Local Embeddings of Metric Spaces

Extended Abstract

Ittai Abraham
Hebrew University
ittaia@cs.huji.ac.il

Yair Bartal*
Hebrew University and Caltech
yair@cs.huji.ac.il

Ofer Neiman†
Hebrew University
neiman@cs.huji.ac.il

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 we initiate the study of *local embeddings* of metric spaces and provide embeddings with distortion depending solely on the local structure of the space.

Categories and Subject Descriptors: F.2.0 [Theory of Computation]: Analysis of Algorithms and Problem Complexity – General

General Terms: Algorithms, Theory

Keywords: Metric Embedding

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. [20]). 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 [13]. 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 [26, 16].

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 [13] and [28] 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 [21] 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 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.* [23] 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 [22] 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 [2, 1]. 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 [22] when indeed preserving local distances is more important than preserving far away distances.

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. A 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 faithfully preserves distances between all nearby neighbor points, i.e. the local structure of the space.

In large scale systems it is often the case that one wants to maintain a compact data structure known as a *Distance Oracle* [26]. More demanding tasks are *name independent compact routing schemes* where the name of the node is independent of its location [6] and *mobile user schemes* which are competitive distributed protocols for routing when the target may be mobile. In all these settings it is clearly desirable to obtain improved results for close by neighbors.

1.1 Local embeddings

We now formally define new notions of *local distortion*. Given a metric space (X, d), let $B(u, r) = \{v \mid d(u, v) \leq r\}$. For any point x let $<_x$ be an 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. 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[1]) 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.*

---

[1]Note that all of the scaling local distortion results in this paper are in fact in the strong sense.

- *Given a distribution $\mathcal{D}$ on maps $f : X \to Y$, we say that $\mathcal{D}$ has probabilistic (strong) $\{k, scaling\}$-local distortion if the appropriate lower bound holds and the appropriate upper bound holds in expectation over $\mathcal{D}$.*

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

DEFINITION 2. *Let (X, d_X) be a metric space on n points with $\min_{x,y \in X}\{d(x, y)\} \geq 1$, 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. All of our scaling results are strong. Theorem 1 shows that any metric space can be embedded into a single ultrametric (and hence a tree metric) with strong scaling local distortion k. Using a variant of Bourgain's embedding shows that strong k-local embeddings with distortion $O(\log k)$ are possible for any fixed k. Using the embeddings of [2] we can also bound the dimension by $O(\log n)$. This is presented in Theorem 2. In Theorem 3 we give an embedding with strong scaling local distortion of $\tilde{O}(\log k)$ using a variation of Bourgain's embedding method. 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 continue with presenting 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 strong scaling local distortion $\tilde{O}(\log k)$, and worst case distortion $O(\log n)$. Our embeddings are based on the existing probabilistic embedding techniques of [7, 8, 18, 9]. Yet, the proof of Theorem 6 differs from all known embeddings into ultrametrics. One indication of this is the fact that all previous embedding are non-contracting, while the probabilistic embedding presented here is *non-expanding* in expectation, and the contraction factors vary for different pairs. Achieving this requires to make subtle modifications of known ultrametric construction algorithms.

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. In Theorem 8 we significantly improve this result and provide a novel form of dimension reduction resulting in embeddings that require only $O(\log k)$ dimensions. Our result requires that the metric space obeys a very weak form of growth bound (formally defined later). Using a subtle argument based on the partition-based embeddings of [2] together with a non-standard variant of the Lovász local lemma, we prove the existence of embeddings with k local distortion $O(\log k)$ in dimension $O(\log k)$.

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. Using embeddings

based on partitions, Theorem 10 provides better local scaling distortion for metrics with improved dimension and embeddings with improved distortion for decomposable (including doubling) metrics. Following [12, 25, 11] we provide local metric Ramsey-type theorems. In particular, we 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, [21] studied a distribution that depends on the distance with exactly this type of local behavior. In Section 9 we study embeddings with *proximity distortion* – in which the distortion bound of a pair x, y is a function of $d(x, y)$. Theorem 12 is our main result using this notion. Using partition-based embeddings [2] we show that embeddings into L_p with scaling proximity distortion $\tilde{O}(\log t)$ are given for decomposable metrics (these include in particular doubling and planar metrics).

In Section 10 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 [27, 25, 11]. For example, we provide distance oracles with linear storage and strong scaling local stretch $\tilde{O}(\log k)$ (that is the stretch for the kth nearest neighbor) under weak growth bound assumptions.

2. PRELIMINARIES

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

DEFINITION 3 (GROWTH BOUND). *Let (X, d) be a metric space and $\chi \geq 1$ a fixed real constant.*

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

Note that the very 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

$$\int_1^\infty \frac{dx}{\vartheta(x)} = 1 . \tag{1}$$

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.

3. LOCAL PROBABILISTIC PARTITIONS

Several of our results use probabilistic partitions [7]. In this section we review some definitions and results concerning these tools, extending the notions from [2].

DEFINITION 4. *The local growth rate of $x \in X$ at radius $r > 0$ for given scale factors $\gamma_1 \geq \gamma_2 > 0$ is defined as $\rho(x, r, \gamma_1, \gamma_2) = |B(x, r\gamma_1)|/|B(x, r\gamma_2)|$. Given a scale factor γ is it useful to denote $\rho(x, r, \gamma) = \rho(x, r, \gamma, 1/\gamma)$. 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 at radius $r > 0$ and scale $\gamma > 0$ is defined as $\bar{\rho}(x, r, \gamma) = \rho(B(x, r), r, \gamma)$.*

CLAIM 1. *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*

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

DEFINITION 5 (PARTITION). *Let (X, d) be a finite metric space. A partition P of X is a collection of disjoint set of non-empty clusters $\mathcal{C}(P) = \{C_1, C_2, \ldots, C_t\}$ such that $X = \cup_j C_j$. The sets C_i are called clusters. For $x \in X$ we denote by $P(x)$ the cluster containing x. Define $\upsilon : P \to X$ as $\upsilon(C_i) = c_i$, where $c_i \in C_i$ is an arbitrary point. Given $\Delta > 0$, a partition is Δ-bounded if for all $1 \leq j \leq t$, $\operatorname{diam}(C_j) \leq \Delta$. 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 $f(x) = f(y)$. For $Z \subseteq X$ we denote by $P_{|Z}$ the restriction of P to points in Z.*

DEFINITION 6 (PROBABILISTIC PARTITION). *A probabilistic partition $\hat{\mathcal{P}}$ consists of a probability distribution over a set of partitions $\mathcal{P}$. $\hat{\mathcal{P}}$ is called Δ-bounded if every $P \in \mathcal{P}$ is Δ-bounded. Given a collection of functions $\eta = \{\eta_P : X \to [0, 1] \mid P \in \mathcal{P}\}$, a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ is called η-padded if for every $x \in X$:*

$$\Pr[B(x, \eta_P(x)) \subseteq P(x)] \geq 1/2$$

DEFINITION 7 (UNIFORMLY PADDED LOCAL PP). *Given $\Delta > 0$, let $\hat{\mathcal{P}}$ be a Δ-bounded probabilistic partition of (X, d). Given collection of functions $\eta = \{\eta_P : X \to [0, 1] | P \in \mathcal{P}\}$ such that η_P is a uniform function with respect to P. We say that $\hat{\mathcal{P}}$ is a η-uniformly padded local probabilistic partition if the event $B(x, \eta_P(x)\Delta) \subseteq P(x)$ occurs with probability $1/2$ and is independent of the structure of the partition outside $B(x, 2\Delta)$. Formally for all $C \subseteq X \setminus B(x, 2\Delta)$ and all partitions P' of C,*

$$\Pr[B(x, \eta_P(x)\Delta) \subseteq P(x) \mid P_{|C} = P'] \geq 1/2$$

DEFINITION 8. *Let (X, d) be a finite metric space. Let $\tau \in (0, 1]$. We say that X admits a τ-decomposition if for every $0 < \Delta \leq \operatorname{diam}(X)$ there exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of X such that $\hat{\mathcal{P}}$ is τ-padded.*

It is known that λ-doubling metrics admit a $\Omega(\log^{-1} \lambda)$-decomposition and K_r-minor excluded graphs admit a $\Omega(r^{-2})$-decomposition.

The following Lemma with local properties is proven in [3].

LEMMA 2 (UNIFORM PADDING LEMMA). *Let (X, d) be a finite metric space. Let $0 < \Delta \leq \operatorname{diam}(X)$. Let $\Gamma \geq 64$.*

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 $\eta = \{\eta_P : X \to (0,1] \mid P \in \mathcal{P}\}$ such that the probabilistic partition $\hat{\mathcal{P}}$ is a η-uniformly padded local probabilistic partition; and the following conditions hold for any $P \in \mathcal{P}$ and any $x \in X$:

1. If $\xi_P(x) = 1$ then: $2^{-7}/\ln\rho(x,8\Delta,\Gamma) \leq \eta_P(x) \leq 2^{-7}$.
2. If $\xi_P(x) = 0$ then: $\eta_P(x) = 2^{-7}$ and $\bar{\rho}(x,8\Delta,\Gamma) < 2$.

Furthermore, if X admits a τ-decomposition then the probabilistic partition $\hat{\mathcal{P}}$ is a η-uniformly padded probabilistic partition; and in addition to conditions 1. and 2. the following additional condition holds for any $P \in \mathcal{P}$ and any $x \in X$:

3. $\eta_P(x) \geq \tau/2$.

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 [7, 12, 19]:

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

PROOF. Let $G = (V,E,w)$ be some graph with weights w on the edges, whose shortest-path metric is the metric of X, and let $M = (V,E',w)$ be its minimum spanning tree. We now define a recursive process for constructing the ultrametric T given M:

Let $e \in E'$ be the largest edge in M. Create the root of T with label $w(e_1)$, the root will have 2 children which are the ultrametrics created recursively on the spanning trees M_1 and M_2, which are the 2 connected components of $M \setminus \{e\}$. Note that the removal of e divides M into 2 disjoint trees M_1, M_2, and each is indeed a minimum spanning tree on its vertices.

From the choice of e any $e' \in M \setminus \{e\}$ satisfy $w(e') \leq w(e)$, hence T is indeed an ultra-metric. It remains to show that the k-local distortion is at most k, for all $k > 0$ simultaneously. Let $u,v \in X$ such that $v \in B(u,r_k(u))$, and assume they were first separated in the i-th step, meaning that if $M^{(i)}$ is the MST of a connected component containing u and v after $i-1$ recursive applications of the process, then their distance in T is fixed to be $\ell = w(e_i)$ (where e_i is the largest edge in $M^{(i)}$). Removing e_i from $M^{(i)}$ divides it to $M_1^{(i)}$ and $M_2^{(i)}$.

A known property of minimum spanning trees is that the largest edge in any cycle of G cannot be in M, hence it can be inferred that there is no edge shorter than ℓ connecting $M_1^{(i)}$ to $M_2^{(i)}$, hence $d_T(u,v) = \ell \leq d_X(u,v)$. Assume by contradiction that $d_X(u,v) > k \cdot d_T(u,v) = k \cdot \ell$, then for each $i = 0,\ldots,k$ let $B_i = B_X(u, i \cdot \ell)$ and consider the shells $S_i = B_{i+1} \setminus B_i$. Since there are at most k vertices in $B(u, d_X(u,v))$ and $d_X(u,v) > k \cdot \ell$, there are at most $k-1$ points other than u in $B(u, k \cdot \ell)$. Hence there exists $i \in [k]$ such that S_i is empty. Therefore to connect u and v by a path in $M^{(i)}$ we must use an edge greater than ℓ in $M^{(i)}$ - contradiction to the choice of e_i. $\square$

To get a strong scaling local distortion, one simply replace in the lower bound analysis $d_X(u,v)$ with $r_{k'}(u)$ for some $k' \leq k$.

4.2 Embedding into ℓ_p with k-local distortion

THEOREM 2. *For any finite metric space X on n points there exists an embedding into ℓ_p with strong k-local distortion $O(\frac{\log k}{p})$ and dimension $O(2^p \log n)$.*

The embedding is a variant of the partition-based embeddings of [10, 2]. A simpler proof based on Bourgain's embedding [14] and Matoušek's improvements for ℓ_p [24] gives a slightly larger dimension of $O(2^p \log n \log k)$. The most basic result is obtained simply by taking the first $\theta(\log k)$ coordinates of Bourgain's embedding. We defer the proof to the full paper. We give stronger results in Section 6 where the dimension is only $O(\log k)$ given a very weak growth bound.

4.3 Embedding into ℓ_p with scaling local distortion

Our embedding is based on Bourgain's embedding [14] and Matoušek's improvements for ℓ_p [24]. We use a function $\vartheta \in \Xi$ to scale each coordinate as described below.

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 $O((1/p)\log n)$ and dimension $O(2^p \log^2 n)$.*

PROOF. Let $s = 2^p$, $t = \log_s n$, $T = \{i \mid 1 \leq i \leq t\}$, $q = O(s\log n)$ and $Q = \{j \mid 1 \leq j \leq q\}$. Choose random subsets A_{ij} for $i \in T$, $j \in Q$, such that each point included in A_{ij} independently with probability $\frac{1}{s^i}$. We now define the embedding $\phi : X \to \ell_p^{t \cdot q}$ by defining for each $i \in T$, $j \in Q$ a function $\phi_{i,j} : X \to \mathbb{R}_+$ by $\phi_{i,j}(u) = \frac{d(u,A_{ij})}{\vartheta(i)^{1/p}}$, and $\phi(u) = \bigoplus_{i=1}^t \bigoplus_{j=1}^q \phi_{i,j}(u)$

For any k, let $u,v \in X$ be such that $v \in B(u,r_k(u))$. Let $r_{s^i} = \max\{r_{s^i}(u), r_{s^i}(v)\}$ and set $\delta_i = r_{s^i} - r_{s^{i-1}}$. Let t' be the smallest such that $r_{s^{t'}} + r_{s^{t'-1}} \geq d(u,v)/4$. If $r_{s^{t'}} \geq d(u,v)/2$ set $r_{s^{t'}} = d(u,v)/2$. Notice that $t' \leq \lceil \log_s k \rceil$ and that $r_{s^{t'}} \geq d(u,v)/8$, hence $\sum_{i=1}^{t'} \delta_i = r_{s^{t'}} \geq d(u,v)/8$. As $v \in B(u,r_k(u))$ we have for all $1 \leq i \leq t'$ that $s^i \leq k$. By standard arguments it can shown that with constant probability for any such pair u,v and scale $i \in T$ there exists a subset $J = J(u,v,i) \subseteq Q$ such that $|J| \geq \frac{\log n}{16}$ and for any $j \in J : |d(u,A_{ij}) - d(v,A_{ij})| \geq \delta_i$.

$$\|\phi(u) - \phi(v)\|_p^p = \sum_{j=1}^q \sum_{i=1}^t \left|\frac{d(u,A_{ij}) - d(v,A_{ij})}{\vartheta(i)^{1/p}}\right|^p$$
$$\leq q \cdot d(u,v)^p \sum_{i=1}^t \frac{1}{\vartheta(i)} = O(q) \cdot d(u,v)^p$$

$$\|\phi(u) - \phi(v)\|_p^p = \sum_{j=1}^{q}\sum_{i=1}^{t}\left|\frac{d(u,A_{ij}) - d(v,A_{ij})}{\vartheta(i)^{1/p}}\right|^p$$

$$\geq \frac{\log n}{16\vartheta(\log_s k)}\sum_{i=1}^{t'}\delta_i^p$$

$$\geq \frac{\log n}{16(t')^{p-1}\vartheta(\log_s k)}\left(\sum_{i=1}^{t'}\delta_i\right)^p$$

$$\geq \frac{\log n}{16(t')^{p-1}\vartheta(\log_s k)}(d(u,v)/8)^p,$$

after appropriate scaling we get the claimed local distortion. $\square$

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. $\square$

We showed that the embedding has scaling local distortion, the strong version follows similarly.

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 the bound on the distortion in Theorem 3 can be upper bounded by $O\left(\frac{\log k}{p}(\log\log k)^{\frac{1+\epsilon}{p}}\right)$.

4.4 Lower dimension for 1-local distortion

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

PROOF SKETCH. 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 the outgoing degree of each node is one, each connected component in G has at most one cycle. Fix some component H and, let r_H be an arbitrary node of H, and if there is a cycle, let w_H be the farthest point on the cycle from r_H (breaking ties arbitrarily).

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. It can be checked that f is non-expansive and that the distortion of nearest neighbors is at most $\sqrt[p]{3}$ $\square$

5. PROBABILISTIC LOCAL EMBEDDING INTO ULTRAMETRICS

Probabilistic embedding of metrics into ultrametrics [7] has many applications in online and approximation algorithms. The basic theorem states that every metric space probabilistically embeds into an ultrametrics with $O(\log n)$ distortion [8, 18, 9]. Here we extend this result to local embeddings.

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 $O(\log k)$.*

We defer the proof to the full version.

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 $O(\log n)$.*

We recall the following lemma implicitly proved in [15, 18] (a similar lemma appears in [2]).

LEMMA 3. *Given a finite metric space (X,d) and $0 < \Delta \leq \text{diam}(X)$, there exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of X such that for any $x \in X$ and any $0 < \eta \leq 1/8$:*

$$\Pr[B(x,\eta\Delta) \subseteq P(x)] \geq 1 - \eta\log(\rho(x,\Delta,8,1)).$$

Let $\Delta_0 = \text{diam}(X)$ and for any integer $i > 0$ let $\Delta_i = \Delta_0 2^{-i}$. For all $i > 0$ create a Δ_i-bounded probabilistic partition $\hat{\mathcal{P}}_i$ as in Lemma 3. Since every cluster $C \in \mathcal{P}_i$ in the partition is created by a ball with radius in the range $[\Delta_i/4, \Delta_i/2]$, around a center point c, assume w.l.o.g that $v(C) = c$, and define for each $P_i \in \mathcal{P}_i$ and any cluster $C \in P_i$

$$b(C) = \log(|B(v(C), 2\Delta_i)|).$$

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 4. *For all $i > 0$, if $C \in P_i$, $D \in P_{i-1}$ and $C \cap D \neq \emptyset$ then $\alpha(D) \leq 2\alpha(C)$.*

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

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 $\mathcal{Q}^{(1)}, \ldots, \mathcal{Q}^{(\log\Delta_0)}$ 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 $\mathcal{Q}^{(i-1)} = \{Q_1^{(i-1)}, \ldots, Q_{i-1}^{(i-1)}\}$ and P_i we define a hierarchical partition $\mathcal{Q}^i = \{Q_1^{(i)}, \ldots, Q_i^{(i)}\}$ 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 labelled tree T from Q and its labels in the natural manner. 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 $\alpha(D) \geq \alpha(C)$.

CLAIM 5. *For any pair* (x, y), *such that* $y \in B(x, r_k(x))$:
$d_T(x, y) \geq \Omega\left(\frac{d(x,y)}{\vartheta(\log k)}\right)$.

PROOF. Let i be the smallest integer such that $3\Delta_i \leq d(x, y)$, it is immediate that (x, y) are separated in the partition P_i. Let $v = \upsilon(P_i(x))$, then

$$B(v, 2\Delta_i) \subseteq B(x, 3\Delta_i) \subseteq B(x, d(x,y)) \subseteq B(x, r_k(x)),$$

which implies that $b(P_i(x)) \leq \log k$, hence $\alpha(P_i(x)) \geq \frac{\Delta_i}{\vartheta(\log k)} \geq \Omega\left(\frac{d(x,y)}{\vartheta(\log k)}\right)$. Now we claim that $\alpha(Q_i(x)) \geq \alpha(P_i(x))$, which holds since if 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 smaller than the radius of $P_i(x)$ therefore it cannot contain y. This implies the claim. $\square$

CLAIM 6. *For any pair* (x, y): $\mathbb{E}[d_T(x, y)] \leq O(d(x, y))$.

PROOF. Fix some partition $P \in \mathcal{P}$. For every $x, y \in X$, let $\alpha_i(x, y) = \max\{\alpha(P_i(x)), \alpha(P_i(y))\}$. Define the events

$$\mathcal{C}_i(x, y) = \{P_i(x) \neq P_i(y)\},$$
$$\mathcal{M}_i(x, y) = \left\{\mathcal{C}_i(x, y) \wedge \bigwedge_{j > i} \alpha_i(x, y) \geq \alpha_j(x, y)\right\}.$$

Notice that if for some scale i event $\mathcal{C}_i(x, y)$ holds but $\mathcal{M}_i(x, y)$ does not then consider the scale $j > i$ maximizing $\alpha_j(x, y)$ (w.l.o.g $\alpha_j(x, y) = \alpha(P_j(x))$). Then some part of $P_i(x)$ that contains x will be replaced by some part of $P_j(x)$ containing x, and the event that $Q_i(x) \neq Q_i(y)$ will depend only on the event $P_j(x) \neq P_j(y)$. So when considering the sum over all scales we need not take into account scale i since we already accounted for scale j. Let $b_i = b_i(x) = \log|B(x, \Delta_i)|$, b_i is monotonic decreasing with i. Note that we can bound $\mathcal{C}_i(x, y)$ by the probability that either $B(x, d(x, y))$ or $B(y, d(x, y))$ is cut (using Lemma 3), hence if $\alpha_i(x, y) = \alpha(P_i(x))$ we will take $B(x, d(x, y))$ and vice versa. this implies that we can assume w.l.o.g that $\alpha_i(x, y) = \alpha(P_i(x))$. For all $i > 0$ since $d(x, \upsilon(P_i(x))) \leq \Delta_i$ it follows that $b(P_i(x)) \geq b_i$. Let ℓ be the largest integer such that $\Delta_\ell > 8d(x, y)$.

$$\mathbb{E}_T[d_T(x, y)]$$
$$\leq \sum_{i > 0} \Pr[\mathcal{M}_i(x, y)] \mathbb{E}[\alpha_{i-1}(x, y) \mid \mathcal{M}_i(x, y)]$$
$$\leq \sum_{i > 0} \Pr[\mathcal{C}_i(x, y)] 2\mathbb{E}[\alpha_i(x, y) \mid \mathcal{M}_i(x, y)]$$
$$\leq 4d(x, y) \sum_{i=1}^{\ell} \frac{b_i - b_{i+3}}{\Delta_i} \mathbb{E}\left[\frac{\Delta_i}{\vartheta(b(P_i(x)))} \mid \mathcal{M}_i(x, y)\right] + 2\sum_{i > \ell} \Delta_i$$
$$\leq 4d(x, y) \sum_{i=1}^{\ell} \sum_{j=b_{i+3}+1}^{b_i} 1/\vartheta(b_i) + 2\Delta_\ell$$
$$\leq 2^6 d(x, y) \sum_{j > 0} 1/\vartheta(j) = O(d(x, y)).$$

The second inequality follows from Claim 4, the third inequality follows from Lemma 3. This concludes the proof of scaling local distortion $O(\vartheta(\log k))$.

By using $\bar{\vartheta}(x) = \min\{\vartheta(x), \log n\}$ we can show that the worst case distortion can be bounded by $O(\log n)$. $\square$

We showed that the embedding has scaling local distortion, the strong version follows similarly.

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 [5, 17]. In [4] 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). 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 metric space* (X, d) *derived from a graph* G *such that any embedding into a distribution of spanning trees of* G *will incur* 1-*local distortion of* $\Omega(\log n)$.

6. EMBEDDING INTO ℓ_P WITH K-LOCAL DISTORTION AND LOW DIMENSION

To achieve low dimension for k-local embeddings into ℓ_p we use local probabilistic partitions, a method of embedding based on uniform probabilistic partitions [2], and the Lovász Local Lemma. In the full version of the paper we give an algorithmic version of this theorem.

THEOREM 8. *For any finite metric space* (X, d) *on* n *points with a* χ *very weak growth bound there exists an embedding into* ℓ_p *with* k-*local distortion* $O(\log k)$ [2] *and dimension* $O(\log k)$.

Let $D = c' \ln k$, for some constant c' to be defined later. Assume that $k \geq 2$ and $\chi \geq 2$. We will define an embedding $f : X \to \ell_p^D$ with k-local distortion $O(\log k)$. We define f by defining for each $1 \leq t \leq D$, a function $f^{(t)} : X \to \mathbb{R}_+$, and letting $f = D^{-1/p} \bigoplus_{1 \leq t \leq D} f^{(t)}$. Fix t, $1 \leq t \leq D$. In what follows we define $f^{(t)}$. Let $\Delta_0 = \text{diam}(X)$, let $I = \{i \in \mathbb{N} \mid i \in [1, \log_8 \Delta_0]\}$ and for every $i \in I$ let $\Delta_i = \Delta_0 8^{-i}$. We construct for all $i \in I$ a uniformly Δ_i-bounded η_i-padded probabilistic partition $\hat{\mathcal{P}}_i$ as in Lemma 2 with parameter $\Gamma = 8^3$, and let ξ_i be as defined in the lemma. Since every cluster $C \in \mathcal{P}_i$ in the partition is created by a ball with radius in the range $[\Delta_i/4, \Delta_i/2]$, around a center point c, assume w.l.o.g that $\upsilon(C) = c$, Denote by Ω the probability space of all possible embeddings f. Now for every $i \in I$ fix a partition $P_i \in \mathcal{P}_i$. We define the embedding by defining the coordinates for each $x \in X$. Let $\ell(x) \in I$ be the minimal such that $|B(\upsilon(P_{\ell(x)}(x)), 9\Gamma\Delta_{\ell(x)})| \leq k^{\chi^7}$.

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

$$\bar{\xi}_{P,i}(x) = \left\{\begin{array}{cc} 0 & i < \ell(x) \\ \xi_{P,i}(x) & \text{otherwise} \end{array}\right. .$$

Define for $x \in X$, $0 < i \in I$, $\phi_i^{(t)} : X \to \mathbb{R}_+$, by $\phi_i^{(t)}(x) = \bar{\xi}_{P,i}(x)\eta_{P,i}(x)^{-1}$. Lemma 2 and the definition of $\ell(x)$ ensures that $\bar{\xi}_i$ and η_i are uniform functions with respect to $\mathcal{P}_i$. By the properties of Lemma 2 and the definition of $\ell(x)$, $\bar{\xi}_{P,i}(x)$ it can be shown that:

---

[2] In fact we show this bound for all $u, v \in X$ such that $v \in B(u, r_k(u))$

CLAIM 7. *For any $x \in X$, $t \in [D]$,*

$$\sum_{i \in I} \phi_i^{(t)}(x) \le 2^{10} \chi^7 \log k.$$

For each $0 < i \in I$ we define a function $f_i^{(t)} : X \to \mathbb{R}_+$ and for $x \in X$, let $f^{(t)}(x) = \sum_{i \in I} f_i^{(t)}(x)$. 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 $f_i^{(t)}(x) = \sigma_i^{(t)}(P_i(x)) \cdot \min\{\phi_i^{(t)}(x) \cdot d(x, X \setminus P_i(x)), \Delta_i\}$.

The following claim was proved in [2]

CLAIM 8. *For any $x, y \in X$ and $0 < i \in I$: $f_i^{(t)}(x) - f_i^{(t)}(y) \le \min\{\phi_i^{(t)}(x) \cdot d(x,y), \Delta_i\}$.*

From Claims 7 and 8 we get

LEMMA 9. *There exists a universal constant $C_1 = C_1(\chi) > 0$ such that for any $x, y \in X$:*

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

6.1 Lower bound analysis

Let $N_i = \{(x,y) \mid x, y \in X \wedge y \in B(x, r_k(x)) \wedge d(x,y) \in [\Delta_{i-2}, \Delta_{i-3}]\}$ and let $N = \bigcup_{i>0} N_i$. For all $i > 0$, $x, y \in N_i$ and $t \in [D]$ let $\mathcal{Z}_{(i,x,y,t)}$ be the event that

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

For any embedding f and $i \in I$ define a function $g_i : N_i \to 2^D$ as follows

$$g_i(x,y) = \left\{t \in [D] \mid \mathcal{Z}_{(i,x,y,t)}\right\},$$

and let $\mathcal{Z}_{(i,x,y)}$ be the event that $|g_i(x,y)| \ge D/16$. Notice that for both events $\mathcal{Z}_{(i,x,y)}, \mathcal{Z}_{(i,x,y,t)}$ we could have omitted the index i since it can be inferred uniquely from $d(x,y)$, we explicitly write this index for ease of presentation. Let $\mathcal{Z} = \bigcap_{(x,y) \in N} \mathcal{Z}_{(i,x,y)}$, then

LEMMA 10. $\Pr[\mathcal{Z}] > 0$.

Before showing the proof of this Lemma, let us see that it is sufficient to show the lower bound.

CLAIM 11. *Fix $x, y \in N_i$ and $t \in g_i(x,y)$, then $|f^{(t)}(x) - f^{(t)}(y)| > d(x,y)/2^{11}$.*

PROOF. By Claim 8 and since Δ_j is a geometric series $|\sum_{j>i} f_j^{(t)}(x) - f_j^{(t)}(y)| \le \sum_{j>i} \Delta_j \le \Delta_i/4...$
Since $\mathcal{Z}_{(i,x,y,t)}$ holds $|\sum_{j \le i} f_j^{(t)}(x) - f_j^{(t)}(y)| \ge \Delta_i/2$ so $|\sum_{j>0} f_j^{(t)}(x) - f_j^{(t)}(y)| \ge \Delta_i/4 \ge d(x,y)/2^{11}$. $\square$

LEMMA 12. *There exists a universal constant C_2 and an embedding $f \in \Omega$ such that for any $x, y \in N$:*

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

PROOF. Using Lemma 10, let $f \in \Omega$ such that event $\mathcal{Z}$ took place. Consider any $(x,y) \in N_i$. Using Claim 11 and the fact that $|g_i(x,y)| \ge D/16$

$$
\begin{aligned}
\|f(x) - f(y)\|_p^p &= \frac{1}{D} \sum_{t \in D} |f^{(t)}(x) - f^{(t)}(y)| \\
&\ge \frac{1}{D} \sum_{t \in g_i(x,y)} |f^{(t)}(x) - f^{(t)}(y)| \\
&\ge (1/16)\frac{d(x,y)}{2^{11}}.
\end{aligned}
$$

$\square$

PROOF OF LEMMA 10. We shall make use the following variation of the Local Lemma due to Erdos and Lovász.

LEMMA 13 (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 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 $(\mathcal{A}_i, \mathcal{A}_j) \in E$ 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 $Q \subseteq \{j : (\mathcal{A}_i, \mathcal{A}_j) \notin E \wedge c(\mathcal{A}_i) \ge c(\mathcal{A}_j)\}$. If $ep(d+1) \le 1$, then*

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

Define a graph $G = (V, E)$ where $V = \{\mathcal{Z}_{(i,x,y)} \mid i > 0 \wedge (x,y) \in N_i\}$, define the ranking function as $c(\mathcal{Z}_{(i,x,y)}) = i$ and

$$\left(\mathcal{Z}_{(i,x,y)}, \mathcal{Z}_{(i',x',y')}\right) \in E \Leftrightarrow$$
$$d(x,x') \le 16r_k(x) \wedge c(\mathcal{Z}_{(i,x,y)}) \le c(\mathcal{Z}_{(i',x',y')}).$$

Note that the definition is not symmetric, G is a directed graph, and that the rating matches the requirements of Lemma 13.

CLAIM 14. *The out-degree of G is bounded by k^{χ^5}.*

PROOF. Fix any vertex $\mathcal{Z}_{(i,x,y)} \in V$. By the very weak growth bound condition $|B(x, 16r_k(x))| \le |B(x, r_k(x))|^{\chi^4} = k^{\chi^4}$, and for each $x' \in B(x, 16r_k(x))$ there are at most k possible values of y' such that $(x', y') \in N$, hence the out degree is bounded by k^{χ^5}. $\square$

CLAIM 15. *Let $(x,y) \in N_i$. If $(x', y') \in N_i$ such that $(\mathcal{Z}_{(i,x,y)}, \mathcal{Z}_{(i,x',y')}) \notin E$ then $d(\{x,y\}, \{x',y'\}) > 2\Delta_i$.*

PROOF. By definition of N_i both $d(x,y), d(x',y') < \Delta_{i-3}$. By definition of G, $d(x,x') > 16r_k(x)$. As $y \in r_k(x)$ we have $d(x,y) \le r_k(x)$, hence $d(x',y) \ge d(x,x') - d(x,y) \ge 15d(x,y) > 2\Delta_i$, and $d(y',y) \ge d(x',y) - d(y',x') \ge 15d(x,y) - \Delta_{i-3} \ge 15\Delta_{i-2} - 8\Delta_{i-2} > 2\Delta_i$. The proof for x is similar and easier. $\square$

We now show a claim about event $\mathcal{Z}_{(i,x,y,t)}$ which lies in the heart of using the local Lemma.

CLAIM 16. *For all $(x,y) \in N_i$:*

$$\Pr\left[\neg\mathcal{Z}_{(i,x,y,t)} \mid \bigwedge_{(i',x',y') \in Q} \mathcal{Z}_{(i',x',y')}\right] \le 7/8$$

for all $Q \subseteq \{(i', x', y') : (\mathcal{Z}_{(i,x,y)}, \mathcal{Z}_{(i',x',y')}) \notin E \wedge c(\mathcal{Z}_{(i,x,y)}) \ge c(\mathcal{Z}_{(i',x',y')})\}$.

PROOF. Fix $(x,y) \in N_i$, $t \in [D]$. For $i' < i$ and any $x',y' \in N_{i'}$ the events $\mathcal{Z}_{(i',x',y')}$ depend only on the first i' scales of the probabilistic partition, so the padding in scale i and choice of σ_i will be independent of these events.

Otherwise assume $i' = i$. For any $x',y' \in N_i$ such that $(\mathcal{Z}_{(i,x,y)}, \mathcal{Z}_{(i,x',y')}) \notin E$, by Claim 15 $d(\{x,y\},\{x',y'\}) > 2\Delta_i$. 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 we have that the padding in scale i for x,y is independent of the padding for x',y'.

Even though the event $\mathcal{Z}_{(i,x,y)}$ depend on scales $j < i$, we will show that there is probability at least $1/8$ to succeed no matter what partitions were created in scales $j < i$:

We claim that $|B(\upsilon(P_i(y)), 9\Gamma\Delta_i)| \leq k^{\chi^7}$. As $d(x,\upsilon(P_i(y))) \leq d(x,y) + \Delta_i \leq 2d(x,y)$, due the very weak growth bound assumption, $|B(\upsilon(P_i(y)), 9\Gamma\Delta_i)| \leq |B(x, 72\Delta_{i-2} + 2d(x,y))| \leq |B(x, 74d(x,y))| \leq |B(x, 2^7 r_k(x))| \leq |B(x, r_k(x))|^{\chi^7} = k^{\chi^7}$. The same argument holds for $|B(\upsilon(P_i(x)), 9\Gamma\Delta_i)|$. This suggests that $\bar{\xi}_i(y) = \xi_i(y)$ and $\bar{\xi}_i(x) = \xi_i(x)$.

By Claim 1 $\max\{\bar{\rho}(x,\Delta_{i-1},\Gamma), \bar{\rho}(y,\Delta_{i-1},\Gamma)\} \geq 2$. W.l.o.g $\bar{\rho}(x,\Delta_{i-1},\Gamma) \geq 2$ which suggests that $\bar{\xi}_i(x) = \xi_i(x) = 1$, hence $\phi_i^{(t)}(x) = \eta_{P,i}^{-1}(x)$.

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

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

By Lemma 2, 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 $|f_i^{(t)}(x) - f_i^{(t)}(y)| \geq \min\{\eta_{P,i}^{-1}(x) \cdot d(x, X \setminus P_i(x)), \Delta_i\} \geq \Delta_i$.

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 $\mathcal{Z}_{(x,y,t)}$ holds. $\square$

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

$$\Pr\left[\mathcal{B} < D/16\right] \leq k^{-\chi^6}.$$

Let $p = k^{-\chi^6}$. To conclude, using Claim 16

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

for all $Q \subseteq \{(i',x',y') \mid (\mathcal{Z}_{(i,x,y)}, \mathcal{Z}_{(i',x',y')}) \notin E \wedge c(\mathcal{Z}_{(i,x,y)}) \geq c(\mathcal{Z}_{(i',x',y')})\}$. Note that $ep(k^{\chi^5}+1) \leq 1$, and apply Lemma 13.

7. PARTITION BASED EMBEDDINGS INTO ℓ_P

For decomposable metrics, we improve the scaling local distortion. Using partition based embeddings [2, 10] we get the following:

THEOREM 9. *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 $O(2^p \log n)$.*

THEOREM 10. *For any metric space X 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 strong 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 $O(\log n \log \Phi)$, where Φ denotes the aspect ratio of X [3].*

8. LOCAL METRIC RAMSEY THEOREMS

In this section we extend the work of [12, 25, 11] to give metric Ramsey theorems with improved local guarantees. The main step is generating metric Ramsey partitions [25] with local guarantees. These are later used to get improved local distance oracles and approximate ranking.

The following lemma follows directly from the uniform padding lemma in [2], a different proof also appears in [25]:

LEMMA 17. *For any metric space (X,d), $\Delta > 0$ there exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of X such that for all $\eta \in (0, 1/C]$ and $x \in X$:*

$$\Pr[B(x,\eta\Delta) \subseteq P(x)] \geq \rho(x,\Delta,\gamma_1,\gamma_2)^{-C\eta},$$

for some constants γ_1, γ_2 and C.

DEFINITION 9. *Let (X,d) be a metrics space and $\vartheta \in \Xi$. Let P be a hierarchical partition of X, let t be a parameter.*

- *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 $2^i \leq r_k(x)$.*
- *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.*

The following Lemmas extends a similar lemma of [25], by giving better padding parameters depending on the locality. We have a k-local and scaling versions:

LEMMA 18. *For any finite metric space (X,d), $k \in \mathbb{N}$ and parameter $t > 1$, there exists a distribution on ultrametrics such that any point $x \in X$ is k-locally padded with parameter t with probability $k^{\Omega(-1/t)}$.*

LEMMA 19. *For any finite metric space (X,d) and parameter $t > 1$, there exists a distribution on ultrametrics such that any point $x \in X$ is completely locally padded with parameter t with probability $n^{\Omega(-1/t)}$.*

Using Lemma 18 and Lemma 19 we obtain the following metric Ramsey theorems:

THEOREM 11. *The following holds for any n point metric space X with weak growth bound χ:*

- *There exists a subspace Y of X of size $n \cdot k^{-O(\chi/t)}$ and an embedding f into an ultrametric U such f has strong k-local distortion t for pairs in Y.[4]*

---

[3]The dimension can be bounded by $O(\log^2 n)$ using a more involved argument.
[4]I.e., for every $x,y \in Y$, $d_U(f(x), f(y)) \geq \min\{d_Y(x,y), r_k(x)\}/t$ (where $r_k(x)$ is defined with respect to X).

- *There exists a subspace Y of X of size $n^{1-O(1/t)}$ and an embedding f into an ultrametric U such f has strong scaling local distortion $\min\{t, \chi \cdot \vartheta(\log k)\}$ for pairs in Y.[5]*

9. EMBEDDING WITH PROXIMITY DISTORTION

In this section we provide proximity distortion results result where the distortion is bounded in terms of the distance between the points (see Definition 2).

For decomposable metrics we have the following result.

THEOREM 12. *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 $O(\log n)$.*

We also show that for growth bounded metrics 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 $x \in X$, $r > 0$: $|B(x, 2r)| \leq 2^\chi |B(x, r)|$. Given $x, y \in X$ such that $d(x, y) \leq t$ we have $|B(x, t)| \leq t^\chi$. Hence we can translate local distortion results into proximity distortion as demonstrated in the following theorem:

CLAIM 20. *Let (X, d) be an χ-growth bounded metric, then there exists an embedding into ℓ_p with scaling proximity distortion $O(\vartheta(\chi \log t))$.*

10. APPLICATIONS

10.1 Small world model

Given a metric space (X, d) and a distribution Π on $\binom{X}{2}$, such that local pairs are given higher probability, then using our embedding techniques yields constant average distortion with respect to Π.

For example, for $\alpha > 0$, our embedding gives constant average distortion for any of Kleinberg's "small world" distributions $\Pi(x, y \mid x) = \frac{k^{-(1+\alpha)}}{\sum_{i=1}^{n} i^{-(1+\alpha)}}$.

LEMMA 21. *Let (X, d) be a metric space, and Π a probability distribution satisfying that given x, the conditional probability $\Pi(x, y \mid x)$ to choose y with $d(x, y) = r_k(x)$ is bounded by $\frac{1}{\vartheta(k) \cdot \vartheta(\log k)}$. Then there exists an embedding f into ℓ_p or a distribution over ultrametrics with $\mathrm{avgdist}^{(\Pi)}(f) = \mathbb{E}_{(x,y) \sim \Pi} \left[\frac{d(x,y)}{\|f(x)-f(y)\|_p} \right] = O(1)$.*

PROOF. Let $f : X \to \ell_p$ be a scaling local embedding with distortion $c \cdot \vartheta(\log k)$.

$\mathrm{avgdist}^{(\Pi)}(f)$

$$= \sum_{x \in X} \Pi(x, \cdot) \sum_{y \in X} \Pi(x, y \mid x) \frac{d(x, y)}{\|f(x) - f(y)\|_p}$$

$$\leq \sum_{x \in X} \Pi(x, \cdot) \sum_{k=1}^{n} \sum_{y \in X} \Pi(x, y \mid x \wedge d(x, y) = r_k(x)) c\vartheta(\log k)$$

$$\leq c \sum_{k=1}^{n} \frac{\vartheta(\log k)}{\vartheta(k) \cdot \vartheta(\log k)} = c \sum_{k=1}^{n} 1/\vartheta(k) = O(1).$$

$\square$

LEMMA 22. *Let (X, d) be a metric space, and Π a probability distribution satisfying that given x, the conditional probability $\Pi(x, y \mid x)$ to choose y with $d(x, y) = r_k(x)$ is $\frac{Z}{d(x,y) \cdot \vartheta(k) \cdot \vartheta(\log k)}$, where Z is a scaling factor. Then there exists an embedding f into ℓ_p or a distribution over ultrametrics with $\mathrm{distavg}^{(\Pi)}(f) = \frac{\mathbb{E}_{(x,y) \sim \Pi}[d(x,y)]}{\mathbb{E}_{(x,y) \sim \Pi}[\|f(x)-f(y)\|_p]} = O(1)$.*

10.2 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, file allocation. 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 [7, 18] we can use the embedding given in Theorem 6. 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 kth 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 $\vartheta(\log k)$.

10.3 Local distance oracles

In [25, 11] metric Ramsey decompositions are used to obtain efficient proximity data structures. Using our local metric Ramsey partitions with the same data structure as [25] we can get local variations on distance oracles under a *weak growth bound* assumption.

CLAIM 23. *Let (X, d) be a metric space with χ weak growth bound, fix some $t > 1$, $x \in X$ and let T be an ultra metric in which x is completely locally padded with parameter t, then for all $y \in X$, if $k = k(y) \in \mathbb{N}$ is such that $d(x, y) = r_k(x)$ then $d(x, y) \leq d_T(x, y) \leq O(d(x, y)\vartheta(\log k))$.*

If the metric space does not have the weak growth bound, we can use our embedding into Euclidean space as distance oracles. The proofs follow from our results on local embedding and partitions, we omit the details.

THEOREM 13. *For any finite metric space there exists the following type of distance oracles:*

1. *For a fixed k: $O(\log k)$ strong local stretch[6] , $O(\log n)$ query time, and $O(n \log n)$ memory.*
2. *Scaling: $O(\vartheta(\log k))$ strong scaling local stretch[7], $O(\log n)$ query time, $O(n \log n)$ memory.*
3. *Scaling: k strong scaling local stretch, $O(1)$ query time, $O(n)$ memory.*

If the metric space has a χ weak growth bound then for any $t > 1$ there exists a distance oracle as follows

---

[5]I.e., for every $x, y \in Y$, and $k \in \mathbb{N}$, if $y \notin B(x, r_k(x))$ then $d_U(f(x), f(y)) \geq r_k(x) / \min\{t, \chi \cdot \vartheta(\log k)\}$ (where $r_k(x)$ is defined with respect to X).

[6]I.e., for any $y \in B(x, r_k(x))$ the stretch is $O(\log k)$ and for any $y \notin B(x, r_k(x))$ the oracles returns a value $\Omega(r_k(x)/\log k)$

[7]I.e., for any $k \in \mathbb{N}$ and $y \notin B(x, r_k(x))$ the oracles returns a value $\Omega(r_k(x)/\log k)$

4. *For a fixed k: $O(t)$ strong local stretch for, $O(1)$ query time, $O(n \cdot k^{\chi/t})$ memory.*

5. *Scaling:* $\min\{O(\chi \cdot \vartheta(\log k)), O(t)\}$ *strong scaling local stretch,* $O(1)$ *query time,* $O(n^{1+1/t})$ *memory.*

10.4 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 $d(x, y) \leq d(x, z)$. An s-approximate ranking is relaxing the last condition to $d(x, y) \leq s \cdot d(x, z)$. It is shown in [25] via a probabilistic construction and a deterministic construction in [11] that for any $s > 1$ there exist a data structure with $O(s)$-approximate ranking which can be pre-processed in $O(sn^{2+1/s} \log n)$ time, uses $O(sn^{1+1/s})$ space, and support queries for 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 according to the locality of the query points. We provide a theorem similar to Theorem 13 for approximate ranking. The complete details appear in the full version of the paper.

11. 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. The is a small gap between our scaling local distortion upper bounds (such as in theorems 3,6) . Is the very weak growth bound assumption in Theorem 8 necessary? Are the weak growth bound assumptions of theorem 13 necessary? Is there a k-local analogue to the Johnson-Lidenstrauss lemma: does every finite metric space in ℓ_2 have a k-local embedding into ℓ_2^d with $(1+\epsilon)$ distortion, where $d = O(\log k/\epsilon^2)$?

12. REFERENCES

[1] Ittai Abraham, Yair Bartal, Hubert T.-H. Chan, Kedar Dhamdhere, Anupam Gupta, Jon M. Kleinberg, Ofer Neiman, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. In *FOCS*, pages 83–100, 2005.

[2] 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.

[3] Ittai Abraham, Yair Bartal, and Ofer Neiman. Embedding metric spaces in their intrinsic dimension. Manuscript, 2007.

[4] Ittai Abraham, Yair Bartal, and Ofer Neiman. Embedding metrics into ultrametrics and graphs into spanning trees with constant average distortion. In *SODA '07 Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007.

[5] 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.

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

[7] Yair Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science* , pages 184–193, Burlington, VT, 1996.

[8] 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.

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

[10] Yair Bartal. On embedding finite metric spaces in low-dimensional normed spaces. Leibnitz Tech. Report., Hebrew University, 2005.

[11] Yair Bartal. Metric Ramsey decompositions and their applications, 2007. Manuscript.

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

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

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

[15] Gruia Calinescu, Howard J. Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *Symposium on Discrete Algorithms*, pages 8–16, 2001.

[16] 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.

[17] 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.

[18] 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, 2003.

[19] 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.

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

[21] 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.

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

[23] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Science*, 102:11623–11628, August 2005.

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

[25] Manor Mendel and Assaf Naor. Ramsey partitions and proximity data structures. In *FOCS '06: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 109–118, Washington, DC, USA, 2006.

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

[27] M. Thorup and U. Zwick. Approximate Distance Oracles. In STOC '01: Proceedings of the 33rd annual ACM symposium on Theory of computing, pages 183-192, Hersonissos, Crete, Greece, July 2001.

[28] 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.