\hat{\chi}_j)>\tau^{-1}$ set $s=j-1$ and stop.
3. Set $\chi_j=\max\{2/\hat{\delta}^{1/4},\hat{\chi}_j\}$.
4. Let $W_{j+1}=W_j\setminus B(v_j,\Delta/4)$.
5. Set $j=j+1$. If $W_j\neq\emptyset$ return to (1).

Now the algorithm for the partition and functions ξ,η is as follows: Let $Z_1=X$. For $j=1,2,3\ldots s$:

1. Let $(S_{v_j},\bar{S}_{v_j})$ be the partition created by invoking Lemma 49 on Z_j with center $v=v_j$ and parameter $\chi=\chi_j$.
2. Set $C_j=S_{v_j}$, $Z_{j+1}=\bar{S}_{v_j}$.
3. For all $x\in C_j$ let $\eta_P(x)=2^{-7}/\max\{\ln\hat{\chi}_j,\ln(1/\hat{\delta})\}$. If $\hat{\chi}_j\ge1/\hat{\delta}$ set $\xi_P(x)=1$, otherwise set $\xi_P(x)=0$.

Fix some $\hat{\delta}\le\delta\le1$. Let $\theta=\delta^{1/4}$. Note that $\theta\ge2\chi_j^{-1}$ for all $j\in[s]$ as required. Recall that $\eta_j=2^{-4}\ln(1/\theta)/\ln\chi_j=2^{-6}\ln(1/\delta)/\ln\chi_j$ (it is easy to verify that $\eta_P(x)\cdot\ln(1/\delta)\le\eta_j$). Observe that some clusters may be empty and that it is not necessarily the case that $v_m\in C_m$.

Phase 2 In this phase we assign any points left un-assigned from *phase 1*. Let $P_2'=\{D_1,D_2,\ldots,D_t\}$ be a Δ-bounded probabilistic partition of X, such that for all $\delta\le1$ satisfying $\ln(1/\delta)\le2^6\tau^{-1}$, P_2' is $(\tau\cdot\ln(1/\delta),\delta)$-padded, this probabilistic partition exists by Definition 10. Let $Z=\bigcup_{i=1}^s C_i$ and $\bar{Z}=X\setminus Z$ (the un-assigned points), then let $P_2=\{D_1\cap\bar{Z},D_2\cap\bar{Z},\ldots,D_t\cap\bar{Z}\}$. For all $x\in\bar{Z}$ let $\eta_P(x)=\tau/2$ and $\xi_P(x)=1$. It can be checked that $\eta_P^{(\delta)}(x)\le\eta_j$ for all $j\in[s]$. Notice that by the stop condition of *phase 1*, $\tau\le2^{-6}/\ln\hat{\chi}_j$, since by definition $\tau\le2^{-6}/\ln(1/\hat{\delta})$ as well follows that for all $x\in\bar{Z}$ and $j\in[s]$, $\eta_P(x)\cdot\ln(1/\delta)\le\eta_j$.

Define $P=P_1\cup P_2$. We now prove the properties in the lemma for some $x\in X$, first consider the sub-partition P_1, and the distribution over the clusters $C_1,C_2,\ldots C_s$ as defined above. For $1\le m\le s$, define the events:

$$\begin{aligned}\mathcal{Z}_m&=\{\forall j,1\le j<m,\ B(x,\eta_j\Delta)\subseteq Z_{j+1}\},\\ \mathcal{E}_m&=\{\exists j,\ m\le j<s\text{ s.t. }B(x,\eta_j\Delta)\bowtie(S_{v_j},\bar{S}_{v_j})|\mathcal{Z}_m\}.\end{aligned}$$

Also let $T=T_x=B(x,\Delta)$. We prove the following inductive claim: For every $1\le m\le s$:

$$\Pr[\mathcal{E}_m]\le(1-\theta)(1+\theta\sum_{j\ge m,v_j\in T}\chi_j^{-1}).\tag{21}$$

Note that $\Pr[\mathcal{E}_s] = 0$. Assume the claim holds for $m+1$ and we will prove for m. Define the events:

$$\begin{aligned}
\mathcal{F}_m &= \{B(x,\eta_m\Delta) \bowtie (S_{v_m}, \bar{S}_{v_m}) | \mathcal{Z}_m\}, \\
\mathcal{G}_m &= \{B(x,\eta_m\Delta) \subseteq \bar{S}_{v_m} | \mathcal{Z}_m\} = \{\mathcal{Z}_{m+1}|\mathcal{Z}_m\}, \\
\bar{\mathcal{G}}_m &= \{B(x,\eta_m\Delta) \nsubseteq \bar{S}_{v_m} | \mathcal{Z}_m\} = \{\mathcal{Z}_{m+1}|\mathcal{Z}_m\}.
\end{aligned}$$

First we bound $\Pr[\mathcal{F}_m]$. Recall that the center v_m of C_m and the value of χ_m are determined deterministically. The radius r_m is chosen from the interval $[\Delta/4, \Delta/2]$. Since $\eta_m \le 1/2$, if $B(x,\eta_m\Delta) \bowtie (S_{v_m}, \bar{S}_{v_m})$ then $d(v_m, x) \le \Delta$, and thus $v_m \in T$. Therefore if $v_m \notin T$ then $\Pr[\mathcal{F}_m] = 0$. Otherwise by Lemma 49

$$\begin{aligned}
&\Pr[\mathcal{F}_m] \qquad (22)\\
&= \Pr[B(x,\eta_m\Delta) \bowtie (S_{v_m}, \bar{S}_{v_m}) | \mathcal{Z}_m] \\
&\le (1-\theta)(\Pr[B(x,\eta_m\Delta) \nsubseteq \bar{S}_{v_m} | \mathcal{Z}_m] + \theta\chi_m^{-1}) \\
&= (1-\theta)(\Pr[\bar{\mathcal{G}}_m] + \theta\chi_m^{-1}).
\end{aligned}$$

Since the choice of radius is the only randomness in the process of creating P_1, the event of padding for $z \in Z$, and the event $B(z, \eta_P(z)\Delta) \cap Z = \emptyset$ for $z \in \bar{Z}$ are independent of all choices of radii for centers $v_j \notin T_z$. That is, for any assignment to clusters of points outside $B(z, 2\Delta)$ (which may determine radius choices for points in $X \setminus B(x,\Delta)$), the padding probability will not be affected. Using the induction hypothesis we prove the inductive claim:

$$\begin{aligned}
\Pr[\mathcal{E}_m] &\le \Pr[\mathcal{F}_m] + \Pr[\mathcal{G}_m]\Pr[\mathcal{E}_{m+1}] \\
&\le (1-\theta)(\Pr[\bar{\mathcal{G}}_m] + \theta\mathbf{1}_{\{v_m\in T\}}\chi_m^{-1}) + \\
&\quad \Pr[\mathcal{G}_m]\cdot(1-\theta)(1+\theta\sum_{j\ge m+1, v_j\in T}\chi_j^{-1}) \\
&\le (1-\theta)(1+\theta\sum_{j\ge m, v_j\in T}\chi_j^{-1}),
\end{aligned}$$

The second inequality follows from (22) and the induction hypothesis. Fix some $x \in X$, $T = T_x$. Observe that for all $v_j \in T$, $d(v_j, x) \le \Delta$, and so we get $B(v_j, 2\Delta/\gamma) \subseteq B(x, 2\Delta)$. On the other hand $B(v_j, 2\gamma\Delta) \supseteq B(x, 2\Delta)$. Note that the definition of W_j implies that if v_j is a center then all the other points in $B(v_j, \Delta/4)$ cannot be a center as well, therefore for any $j \ne j'$, $d(v_j, v_{j'}) > \Delta/4 \ge 4\Delta/\gamma$, so that $B(v_j, 2\Delta/\gamma) \cap B(v_{j'}, 2\Delta/\gamma) = \emptyset$. Hence, we get:

$$\begin{aligned}
\sum_{j\ge 1, v_j\in T}\chi_j^{-1} &\le \sum_{j\ge 1, v_j\in T}\hat{\chi}_j^{-1} \\
&\le \sum_{j\ge 1, v_j\in T}\frac{|B(v_j, 2\Delta/\gamma)|}{|B(v_j, 2\gamma\Delta)|} \\
&\le \sum_{j\ge 1, v_j\in T}\frac{|B(v_j, 2\Delta/\gamma)|}{|B(x, 2\Delta)|} \le 1.
\end{aligned}$$

We conclude from the claim (21) for $m = 1$ that

$$\Pr[\mathcal{E}_1] \le (1-\theta)(1+\theta\cdot\sum_{j\ge 1, v_j\in T}\chi_j^{-1}) \le (1-\theta)(1+\theta) \le 1-\delta^{1/2}.$$

Hence there is probability at least $\delta^{1/2}$ that event $\neg\mathcal{E}_1$ occurs. Given that this happens, we will show that there is probability at least $\delta^{1/2}$ that x is padded. If $x \in Z$, then let $j \in [s]$ such that $P(x) = C_j$, then $\eta_P(x)\cdot\ln(1/\delta) \le \eta_j$ and so $B(x, \eta_P(x)\cdot\ln(1/\delta)\Delta) \subseteq B(x, \eta_j\Delta)$. Note that if $x \in Z$ is padded in P_1 it will be padded in P. If $x \in \bar{Z}$: since for any $j \in [s]$, $\eta_P(x)\cdot\ln(1/\delta) \le \eta_j$ we have that $\neg\mathcal{E}_1$ implies that $B(x, \eta_P(x)\cdot\ln(1/\delta)\Delta) \cap Z = \emptyset$. As P_2 is performed independently of P_1 we have $\Pr[B(x, (\tau/2)\ln(1/\delta)) \subseteq P_2(x)] \ge \delta^{1/2}$, hence

$$\Pr[B(x, (\tau/2)\ln(1/\delta)) \subseteq P(x)] \ge \Pr[B(x, (\tau/2)\ln(1/\delta)) \subseteq P(x) \mid \neg\mathcal{E}_1]\cdot\Pr[\neg\mathcal{E}_1] \ge \delta^{1/2}\cdot\delta^{1/2} = \delta.$$

It follows that $\hat{\mathcal{P}}$ is uniformly padded. Finally, we show the properties stated in the lemma. The first property follows from the stop condition in *phase 1* and from the definition of $\eta_P(x)$. The second property holds: first take $x \in Z$ and let j be such that $x \in C_j$, then $\xi_P(x) = 1$ implies that $\hat{\chi}_j \geq 1/\hat{\delta}$ hence $\eta_P(x) = 2^{-7}/\ln \hat{\chi}_j = 2^{-7}/\ln \rho(v_j, 2\Delta, \gamma)$ and by the minimality of v_j, $\eta_P(x) \geq 2^{-7}/\ln \rho(x, 2\Delta, \gamma)$. By definition $\eta_P(x) \leq 2^{-7}/\ln(1/\hat{\delta})$. If $x \in \bar{Z}$ then $\eta_P(x) = \tau/2$, by the stop condition of *phase 1* $\tau/2 \geq 2^{-7}/\ln \hat{\chi}_j$. Again by definition of $\hat{\delta}$ follows that $\tau/2 \leq 2^{-7}/\ln(1/\hat{\delta})$. As for the third property, which is meaningful only for $x \in Z$, let j such that $x \in C_j$, then $\xi_P(x) = 0$ implies that $\hat{\chi}_j < 1/\hat{\delta}$ hence $\eta_P(x) = 2^{-7}/\ln(1/\hat{\delta})$ and since $d(x, v_j) \leq \Delta$ also $\bar{\rho}(x, 2\Delta, \gamma) \leq \rho(v_j, 2\Delta, \gamma) < 1/\hat{\delta}$.