Advances in Metric Embedding Theory

Ittai Abraham *
Microsoft Research

Yair Bartal†
Hebrew University

Ofer Neiman‡
Princeton University and CCI

Abstract

Metric Embedding plays an important role in a vast range of application areas such as computer vision, computational biology, machine learning, networking, statistics, and mathematical psychology, to name a few. The mathematical theory of metric embedding is well studied in both pure and applied analysis and has more recently been the a source of interest for computer scientists as well. Most of this work is focused on the development of bi-Lipschitz mappings between metric spaces. In this paper we present new concepts in metric embeddings as well as new embedding methods for metric spaces. We focus on finite metric spaces, however some of the concepts and methods are applicable in other settings as well.

One of the main cornerstones in finite metric embedding theory is a celebrated theorem of Bourgain which states that every finite metric space on n points embeds in Euclidean space with $O(\log n)$ distortion. Bourgain's result is best possible when considering the worst case distortion over all pairs of points in the metric space. Yet, it is natural to ask: can an embedding can do much better in terms of the *average distortion*? Indeed, in most practical applications of metric embedding the main criteria for the quality of an embedding is its *average* distortion over all pairs.

In this paper we provide an embedding with *constant* average distortion for arbitrary metric spaces, while maintaining the same worst case bound provided by Bourgain's theorem. In fact, our embedding possesses a much stronger property. We define the ℓ_q*-distortion* of a uniformly distributed pair of points. Our embedding achieves the best possible ℓ_q-distortion for all $1 \leq q \leq \infty$ *simultaneously.*

The results are based on novel embedding methods which improve on previous methods in another important aspect: the *dimension* of the host space. The dimension of an embedding is of very high importance in particular in applications and much effort has been invested in analyzing it. However, no previous result improved the bound on the dimension which can be derived from Bourgain's embedding. Our embedding methods achieve better dimension, and in fact, shed new light on another fundamental question in metric embedding, which is: whether the *embedding dimension* of a metric space is related to its *intrinsic dimension* ? I.e., whether the dimension in which it can be embedded in some real normed space is related to the intrinsic dimension which is reflected by the inherent geometry of the space, measured by the space's

*Email: ittaia@microsoft.com. Most of the research on this paper was carried out while the author was a student at the Hebrew University.

†Email: yair@cs.huji.ac.il. Most of the research was done while the author was supported in part by a grant from the Israeli Science Foundation (195/02) and part of the research was done while the author was at the Center of the Mathematics of Information, Caltech, CA, USA, and was supported in part by a grant from the National Science Foundation (NSF CCF-0652536).

‡Email: oneiman@princeton.edu. Partially funded by NSF Expeditions in Computing award, 2009-2010. 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).

doubling dimension. The existence of such an embedding was conjectured by Assouad[1]. and was later posed as an open problem in several papers. Our embeddings give the first positive result of this type showing any finite metric space obtains a low distortion (and constant *average distortion*) embedding in Euclidean space in dimension proportional to its doubling dimension.

Underlying our results is a novel embedding method. Probabilistic metric decomposition techniques have played a central role in the field of finite metric embedding in recent years. Here we introduce a novel notion of probabilistic metric decompositions which comes particularly natural in the context of embedding. Our new methodology provides a *unified approach* to all known results on embedding of arbitrary finite metric spaces. Moreover, as described above, with some additional ideas they allow to get far stronger results.

The results presented in this paper[2] have been the basis for further developments both within the field of metric embedding and in other areas such as graph theory, distributed computing and algorithms. We present a comprehensive study of the notions and concepts introduced here and provide additional extensions, related results and some examples of algorithmic applications.

Contents

[1]Assouad [Ass83] conjectured that the distortion would be a function only of the doubling dimension as well, but this stronger conjecture was disproved [Sem96].

[2]This paper is a full version based on the conference papers: [ABC+05, ABN06, ABN08]

1 Introduction

The theory of embeddings of finite metric spaces has attracted much attention in recent decades by several communities: mathematicians, researchers in theoretical Computer Science as well as researchers in the networking community and other applied fields of Computer Science.

The main objective of the field is to find *low distortion* embeddings of metric spaces into other more simple and structured spaces.

Given two metric spaces (X, d_X) and (Y, d_Y) an *injective* mapping $f : X \to Y$ is called an *embedding* of X into Y. An embedding is *non-contractive* if for every $u \neq v \in X$: $d_Y(f(u), f(v)) \geq d_X(u, v)$. The *distortion* of a non-contractive embedding f is: $\text{dist}(f) = \sup_{u \neq v \in X} \text{dist}_f(u, v)$, where $\text{dist}_f(u, v) = \frac{d_Y(f(u), f(v))}{d_X(u,v)}$. Equivalently, the distortion of a non-contracting embedding is the infimum over values α such that f is α-Lipschitz.

We say that X embeds in Y with distortion α if there exists an embedding of X into Y with distortion α.

In Computer Science, embeddings of finite metric spaces have played an important role, in recent years, in the development of algorithms. More general practical use of embeddings can be found in a vast range of application areas including computer vision, computational biology, machine learning, networking, statistics, and mathematical psychology to name a few.

From a mathematical perspective embeddings of finite metric spaces into normed spaces are considered natural non-linear analogues to the local theory of Banach spaces. The most classic fundamental question is that of embedding metric spaces into Hilbert Space.

Major effort has been put into investigating embeddings into l_p normed spaces (see the surveys [Ind01, Lin02, IM04] and the book [Mat02] for an exposition of many of the known results). The main cornerstone of the field has been the following theorem by Bourgain [Bou85]:

Theorem 1 (Bourgain)**.** *For every n-point metric space there exists an embedding into Euclidean space with distortion $O(\log n)$.*

This theorem has been the basis on which the theory of embedding into finite metric spaces has been built. In [LLR94] it is shown that Bourgain's embedding provides an embedding into l_p with distortion $O(\log n)$, where the dimension of the l_p space is at most $O(\log^2 n)$. In this paper we improve this result in two ways: for any $1 \leq p \leq \infty$ we present an embedding with *average* distortion $O(1)$ into $O(\log n)$ dimensional l_p space, while maintaining $O(\log n)$ distortion.

1.1 On the Average Distortion of Metric Embeddings

The $O(\log n)$ distortion guaranteed by Bourgain's theorem is existentially tight. A nearly matching bound was already shown in Bourgain's paper and later Linial, London and Rabinovich [LLR94] proved that embedding the metrics of constant-degree expander graphs into Euclidean space requires $\Omega(\log n)$ distortion.

Yet, this lower bound on the distortion is a *worst case* bound, i.e., it means that there *exists* a pair of points whose distortion is large. However, the *average case* is often more significant in terms of evaluating the quality of the embedding, in particular in relation to practical applications. Formally, the *average distortion* of an embedding f is defined as: $\mathrm{avgdist}(f) = \frac{1}{\binom{n}{2}} \sum_{u \neq v \in X} \mathrm{dist}_f(u,v)$. See Section 1.6 for discussion on other related notions.

Indeed, in most real-world applications of metric embeddings *average distortion* and similar notions are used for evaluating the embedding's performance in practice, for example see [HS03, HFC00, AS03, HBK+03, ST04, TC04]. Moreover, in some cases it is desired that the average distortion would be small and the worst case distortion would still be reasonably bounded as well. While these papers provide some indication that such embeddings are possible in practice, the classic theory of metric embedding fails to address this natural question.

In particular, applying Bourgain's embedding to the metric of a constant-degree expander graph results in $\Omega(\log n)$ distortion for a *constant fraction* of the pairs[3].

In this paper we prove the following theorem which provides a qualitative strengthening of Bourgain's theorem:

Theorem 2 (Average Distortion)**.** *For every n-point metric space there exists an embedding into $O(\log n)$ dimensional Euclidean space with distortion $O(\log n)$ and average distortion $O(1)$.*

In fact our results are even stronger. For $1 \leq q \leq \infty$, define the *ℓ_q-distortion* of an embedding f as:

$$\mathrm{dist}_q(f) = \|\mathrm{dist}_f(u,v)\|_q^{(\mathcal{U})} = \mathbb{E}[\mathrm{dist}_f(u,v)^q]^{1/q},$$

where $\|\cdot\|_q^{(\mathcal{U})}$ denotes the *normalized* q norm over the distribution $(\mathcal{U})$, defined as in the equation above, for $q < \infty$, where the expectation is taken according to the uniform distribution $\mathcal{U}$ over $\binom{X}{2}$. For $q = \infty$ we have: $\mathrm{dist}_\infty(f) = \|\mathrm{dist}_f(u,v)\|_\infty^{(\mathcal{U})} = \max_{u,v \in X} \mathrm{dist}_f(u,v)$. The classic notion of distortion is expressed by the ℓ_∞-distortion and the average distortion is expressed by the ℓ_1-distortion. Theorem 2 follows from the following:

Theorem 3 (ℓ_q-Distortion)**.** *For every n-point metric space (X,d) there exists an embedding f of X into $O(\log n)$ dimensional Euclidean space such that for any $1 \leq q \leq \infty$, $\mathrm{dist}_q(f) = O(\min\{q, \log n\})$.*

Both of these theorems follow from Theorem 10, which is proven in Section 4.

A variant of average distortion that is natural is what we call *distortion of average*: $\mathrm{distavg}(f) = \frac{\sum_{u \neq v \in X} d_Y(f(u),f(v))}{\sum_{u \neq v \in X} d(u,v)}$, which can be naturally extended to its ℓ_q-normed extension termed *distortion of ℓ_q-norm*. Theorems 2 and 3 extend to those notions as well.

Besides $q = \infty$ and $q = 1$, the case of $q = 2$ provides a particularly natural measure. It is closely related to the notion of *stress* which is a standard measure in *multidimensional scaling* methods, invented by Kruskal [Kru64] and later studied in many models and variants. Multidimensional

[3]Similar statements hold for the more recent metric embeddings of [Rao99, KLMN04] as well.

scaling methods (see [KW78, HS03]) are based on embedding of a metric representing the relations between entities into low dimensional space to allow feature extraction and are often used for indexing, clustering, nearest neighbor searching and visualization in many application areas [HFC00].

1.2 Low-Dimension Embeddings

Our new embeddings into l_p improve on the previous embedding methods by achieving optimal dimension.

Recall that Bourgain proved that every n point metric space embeds into l_p with $O(\log n)$ distortion. One of the most important parameters of an embedding into a normed space is the *dimension* of the embedding. This is of particular important in applications and has been the main object of study in the paper by Linial, London and Rabinovich [LLR94] In particular, they ask: what is the dimension of the embedding in Theorem 1 ?

For embedding into Euclidean space, this can be answered by applying the Johnson and Lindenstrauss [JL84] dimension reduction lemma which states that any n-point metric space in L_2 can be embedded in Euclidean space of dimension $O(\log n)$ with constant distortion. This reduces the dimension in Bourgain's theorem to $O(\log n)$.

However, dimension reduction techniques[4] *cannot* be used to generalize the low dimension bound to l_p for all p. In particular, while every metric space embeds isometrically in l_∞ there are super constant lower bounds on the distortion of embedding specific metric spaces into low dimensional l_∞ space [Mat96].

This problem has been addressed by Linial, London, and Rabinovich [LLR94] and separately by Matoušek [Mat90] where they observe that the embedding given in Bourgain's proof of Theorem 1 can be used to bound the dimension of the embedding into l_p by $O(\log^2 n)$.

In this paper we prove the following:

Theorem 4. *For any $1 \le p \le \infty$, every n-point metric space embeds in l_p with distortion $O(\log n)$ in dimension $O(\log n)$.*

The proof of Theorem 4 introduces new embedding techniques. In particular, the lower dimension is achieved due to a new technique of summing up the components of the embedding over all scales. This is in contrast to previous embeddings where such components were allocated separate coordinates. This allows us to create an embedding into a *single* dimension that preserves the distortion in expectation. This saves us the extra logarithmic factor in dimension, since logarithmic dimension suffices by a Chernoff-type argument.

Moreover, we show the following trade-off between distortion and dimension , which generalizes Theorem 4 :

Theorem 5. *For any $1 \le p \le \infty$, and integer $D \ge 1$, every n-point metric space embeds in l_p with distortion $O(n^{1/D} \log n)$ in dimension $O(D)$.*

In particular one can choose for any $\theta > 0$, $D = \frac{\log n}{\theta \log \log n}$ and obtain dimension $O(D)$ with almost optimal distortion of $O(\log^{1+\theta} n)$. The bounds in theorems 4 and 5 are tight for all values of n, p and D, as shown in Theorem 13 by examining the metric of an expander.

[4] For $1 \le p < 2$, a combination of lemmas of [JL84] and [JS82] (generalization of [FLM77]'s result for $p = 1$) can be used to obtain an embedding in dimension $O(\log n)$.

Matoušek extended Bourgain's proof to improve the distortion bound into l_p to $O(\lceil \frac{\log n}{p} \rceil)$. He also showed this bound is tight [Mat97]. The dimension obtained in Matoušek's analysis of the embedding into l_p is $e^{O(p)} \log^2 n$. Our methods extend to give the following improvement:

Theorem 6. *For any $1 \le p \le \infty$ and any $1 \le k \le p$, every n-point metric space embeds in l_p with distortion $O(\lceil \frac{\log n}{k} \rceil)$ in dimension $e^{O(k)} \log n$.*

The bound on the dimension in Theorem 6 is nearly tight (up to lower order terms) as follows from volume arguments by Matoušek [Mat96] (based on original methods of Bourgain [Bou85]).

Theorem 4 and Theorem 5 are proven in Section 4, in particular by Corollary 19, and the proof of Theorem 6 is implied by the proof of Theorem 10, which is proven in the same section.

1.3 Infinite Compact Spaces

It is well known that infinite metric spaces may require infinite distortion when embedded into Euclidean space, this is also implied by Bourgain's result - the distortion tends to infinity with the cardinality of (X, d). However, our bound on the average distortion (and in general the ℓ_q-distortion) does not depend on the size of (X, d), hence we can apply our embedding technique to infinite compact metric spaces as well.

For a compact metric space (X, d) equipped with a measure[5] σ we define the product distribution $\Pi = \Pi(\sigma)$ over $X \times X$ as $\Pi(x, y) = \sigma(x)\sigma(y)$. Define the ℓ_q-distortion of an embedding f for $1 \le q < \infty$ as:

$$\text{dist}_q(f) = \mathbb{E}_{(x,y)\sim\Pi}[\text{dist}_f(x, y)^q]^{1/q}.$$

Theorem 7. *For any $q \ge 1$, $p \ge 1$, any compact metric space (X, d) and any probability measure σ over X, there is a mapping $f : X \to l_p$ with $\text{dist}_q(f) = O(q)$, for every $1 \le q < \infty$.*

In particular the embedding has constant average distortion.

1.4 Intrinsic Dimension

Metric embedding has important applications in many practical fields. Finding compact and faithful representations of large and complex data sets is a major goal in fields like data mining, information retrieval and learning. Many real world measurements are of intrinsically low dimensional data that lie in extremely high dimensional space.

Given a metric space with high *intrinsic dimension*, there is an obvious lower bound of $\Omega(\log_\alpha n)$ on the dimension for embedding this metric space into Euclidean space with distortion α (see [Mat02]). The intrinsic dimension of a metric space X is naturally measured by the doubling constant of the space: the minimum λ such that every ball can be covered by λ balls of half the radius. The doubling dimension of X is defined as $\dim(X) = \log_2 \lambda$. The doubling dimension of a metric space is the minimal dimension in which a metric space can be embedded into a normed space in a sense that embedding into less dimensions may cause arbitrarily high distortion.

A fundamental question in the theory of metric embedding is the relationship between the *embedding dimension* of a metric space and its *intrinsic dimension* That is, whether the dimension in which it can be embedded in some real normed space is implied by the intrinsic dimension which is reflected by the inherent geometry of the space.

[5] We may assume w.l.o.g that σ is a probability measure

Variants of this question were posed by Assouad [Ass83] as well as by Linial, London and Rabinovich [LLR94], Gupta, Krauthgamer and Lee [GKL03], and mentioned in [Mat05]. Assouad [Ass83] proved that for any $0 < \gamma < 1$ there exist numbers $D = D(\lambda, \gamma)$ and $C = C(\lambda, \gamma)$ such that for any metric space (X, d) with $\dim(X) = \lambda$, its "snowflake" version (X, d^γ) can be embedded into a D-dimensional Euclidean space with distortion at most C. Assouad conjectured that similar results are possible for $\gamma = 1$, however this conjecture was disproved by Semmes [Sem96]. Gupta, Krauthgamer and Lee [GKL03] initiated a comprehensive study of embeddings of doubling metrics. They analyzed the Euclidean distortion of the Laakso graph, which has constant doubling dimension, and show a lower bound of $\Omega(\sqrt{\log n})$ on the distortion. They also show a matching upper bound on the distortion of embedding doubling metrics, more generally the distortion is $O(\log^{1/p} n)$ for embedding into l_p. The best dependency on $\dim(X)$ of the distortion for embedding doubling metrics is given by Krauthgamer *et. al.* [KLMN04]. They show an embedding into l_p with distortion $O((\dim(X))^{1-1/p}(\log n)^{1/p})$, and dimension $O(\log^2 n)$.

However, all known embeddings for general spaces [Bou85, Mat96, LLR94, ABN06], and even those that were tailored specifically for bounded doubling dimension spaces [GKL03, KLMN04] require $\Omega(\log n)$ dimensions. In this paper we give the first general low-distortion embeddings into a normed space whose dimension depends only on $\dim(X)$.

Theorem 8. *There exists a universal constant C such that for any n-point metric space (X, d) and any $C/\log\log n < \theta \le 1$, there exists an embedding $f : X \to l_p^D$ with distortion $O\left(\log^{1+\theta} n\right)$ where $D = O\left(\frac{\dim(X)}{\theta}\right)$.*

We present additional results in Section 1.12, including an embedding into $\tilde{O}(\dim(X))$ [6] dimensions with constant average distortion and an extension of Assouad's result.

1.5 Novel Embedding Methods

There are few general methods of embedding finite metric spaces that appear throughout the literature. One is indeed the method introduced in Bourgain's proof (which itself is based on a basic approach attributed to Fréchet). This may be described as a Fréchet-style embedding where coordinates are defined as distances to randomly chosen sets in the space. Some examples of its use include [Bou85, LLR94, Mat90, Mat97], essentially providing the best known bounds on embedding arbitrary metric spaces into l_p.

The other embedding method which has been extensively used in recent years, is based on *probabilistic partitions* of metric spaces [Bar96], originally defined in the context of *probabilistic embedding* of metric spaces. Probabilistic partitions for arbitrary metric spaces were also given in [Bar96] and similar constructions appeared in [LS91].

The probabilistic embedding of [Bar96] (and later improvements in [Bar98, FRT03, Bar04]) provide in particular embeddings into L_1 and serve as the first use of probabilistic partitions in the context of embeddings into normed spaces. A partition is simply a collection of disjoint clusters of points whose union cover the entire space, and probabilistic partition is a distribution over such collections.

A major step was done in a paper by Rao [Rao99] where he shows that a certain padding property of such partitions can be used to obtain embeddings into L_2. Informally, a probabilistic

[6] By $\tilde{O}(N)$ we mean $N \cdot \log^{O(1)} N$.

partition is *padded* if every ball of a certain radius depending on some *padding parameter* has a good chance of being contained in a cluster. Rao's embedding defines coordinates which may be described as the distance from a point to the edge of its cluster in the partition and the padding parameter provides a lower bound on this quantity (with some associated probability). While Rao's original proof was done in the context of embedding planar metrics, it has since been observed by many researchers that his methods are more general and in fact provide the first decomposition-based embedding into l_p, for $p > 1$. However, the resulting distortion bound still did not match those achievable by Bourgain's original techniques.

This gap has been recently closed by Krauthgamer et. al [KLMN04]. Their embedding method is based on the probabilistic partition of [FRT03], which in turn is based on an algorithm of [CKR01] and further improvements by [FHRT03]. In particular, the main property of the probabilistic partition of [FRT03] is that the padding parameter is defined separately at each point of the space and depends in a delicate fashion on the growth rate of the space in the local surrounding of that point.

This paper introduces *novel probabilistic partitions* with even more refined properties which allow stronger and more general results on embedding of finite metric spaces.

Probabilistic partitions were also shown to play a fundamental role in the Lipschitz extension problem [LN05]. Partition based embeddings also play a fundamental role in the recently developed metric Ramsey theory [BBM06, BLMN05c, MN06]. In [BLMN05a] it is shown that the standard Fréchet style embeddings do not allow similar results. One indication that our approach significantly differs from the previous embedding methods discussed above is that our new theorems crucially rely on the use of non-Fréchet embeddings.

The main idea is the construction of *uniformly padded* probabilistic partitions. That is the padding parameter is *uniform* over all points within a cluster. The key is that having this property allows partition-based embeddings to use the value of the padding parameter in the definition of the embedding in the most natural way. In particular, the most natural definition is to let a coordinate be the distance from a point to the edge of the cluster (as in [Rao99]) multiplied by the inverse of the padding parameter. This provides an alternate embedding method with essentially similar benefits as the approach of [KLMN04].

We present a construction of uniformly padded probabilistic partitions which still posses intricate properties similar to those of [FRT03]. The construction is mainly based on a decomposition lemma similar in spirit to a lemma which appeared in [Bar04], which by itself is a generalization of the original probabilistic partitions of [Bar96, LS91]. However the proof that the new construction obeys the desired properties is quite technically involved and requires several new ideas that have not previously appeared.

We also give constructions of *uniformly padded hierarchical probabilistic partitions*. The idea is that these partitions are padded in a hierarchical manner – a much stronger requirement than for only a single level partition. Although these are not strictly necessary for the proof of our main theorems they capture a *stronger* property of our partitions and play a central role in showing that arbitrary metric spaces embed in l_p with constant average distortion, while maintaining the best possible worst case distortion bounds. The embeddings in this paper demonstrate the versatility of these techniques and further applications that appeared subsequent to this work are discussed in Subsection 1.16.

1.6 Related Work

Average distortion. Related notions to the ones studied in this paper have been considered before in several theoretical papers. Most notably, Yuri Rabinovich [Rab03] studied the notion of distortion of average[7] motivated by its application to the Sparsest Cut problem. This however places the restriction that the embedding is Lipschitz or *non-expansive*. Other recent papers have address this version of distortion of average and its extension to weighted average. In particular, it has been recently shown (see for instance [FHL05]) that the work of Arora, Rao and Vazirani on Sparsest Cut [ARV04] can be rephrased as an embedding theorem using these notions.

In his paper, Rabinovich observes that for Lipschitz embeddings the lower bound of $\Omega(\log n)$ still holds. It is therefore *crucial* in our theorems that the embeddings are *co-Lipschitz*[8] (a notion defined by Gromov [Gro83]) (and w.l.o.g *non-contractive*).

To the best of our knowledge the only paper addressing such embeddings prior to this work is by Lee, Mendel and Naor [LMN05] where they seek to bound the *average distortion* of embedding n-point L_1 metrics into Euclidean space. However, even for this special case they do not give a constant bound on the average distortion[9].

Network embedding. Our work is largely motivated by a surge of interest in the networking community on performing *passive distance estimation* (see e.g. [FJJ+01, NZ02, LHC05, CDK+04, ST04, CCRK04]), assigning nodes with short labels in such a way that the network latency between nodes can be approximated efficiently by extracting information from the labels without the need to incur active network overhead. The motivation for such labeling schemes are many emerging large-scale decentralized applications that require *locality awareness*, the ability to know the relative distance between nodes. For example, in peer-to-peer networks, finding the nearest copy of a file may significantly reduce network load, or finding the nearest server in a distributed replicated application may improve response time. One promising approach for distance labeling is *network embedding* (see [CDK+04]). In this approach nodes are assigned coordinates in a low dimensional Euclidean space. The node coordinates form simple and efficient *distance labels*. Instead of repeatedly measuring the distance between nodes, these labels allow to extract an approximate measure of the latency between nodes. Hence these network coordinates can be used as an efficient building block for locality aware networks that significantly reduce network load.

In networking embedding, a natural measure of efficiency is the embedding performance on average. Where the notion of average distortion comes in several variations are possible in terms of the definitions given above. The phenomenon observed in measurements of network distances is that the average distortion of network embeddings was bounded by a small constant. Our work gives the *first* full theoretical explanation for this intriguing phenomenon.

Embedding with relaxed guaranties. The theoretical study of such phenomena was initiated by the work of Kleinberg, Slivkins and Wexler [KSW09]. They mainly focus on the fact reported in the networking papers that the distortion of almost all pairwise distances is bounded by some small constant. In an attempt to provide theoretical justification for such phenomena [KSW09] define the notion of a $(1-\epsilon)$-partial embedding[10] where the distortion is bounded for at least some $(1-\epsilon)$ fraction of the pairwise distances. They obtained some initial results for metrics

[7] Usually this notion was called average distortion but the name is somewhat confusing.

[8] We use the term *co-Lipschitz* to mean a function $f : X \to Y$ such that for all $u, v \in X$: $d_Y(f(u), f(v)) \geq c \cdot d_X(u, v)$ for some universal constant c. Up to scaling we can assume such a function to be non-contractive.

[9] The bound given in [LMN05] is $O(\sqrt{\log n})$ which applies to a somewhat weaker notion.

[10] Called "embeddings with ϵ-slack" in [KSW09].

which have constant doubling dimension [KSW09]. In Abraham et. al. [ABC$^+$05] it was shown that any finite metric space has a $(1-\epsilon)$-partial embedding into Euclidean space with $O(\log \frac{2}{\epsilon})$ distortion.

While this result is very appealing it has the disadvantage of lacking any promise for some fraction of the pairwise distances. This may be critical for applications - that is we really desire an embedding which in a sense does *"as well as possible"* for all distances. To define such an embedding [KSW09] suggested a stronger notion of *scaling distortion*[11]. An embedding has scaling distortion of $\alpha(\epsilon)$ if it provides this bound on the distortion of a $(1-\epsilon)$ fraction of the pairwise distances, *for any* ϵ. In [KSW09], such embeddings with $\alpha(\epsilon) = O(\log \frac{2}{\epsilon})$ were shown for metrics of bounded growth dimension, this was extended in [ABC$^+$05] to metrics of bounded doubling dimension. In addition [ABC$^+$05] give a rather simple probabilistic embedding with scaling distortion, implying an embedding into (high-dimensional) L_1 (see also Section 9 of this paper).

The most important question arising from the work of [KSW09, ABC$^+$05] is whether embeddings with small scaling distortion exist for embedding arbitrary metrics into Euclidean space. We give the following theorem[12] which lies at the heart of the proof o Theorem 3:

Theorem 9. *For every finite metric space (X,d), there exists an embedding of X into Euclidean space with scaling distortion $O(\log \frac{2}{\epsilon})$ and dimension $O(\log n)$.*

This theorem is proved by Corollary 19 in Section 4.

1.7 Additional Results and Applications

In addition to our main result, the paper contains several other contributions: we extend the results on average distortion to weighted averages. We show the bound is $O(\log \Phi)$ where Φ is the effective aspect ratio of the weight distribution.

Then we demonstrate some basic algorithmic applications of our theorems, mostly due to their extensions to general weighted averages. Among others is an application to *uncapacitated quadratic assignment* [PRW94, KT02]. We also extend our concepts to analyze Distance Oracles of Thorup and Zwick [TZ05] providing results with strong relation to the questions addressed by [KSW09]. We however feel that our current applications do not make full use of the strength of our theorems and techniques and it remains to be seen if such applications will arise.

In the next few sections of the introduction we formally define all notions we use or introduce in this paper and provide formal statements of our theorems.

1.8 Novel Notions of Distortion

Given two metric spaces (X, d_X) and (Y, d_Y) an *injective* mapping $f : X \to Y$ is called an *embedding* of X into Y. An embedding f is called c-co-Lipschitz [Gro83] if for any $u \neq v \in X$: $d_Y(f(u), f(v)) \geq c \cdot d_X(u,v)$ and *non-contractive* if $c = 1$. In the context of this paper we will restrict attention to co-Lipschitz embeddings, which due to scaling may be further restricted to *non-contractive* embeddings. This has no difference for the classic notion of distortion but has a crucial role for the results presented in this paper. We will elaborate more on this issue in the sequel.

[11] Called "gracefully degrading distortion" in [KSW09].

[12] In fact in this theorem the definition of scaling distortion is even stronger. This is explained in detail in the appropriate section.

For a non-contractive embedding f, define the distortion function of f, $\mathrm{dist}_f : \binom{X}{2} \to \mathbb{R}^+$, where for $u \neq v \in X$: $\mathrm{dist}_f(u,v) = \frac{d_Y(f(u),f(v))}{d_X(u,v)}$. The distortion of f is defined as $\mathrm{dist}(f) = \sup_{u \neq v \in X} \mathrm{dist}_f(u,v)$.

Definition 1 (ℓ_q-Distortion)**.** Given a distribution Π over $\binom{X}{2}$ define for $1 \leq q \leq \infty$ the ℓ_q-*distortion* of f with respect to Π:

$$\mathrm{dist}_q^{(\Pi)}(f) = \|\mathrm{dist}_f(u,v)\|_q^{(\Pi)} = \mathbb{E}_\Pi[\mathrm{dist}_f(u,v)^q]^{1/q},$$

where $\|\cdot\|_q^{(\Pi)}$ denotes the *normalized* q norm over the distribution (Π), defined as in the equation above for $q < \infty$. For $q = \infty$ we have: $\mathrm{dist}_\infty(f) = \|\mathrm{dist}_f(u,v)\|_\infty^{(\Pi)} = \sup_{\pi(u,v)\neq 0} \mathrm{dist}_f(u,v)$, where π denotes Π's probability function. Let $\mathcal{U}$ denote the uniform distribution over $\binom{X}{2}$. The ℓ_q-*distortion* of f is defined as: $\mathrm{dist}_q(f) = \mathrm{dist}_q^{(\mathcal{U})}(f)$.

In particular the classic distortion may be viewed as the ℓ_∞-distortion: $\mathrm{dist}(f) = \mathrm{dist}_\infty(f)$. An important special case of ℓ_q-distortion is when $q = 1$:

Definition 2 (Average Distortion)**.** Given a distribution Π over $\binom{X}{2}$ define the *average distortion* of f with respect to Π as: $\mathrm{avgdist}^{(\Pi)}(f) = \mathrm{dist}_1^{(\Pi)}(f)$, and the *average distortion* of f is given by: $\mathrm{avgdist}(f) = \mathrm{dist}_1(f)$.

Another natural notion is the following:

Definition 3 (Distortion of ℓ_q-Norm)**.** Given a distribution Π over $\binom{X}{2}$ define the *distortion of ℓ_q-norm* of f with respect to Π:

$$\mathrm{distnorm}_q^{(\Pi)}(f) = \frac{\mathbb{E}_\Pi[d_Y(f(u),f(v))^q]^{1/q}}{\mathbb{E}_\Pi[d_X(u,v)^q]^{1/q}},$$

for $q < \infty$, and for $q = \infty$ we have: $\mathrm{distnorm}_\infty(f) = \|\mathrm{distnorm}_f(u,v)\|_\infty^{(\Pi)} = \frac{\sup_{\pi(u,v)\neq 0} d_Y(f(u),f(v))}{\sup_{\pi(u,v)\neq 0} d_X(u,v)}$, where π denotes Π's probability function. Finally, let $\mathrm{distnorm}_q(f) = \mathrm{distnorm}_q^{(\mathcal{U})}(f)$.

Again, an important special case of distortion of ℓ_q-norm is when $q = 1$:

Definition 4 (Distortion of Average)**.** Given a distribution Π over $\binom{X}{2}$ define the *distortion of average* of f with respect to Π as: $\mathrm{distavg}^{(\Pi)}(f) = \mathrm{distnorm}_1^{(\Pi)}(f)$ and the *distortion of average* of f is given by: $\mathrm{distavg}(f) = \mathrm{distnorm}_1(f)$.

For simplicity of the presentation of our main results we use the following notation: $\mathrm{dist}_q^{*(\Pi)}(f) = \max\{\mathrm{dist}_q^{(\Pi)}(f), \mathrm{distnorm}_q^{(\Pi)}(f)\}$, $\mathrm{dist}_q^*(f) = \max\{\mathrm{dist}_q(f), \mathrm{distnorm}_q(f)\}$, and $\mathrm{avgdist}^*(f) = \max\{\mathrm{avgdist}(f), \mathrm{distavg}(f)\}$.

Definition 5. A probability distribution Π over $\binom{X}{2}$, with probability function $\pi : \binom{X}{2} \to [0,1]$, is called *non-degenerate* if for every $u \neq v \in X$: $\pi(u,v) > 0$. The *aspect ratio* of a non-degenerate probability distribution Π is defined as:

$$\Phi(\Pi) = \frac{\max_{u \neq v \in X} \pi(u,v)}{\min_{u \neq v \in X} \pi(u,v)}.$$

In particular $\Phi(\mathcal{U}) = 1$. If Π is *not* non-degenerate then $\Phi(\Pi) = \infty$.

For an *arbitrary* probability distribution Π over $\binom{X}{2}$, define its *effective aspect ratio* as:[13] $\hat{\Phi}(\Pi) = 2\min\{\Phi(\Pi), \binom{n}{2}\}$.

Theorem 10 (Embedding into l_p)**.** *Let (X, d) an n-point metric space, and let $1 \le p \le \infty$. There exists an embedding f of X into l_p of dimension $e^{O(p)} \log n$, such that for every $1 \le q \le \infty$, and any distribution Π over $\binom{X}{2}$: $\mathrm{dist}_q^{*(\Pi)}(f) = O(\lceil(\min\{q, \log n\} + \log \hat{\Phi}(\Pi))/p\rceil)$. In particular, $\mathrm{avgdist}^{*(\Pi)}(f) = O(\lceil \log \hat{\Phi}(\Pi)/p\rceil)$. Also: $\mathrm{dist}(f) = O(\lceil(\log n)/p\rceil)$, $\mathrm{dist}_q^*(f) = O(\lceil q/p\rceil)$ and $\mathrm{avgdist}^*(f) = O(1)$.*

Theorem 14, Lemma 2 and Theorem 15 show that all the bounds in the theorem above are tight.

The proof of Theorem 10 follows directly from results on embedding with scaling distortion, discussed in the next paragraph.

1.9 Partial Embedding and Scaling Distortion

Following [KSW09] we define:

Definition 6 (Partial Embedding)**.** Given two metric spaces (X, d_X) and (Y, d_Y), a *partial embedding* is a pair (f, G), where f is a non-contractive embedding of X into Y, and $G \subseteq \binom{X}{2}$. The distortion of (f, G) is defined as: $\mathrm{dist}(f, G) = \sup_{\{u,v\}\in G} \mathrm{dist}_f(u, v)$.

For $\epsilon \in (0, 1)$, a $(1-\epsilon)$*-partial embedding* is a partial embedding such that $|G| \ge (1-\epsilon)\binom{n}{2}$.[14]

Next, we would like to define a special type of $(1-\epsilon)$-partial embeddings. Let $\hat{G}(\epsilon) = \{\{x, y\} \in \binom{X}{2} \mid \min\{|B(x, d(x, y))|, |B(y, d(x, y))|\} \ge \epsilon n/2\}$. A *coarsely* $(1-\epsilon)$-partial embedding f is a partial embedding $(f, \hat{G}(\epsilon))$[15].

Definition 7 (Scaling Distortion)**.** Given two metric spaces (X, d_X) and (Y, d_Y) and a function $\alpha : (0, 1) \to \mathbb{R}^+$, we say that an embedding $f : X \to Y$ has *scaling distortion* α if for any $\epsilon \in (0, 1)$, there is some set $G(\epsilon)$ such that $(f, G(\epsilon))$ is a $(1-\epsilon)$-partial embedding with distortion at most $\alpha(\epsilon)$. We say that f has *coarsely* scaling distortion if for every ϵ, $G(\epsilon) = \hat{G}(\epsilon)$.

We can extend the notions of partial embeddings and scaling distortion to probabilistic embeddings. For simplicity we will restrict to coarsely partial embeddings.[16]

Definition 8 (Partial/Scaling Prob. Embedding)**.** Given (X, d_X) and a set of metric spaces $\mathcal{S}$, for $\epsilon \in (0, 1)$, a *coarsely* $(1-\epsilon)$*-partial probabilistic embedding* consists of a distribution $\hat{\mathcal{F}}$ over a set $\mathcal{F}$ of *coarsely* $(1-\epsilon)$-partial embeddings from X into $Y \in \mathcal{S}$. The distortion of $\hat{\mathcal{F}}$ is defined as: $\mathrm{dist}(\hat{\mathcal{F}}) = \sup_{\{u,v\}\in\hat{G}(\epsilon)} \mathbb{E}_{(f,\hat{G}(\epsilon))\sim\hat{\mathcal{F}}}[\mathrm{dist}_f(u, v)]$.

The notion of scaling distortion is extended to probabilistic embedding in the obvious way.

We observe the following relation between partial embedding, scaling distortion and the ℓ_q-distortion (we note that a similar relation holds in the other direction as well given by Lemma 2).

[13]The factor of 2 in the definition is placed solely for the sake of technical convenience.

[14]Note that the embedding is *strictly* partial only if $\epsilon \ge 1/\binom{n}{2}$.

[15]It is elementary to verify that indeed this defines a $(1-\epsilon)$-partial embedding. We also note that in most of the proofs we can use a max rather than min in the definition of $\hat{G}(\epsilon)$. However, this definition seems more natural and of more general applicability.

[16]Our upper bounds use this definition, while our lower bounds hold also for the non-coarsely case.

Lemma 1 (Scaling Distortion vs. ℓ_q-Distortion)**.** *Given an n-point metric space (X, d_X) and a metric space (Y, d_Y). If there exists an embedding $f : X \to Y$ with scaling distortion α then for any distribution Π over $\binom{X}{2}$:*[17]

$$\mathrm{dist}_q^{(\Pi)}(f) \leq \left(2 \int_{\frac{1}{2}\binom{n}{2}^{-1}\hat{\Phi}(\Pi)}^{1} \alpha(x\hat{\Phi}(\Pi)^{-1})^q dx \right)^{1/q} + \alpha(\hat{\Phi}(\Pi)^{-1}).$$

In the case of coarsely *scaling distortion this bound holds for* $\mathrm{dist}_q^{*(\Pi)}(f)$.

Combined with the following theorem we obtain Theorem 10. We note that when applying the lemma we use $\alpha(\epsilon) = O(\log \frac{2}{\epsilon})$ and the bounds in the theorem mentioned above follow from bounding the corresponding integral.

Theorem 11 (Scaling Distortion Theorem into l_p)**.** *Let $1 \leq p \leq \infty$. For any n-point metric space (X, d) there exists an embedding $f : X \to l_p$ with coarsely scaling distortion $O(\lceil(\log \frac{2}{\epsilon})/p\rceil)$ and dimension $e^{O(p)} \log n$.*

This theorem is proved in Section 4.3.

1.10 Infinite Compact Spaces

For embedding of infinite compact spaces we require slightly different definitions. Let (X, d) be a compact metric space, equipped with a probability measure σ (in compact space every measure is equivalent to a probability measure). Define the product distribution $\Pi = \Pi(\sigma)$ over $X \times X$ as $\Pi(x, y) = \sigma(x)\sigma(y)$. Now for $1 \leq q < \infty$, the ℓ_q-distortion of an embedding f will be defined with respect to Π

$$\mathrm{dist}_q(f) = \mathbb{E}_{(x,y)\sim\Pi}[\mathrm{dist}_f(x, y)^q]^{1/q}.$$

The definition of $\hat{G}(\epsilon)$ for coarse scaling embedding will become
$\hat{G}(\epsilon) = \left\{(x, y) \in \binom{X}{2} \mid \min\{\sigma(B(x, d(x, y))), \sigma(B(y, d(x, y)))\} \geq \epsilon/2\right\}$.

In order to prove Theorem 7 we again will show an embedding with scaling distortion.

Theorem 12 (ℓ_q-Distortion for Compact Spaces)**.** *Let $1 \leq p \leq \infty$ and let (X, d) be a compact metric space. There exists an embedding $f : X \to l_p$ having coarsely scaling distortion $O(\lceil(\log \frac{2}{\epsilon})\rceil)$. For any $1 \leq q < \infty$, the ℓ_q-distortion of this embedding is: $\mathrm{dist}_q(f) = O(q)$.*

1.11 Lower Bounds

In Section 11 We show that our results are tight. First we show that the distortion-dimension tradeoff of Theorem 5 is indeed tight.

Theorem 13. *For any $1 \leq p < \infty$ and any $\theta > 0$, if the metric of an n-node constant degree expander embeds into l_p with distortion $O(\log^{1+\theta} n)$ then the dimension of the embedding is $\Omega(\log n/\lceil \log(\min\{p, \log n\}) + \theta \log\log n\rceil)$.*

The following theorem shows that the bound on the weighted average distortion (and distortion of average) is tight as well.

[17] Assuming the integral is defined. We note that lemma is stated using the integral for presentation reasons.

Theorem 14. *For any $p \geq 1$ and any large enough $n \in \mathbb{N}$ there exists a metric space (X, d) on n points, and non-degenerate probability distributions Π, Π' on $\binom{X}{2}$ with $\Phi(\Pi) = n$ and $\Phi(\Pi') = n^2$, such that any embedding f of X into l_p will have $\mathrm{dist}_p^{(\Pi)}(f) \geq \Omega(\log(\Phi(\Pi))/p)$, and $\mathrm{distnorm}_p^{(\Pi')}(f) \geq \Omega(\log(\Phi(\Pi'))/p)$.*

The following simple Lemma gives a relation between lower bound on partial embedding and the ℓ_q distortion.

Lemma 2 (Partial Embedding vs. ℓ_q-Distortion)**.** *Let Y be a target metric space, let $\mathcal{X}$ be a family of metric spaces. If for any $\epsilon \in (0, 1)$, there is a lower bound of $\alpha(\epsilon)$ on the distortion of $(1 - \epsilon)$ partial embedding of metric spaces in $\mathcal{X}$ into Y, then for any $1 \leq q \leq \infty$, there is a lower bound of $\frac{1}{2}\alpha(2^{-q})$ on the ℓ_q-distortion of embedding metric spaces in $\mathcal{X}$ into Y.*

Finally we give a lower bound on partial embeddings. In order to describe the lower bound, we require the notion of *metric composition* introduced in [BLMN05c].

Definition 9. Let N be a metric space, assume we have a collection of disjoint metric spaces C_x associated with the elements x of N, and let $\mathcal{C} = \{C_x\}_{x \in N}$. The β-composition of N and $\mathcal{C}$, for $\beta \geq \frac{1}{2}$, denoted $M = \mathcal{C}_\beta[N]$, is a metric space on the disjoint union $\dot{\bigcup}_x C_x$. Distances in $\mathcal{C}$ are defined as follows: let $x, y \in N$ and $u \in C_x, v \in C_y$, then:

$$d_M(u, v) = \begin{cases} d_{C_x}(u, v) & x = y \\ \beta\gamma d_N(x, y) & x \neq y \end{cases}$$

where $\gamma = \frac{\max_{x \in N} \mathrm{diam}(C_x)}{\min_{u,v \in N} d_N(u,v)}$, guarantees that M is indeed a metric space.

Definition 10. Given a class $\mathcal{X}$ of metric spaces, we consider $comp_\beta(\mathcal{X})$, its closure under $\geq \beta$-composition.
$\mathcal{X}$ is called *nearly closed under composition* if for every $\delta > 0$ there exists some $\beta \geq 1/2$, such that for every $X \in comp_\beta(\mathcal{X})$ there is $\hat{X} \in \mathcal{X}$ and an embedding of X into $\hat{X}$ with distortion at most $1 + \delta$.

Among the families of metric spaces that are nearly closed under composition we find the following: tree metrics, any family of metrics that exclude a fixed minor (including planar metrics) and normed spaces. When the size of all the composed metrics C_x is equal, also doubling metrics are nearly closed under composition.

Theorem 15 (Partial Embedding Lower Bound)**.** *Let Y be a target metric space, let $\mathcal{X}$ be a family of metric spaces nearly closed under composition. If for any $k > 1$, there is $Z \in \mathcal{X}$ of size k such that any embedding of Z into Y has distortion at least $\alpha(k)$, then for all $n > 1$ and $\frac{1}{n} \leq \epsilon \leq 1$ there is a metric space $X \in \mathcal{X}$ on n points such that the distortion of any $(1 - \epsilon)$ partial embedding of X into Y is at least $\alpha\left(\lceil \frac{1}{4\sqrt{\epsilon}} \rceil\right)/2$.*

See Corollary 72 for some implication of this Theorem.

1.12 Intrinsic Dimension

The intrinsic dimension of a metric space is naturally measured by its doubling constant:

Definition 11. The doubling constant of a metric space (X,d) is the minimal λ such that for any $x \in X$ and $r > 0$ the ball $B(x,2r)$ can be covered by λ balls of radius r. The *doubling dimension* denoted by $\dim(X)$ is define as $\log_2 \lambda$.

The doubling dimension of a metric space (X,d) provides an inherent bound on the dimension in which the metric can be embedded into some normed space with small distortion. Specifically, a simple volume argument suggests that to embed X into L_2 with distortion α requires at least $\Omega(\dim(X)/\log\alpha)$ dimensions.

Our main theorem is Theorem 8 which states that for any $0 < \theta \leq 1$, every n-point metric space X embeds in l_p in dimension $O(\dim(X)/\theta)$ with distortion $O(\log^{1+\theta} n)$. In addition we have the following results.

We prove the following theorem which shows that Assouad's conjecture is true in the following practical sense: low dimensional data embed into *constant* dimensional space with *constant* average distortion:

Theorem 16. *For any λ-doubling metric space (X,d) there exists an embedding $f : X \to l_p^D$ with coarse scaling distortion $O\left(\log^{26}(\frac{1}{\epsilon})\right)$ where $D = O(\log\lambda \log\log\lambda)$.*

Obtaining bounds on the scaling distortion in a dimension which depends only on $\dim(X)$ is considerably more demanding. The technical difficulties are discussed in Section 6.2

We also show a theorem that strengthens Assouad's result [Ass83], regarding embedding of a "snowflake" of metrics with low doubling dimensions, that is, for a metric (X,d) embed (X,d^α) for some $0 < \alpha < 1$ with distortion and dimension that depend only on the doubling dimension of (X,d).

Theorem 17. *For any n point λ-doubling metric space (X,d), any $0 < \alpha < 1$, any $p \geq 1$, any $\theta \leq 1$ and any $2^{192/\theta} \leq k \leq \log\lambda$, there exists an embedding of (X,d^α) into l_p with distortion $O(k^{1+\theta}\lambda^{1/(pk)}/(1-\alpha))$ and dimension $O\left(\frac{\lambda^{1/k}\ln\lambda}{\alpha\theta} \cdot \left(1 - \frac{\log(1-\alpha)}{\log k}\right)\right)$.*

1.13 Additional Results

1.13.1 Decomposable Metrics

For metrics with a decomposability parameter τ (see Definition 18 for precise definition)[18] we obtain the following theorem, which is the scaling analogous of the main result of [KLMN04].

Theorem 18. *Let $1 \leq p \leq \infty$. For any n-point τ-decomposable metric space (X,d) there exists an embedding $f : X \to l_p$ with coarse scaling distortion $O(\min\{(1/\tau)^{1-1/p}(\log\frac{2}{\epsilon})^{1/p}, \log\frac{2}{\epsilon}\})$ and dimension $O(\log^2 n)$.*

1.13.2 Scaling Embedding into Trees

Definition 12. An ultrametric (X,d) is a metric space satisfying a strong form of the triangle inequality, for all $x,y,z \in X$, $d(x,z) \leq \max\{d(x,y), d(y,z)\}$. In particular, it is also a tree metric.

Theorem 19 (Scaling Probabilistic Embedding)**.** *For any n-point metric space (X,d) there exists a probabilistic embedding into a distribution over ultrametrics with coarse scaling distortion $O(\log\frac{2}{\epsilon})$.*

[18]In particular doubling metrics and planar metrics have constant decomposability parameter

Applying Lemma 1 to Theorem 19 we obtain:

Theorem 20. *Let (X,d) be an n-point metric space. There exists a probabilistic embedding $\hat{\mathcal{F}}$ of X into ultrametrics, such that for every $1 \le q \le \infty$, and any distribution Π over $\binom{X}{2}$:* $\mathrm{dist}_q^{*(\Pi)}(\hat{\mathcal{F}}) = O(\min\{q, \log n\} + \log \hat{\Phi}(\Pi))$.

For $q = 1$ and for a given fixed distribution the following theorem gives a deterministic version of Theorem 20, which follows from the method of [CCG+98] for finding a single ultrametric.

Theorem 21. *Given an arbitrary fixed distribution Π over $\binom{X}{2}$, for any finite metric space (X,d) there exist embeddings f, f' into ultrametrics, such that* $\mathrm{avgdist}^{(\Pi)}(f) = O(\log \hat{\Phi}(\Pi))$ *and* $\mathrm{distavg}^{(\Pi)}(f') = O(\log \hat{\Phi}(\Pi))$.

We note that complementary to these results in was shown in [ABN07a] that any metric space embeds in a single ultrametric with scaling distortion and as a consequence with constant average distortion (see more in Section 1.16).

1.13.3 Partial Embedding Results

Even though partial embeddings are inferior to embeddings with scaling distortion, in a sense that they guarantee distortion bound only on a fraction of pairs, they can be useful since the dimension of the embedding can be much lower. We show general theorems that convert any embedding into l_p into partial embedding, for subset-closed[19] families of metric spaces. For most of the known results for specific families of metric spaces we can use this general reduction to obtain partial embedding results where the role of n is replaced by $1/\epsilon$ in the distortion and in the dimension bounds. In particular, these theorems imply that for any $\epsilon > 0$ and $1 \le p \le \infty$ any metric space has a $(1-\epsilon)$ partial embedding into l_p^D space with distortion $O(\log(2/\epsilon))$ where $D = O(\log(2/\epsilon))$. We also present theorem providing a $(1-\epsilon)$ coarse partial embedding with comparable distortion bound with an overhead of $O(\log n)$ in the dimension[20]. See Section 10 for the specific theorems.

Similar results were obtained independently by [CDG+09].

1.14 Algorithmic Applications

We demonstrate some basic applications of our main theorems. We must stress however that our current applications do not use the full strength of these theorems. Most of our applications are based on the bound given on the *distortion of average* for general distributions of embeddings f into l_p and into ultrametrics with $\mathrm{distavg}^{(\Pi)}(f) = O(\log \hat{\Phi}(\Pi))$. In some of these applications it is crucial that the result holds for all such distributions Π. This is useful for problems which are defined with respect to weights $c(u,v)$ in a graph or in a metric space, where the solution involves minimizing the sum over distances weighted according to c. This is common for many optimization problem either as part of the objective function or alternatively it may come up in the linear programming relaxation of the problem. These weights can be normalized to define the distribution Π. Using this paradigm we obtain $O(\log \hat{\Phi}(c))$ approximation algorithms, improving on the general bound which depends on n in the case that $\hat{\Phi}(c)$ is small. This is the *first* result of this nature.

[19] A family of metrics $\mathcal{X}$ is subset-closed if for all $X \in \mathcal{X}$, any sub-metric $Y \subset X$ satisfies $Y \in \mathcal{X}$

[20] For coarse embeddings $\Omega(\log n)$ dimension is necessary.

We are able to obtain such results for the following group of problems: *general sparsest cut* [LR99, AR98, LLR94, ARV04, ALN05], *multicut* [GVY93], *minimum linear arrangement* [ENRS00, RR98b], *embedding in d-dimensional meshes* [ENRS00, Bar04], *multiple sequence alignment* [WLB+99] and *uncapacitated quadratic assignment* [PRW94, KT02].

We would like to emphasize that the notion of bounded weights is in particular natural in the last application mentioned above. The problem of *uncapacitated quadratic assignment* is one of the most basic problems in operations research (see the survey [PRW94]) and has been one of the main motivations for the work of Kleinberg and Tardos on metric labeling [KT02].

We also give a different use of our results for the problem of *min-sum k-clustering* [BCR01].

1.15 Distance Oracles

Thorup and Zwick [TZ05] study the problem of creating *distance oracles* for a given metric space. A distance oracle is a space efficient data structure which allows efficient queries for the approximate distance between pairs of points.

They give a distance oracle of space $O(kn^{1+1/k})$, query time of $O(k)$ and *worst case* distortion (also called stretch) of $2k-1$. They also show that this is nearly best possible in terms of the space-distortion tradeoff.

We extend the new notions of distortion in the context of distance oracles. In particular, we can define the ℓ_q-distortion of a distance oracle. Of particular interest are the average distortion and distortion of average notions. We also define partial distance oracles, distance oracle scaling distortion, and extend our results to distance labels and distributed labeled compact routing schemes in a similar fashion. Our main result is the following strengthening of [TZ05]:

Theorem 22. *Let (X,d) be a finite metric space. Let $k=O(\ln n)$ be a parameter. The metric space can be preprocessed in polynomial time, producing a data structure of size $O(n^{1+1/k}\log n)$, such that distance queries can be answered in $O(k)$ time. The distance oracle has worst case distortion $2k-1$. Given any distribution Π, its average distortion (and distortion of average) with respect to Π is $O(\log\hat{\Phi}(\Pi))$. In particular the average distortion (and distortion of average) is $O(1)$.*

Our extension of Assouad's theorem can yield an improved distance oracle for metrics with small doubling dimension. Taking $p=\infty$, $\theta=\Theta(1/\log k)$ and $\alpha=1/\log k$ in Theorem 17 yields the following:

Theorem 23. *Let (X,d) be a finite metric space. Let $k=O(\ln n)$ be a parameter. The metric space can be preprocessed in polynomial time, producing a data structure of size $O(n\cdot\lambda^{\frac{\log k}{k}}\log\lambda\log^2 k)$, such that distance queries can be answered in $O(\lambda^{\frac{\log k}{k}}\log\lambda\log^2 k)$ time, with worst case distortion $O(k)$.*

This distance oracle improves known constructions when $\dim(X)=o(\log_k n)$.

1.16 Subsequent Research and Impact

The results presented in this paper have been the basis for further developments both within the field of metric embedding and in other areas such as graph theory, distributed computing and algorithms. We give here a partial survey of such subsequent work.

Subsequent to [ABN06], the idea of average distortion and related notions have been further studied in different contexts. Chan et al. [CDG06] proved results on spanners with slack. Given a graph they find a sparse graph (with linear number of edges) that preserves the distances in the original graph. Their main result is partial and scaling spanners with $O(\log 1/\epsilon)$ distortion. Results of similar flavor were obtained for compact routing problems in the context of distributed networks [Din07, KRXY07].

Simultaneously and independently to our results on doubling metrics [ABN08], Chan, Gupta and Talwar [CGT08] obtained a result similar in spirit to our results of Section 6. Our main result on embedding metric spaces in their intrinsic dimension was used in [LS08] in the context of Kleinberg's the small world random graph model [Kle00].

The probabilistic partitions of [ABN06] were later used and refined in [ABN07b] and [ABN08] . For our main theorem of embedding with constant average distortion into l_p we did not require the partitions to possess the local property defined in the sequel. However for the results for doubling metrics and also for the local embedding results of [ABN07b], as well as in the work of [BRS11, GK11] this extra property is required, hence we present here our most general partitions.

Another application of the probabilistic partitions of [ABN06] is for constructing low stretch spanning trees, where for a given graph we wish to find a spanning tree with low average stretch (which is simply the average distortion over the edges of the graph). In [ABN08] we show a nearly tight result of $\tilde{O}(\log n)$. One of the ingredients of the construction is a generalization of the [ABN06] probabilistic partitions of metric spaces to graphs, *i.e.* the clusters of the partition are connected components of the graph. One of the main applications of low stretch spanning trees is solving *sparse symmetric diagonally dominant linear systems of equations*. This approach was suggested by Boman, Hendrickson and Vavasis [BHV08] and later improved by Spielman and Teng [ST04] to a near linear time solver. The best result of this type is due to Koutis, Miller and Peng [KMP]. A fast construction of a low stretch spanning tree is a basic component of these solvers, and the construction of [ABN08] plays a crucial role in the result of [KMP]. Another important application of this construction is for graph sparsification [KMST10].

In [ABN07a] we show that any metric can be embedded into a single ultrametric and any graph contains a spanning tree with *constant* average distortion (and ℓ_2-distortion of $O(\sqrt{\log n})$). Elkin et al. [ELR07] have shown that this implies a strictly fundamental cycle basis of length $O(n^2)$ for any unweighted graph, proving the conjecture of Deo et al. [DPK82].

Extending the main embedding result given here, [ABNS10] show an embedding that preserves not only pairwise distances but also volumes of sets of size k with average volume-distortion $O(\log k)$.

In [ABN07b] and [ABN09] our embedding techniques were used to obtain local embeddings were the distortion and dimension depend solely on the size of the local neighborhood in which the embedding preserves distances.

Recently, a local dimension reduction in Euclidean space was given [BRS11] which breaks the Johnson-Lindenstrauss [JL84] dimension bound. Similar techniques [GK11, BRS11] provide Assouad-type dimension reduction theorems. These results make use of various ideas among which are some of our notions and techniques.

1.17 Organization of the Paper

In Section 3 we define the new probabilistic partitions, including a uniform padding lemma, a lemma for decomposable metrics and a hierarchical lemma.

Section 4 contains the proof of our main embedding result. In Section 4.1 we present the main technical lemma, that gives an embedding into the line. This embedding is used in Section 4.2 to prove Theorem 9, which by Lemma 1 implies $O(1)$ average distortion, and ℓ_q-distortion of $O(q)$ as stated in Theorems 2, 3. The proof of Theorems 4, 5 is shown as well in that section. In Section 4.3 we extend the previous result for embedding into l_p proving Theorem 11 which implies also Theorem 6.

In Section 5 we show how to extend the embedding for infinite compact metric spaces, proving Theorem 12 and showing how it implies Theorem 7.

In Section 6 we prove the theorems regarding the intrinsic dimension of metric spaces, that are described in Section 1.12, in particular Theorem 8. In Section 7 we prove Theorem 18, a generalization for decomposable metrics.

In Section 8 we prove Lemma 1 and Lemma 62, showing the relation between scaling distortion and our notions of average distortion.

Next in Section 9 we prove Theorem 19 - a probabilistic embedding into distribution of trees with scaling distortion.

In Section 10 we include some results on partial embedding.

In Section 11 we prove all the lower bound results mentioned in Section 1.11, including Theorem 13.

Finally, in Section 12 we show some algorithmic applications of our methods and in Section 13 we show how our results can be used as distance oracles.

2 Preliminaries

Consider a finite metric space (X, d) and let $n = |X|$. The *diameter* of X is denoted $\mathrm{diam}(X) = \max_{x,y\in X} d(x, y)$. For a point x and $r \geq 0$, the ball at radius r around x is defined as $B_X(x, r) = \{z \in X | d(x, z) \leq r\}$. We omit the subscript X when it is clear from the context. The notation $B^\circ(x, r) = \{z \in X | d(x, z) < r\}$ stands for strict inequality. For any $\epsilon > 0$ let $r_\epsilon(x)$ denote the minimal radius r such that $|B(x, r)| \geq \epsilon n$.

2.1 Local Growth Rate

The following definition and property below are useful for the properties of our partitions described in Section 3.

Definition 13. The local growth rate of $x \in X$ at radius $r > 0$ for given scales $\gamma_1, \gamma_2 > 0$ is defined as

$$\rho(x, r, \gamma_1, \gamma_2) = |B(x, r\gamma_1)|/|B(x, r\gamma_2)|.$$

Given a subspace $Z \subseteq X$, the minimum local growth rate of Z at radius $r > 0$ and scales $\gamma_1, \gamma_2 > 0$ is defined as $\rho(Z, r, \gamma_1, \gamma_2) = \min_{x\in Z} \rho(x, r, \gamma_1, \gamma_2)$. The minimum local growth rate of $x \in X$ at radius $r > 0$ and scales $\gamma_1, \gamma_2 > 0$ is defined as $\bar{\rho}(x, r, \gamma_1, \gamma_2) = \rho(B(x, r), r, \gamma_1, \gamma_2)$.

Claim 3. *Let $x, y \in X$, let $\gamma_1, \gamma_2 > 0$ and let r be such that $2(1 + \gamma_2)r < d(x, y) \leq (\gamma_1 - \gamma_2 - 2)r$, then*

$$\max\{\bar{\rho}(x, r, \gamma_1, \gamma_2), \bar{\rho}(y, r, \gamma_1, \gamma_2)\} \geq 2.$$

Proof. Let $B_x = B(x, r(1 + \gamma_2))$, $B_y = B(y, r(1 + \gamma_2))$, and assume w.l.o.g that $|B_x| \leq |B_y|$. As $r(1 + \gamma_2) < d(x, y)/2$ we have $B_x \cap B_y = \emptyset$. Note that for any $x' \in B(x, r)$, $B(x', r\gamma_2) \subseteq B_x$, and similarly for any $y' \in B(y, r)$, $B(y', r\gamma_2) \subseteq B_y$. On the other hand $B(x', r\gamma_1) \supseteq B_x \cup B_y$, since for any $y' \in B_y$, $d(x', y') \le d(x', x) + d(x, y) + d(y, y') \le r + r(\gamma_1 - \gamma_2 - 2) + r(1 + \gamma_2) = r\gamma_1$. We conclude that

$$\rho(x', r, \gamma_1, \gamma_2) = |B(x', r\gamma_1)|/|B(x', r\gamma_2)| \ge (|B_x| + |B_y|)/|B_x| \ge 2.$$

$\square$

3 Partition Lemmas

In this section we show the main tool of our embedding: uniformly padded probabilistic partitions. We give several versions of these partitions, first a general one, then an extension of it to decomposable metrics (defined formally in the sequel), and finally a hierarchical construction of partitions. These partitions will be used in almost all the embedding results.

Definition 14 (Partition)**.** A partition P of X is a collection of pairwise disjoint sets $\mathcal{C}(P) = \{C_1, C_2, \dots, C_t\}$ 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 $j \in [t]$, $\mathrm{diam}(C_j) \le \Delta$. For $Z \subseteq X$ we denote by $P[Z]$ the restriction of P to points in Z.

Definition 15 (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 16 (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 17 (Uniformly Padded Local PP)**.** Given $\Delta > 0$ and $0 < \delta \le 1$, let $\hat{\mathcal{P}}$ be a Δ-bounded probabilistic partition of (X, d). Given collection of functions $\eta = \{\eta_P : X \to [0, 1] | P \in \mathcal{P}\}$, 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,

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

Let $0 < \hat{\delta} \le 1$. We say that $\hat{\mathcal{P}}$ is *strong* $(\eta, \hat{\delta})$-locally padded if for any $\hat{\delta} \le \delta \le 1$, $\hat{\mathcal{P}}$ is $(\eta \cdot \ln(1/\delta), \delta)$-padded.

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

The following lemma is a generalization of a decomposition lemma that appeared in [Bar04], which by itself is a generalization of the original probabilistic partitions of [Bar96, LS91]. For sets $A, B, C \subseteq X$ we denote by $A \bowtie (B, C)$ the property that $A \cap B \neq \emptyset$ and $A \cap C \neq \emptyset$.

Lemma 4 (Probabilistic Decomposition)**.** *For any metric space (Z, d), point $v \in Z$, real parameters $\chi \ge 2, \Delta > 0$, let r be a random variable sampled from a truncated exponential density function with parameter $\kappa = 8\ln(\chi)/\Delta$*

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

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

$$\Pr\left[B(x,\eta\Delta) \bowtie (S,\bar{S})\right] \le (1-\theta)\left(\Pr\left[B(x,\eta\Delta) \not\subseteq \bar{S}\right] + \frac{2\theta}{\chi}\right).$$

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

Proof. Let $x \in Z$. Let $a = \inf_{y\in B(x,\eta\Delta)}\{d(v,y)\}$ and $b = \sup_{y\in B(x,\eta\Delta)}\{d(v,y)\}$. By the triangle inequality: $b - a \le 2\eta\Delta$. We have:

$$\begin{aligned}
&\Pr[B(x,\eta\Delta) \bowtie (S,\bar{S})] = \\
&\quad \int_a^b f(r)dr = (\tfrac{\chi^2}{1-\chi^{-2}})\chi^{-\frac{8a}{\Delta}}(1-\chi^{-8\frac{b-a}{\Delta}}) \\
&\quad \le (\tfrac{\chi^2}{1-\chi^{-2}})\chi^{-\frac{8a}{\Delta}}(1-\theta),
\end{aligned} \tag{1}$$

which follows since:

$$\frac{8(b-a)}{\Delta} \le \frac{16\eta\Delta}{\Delta} = 16\eta = \ln_\chi(1/\theta).$$

$$\begin{aligned}
&\Pr[B(x,\eta\Delta) \not\subseteq \bar{S}] = \\
&\quad \int_a^{\Delta/2} f(r)dr = (\tfrac{\chi^2}{1-\chi^{-2}})(\chi^{-\frac{8a}{\Delta}} - \chi^{-4}).
\end{aligned} \tag{2}$$

Therefore we have:

$$\begin{aligned}
&\Pr[B(x,\eta\Delta) \bowtie (S,\bar{S})] - (1-\theta)\cdot\Pr[B(x,\eta\Delta) \not\subseteq \bar{S}] \\
&\quad \le (1-\theta)(\tfrac{\chi^2}{1-\chi^{-2}})\chi^{-4} \le (1-\theta)\cdot 2\chi^{-2},
\end{aligned}$$

where in the last inequality we have used the assumption that $\chi \ge 2$. Since $\chi^{-1} \le \theta$, this completes the proof of the lemma. □

3.1 Uniform Padding Lemma

The following lemma describes the uniform probabilistic partition, the uniformity is with respect to η - the padding parameter, which will the same for all points that are in the same cluster. This η will actually be a function of the local growth rate of a single point, "the center" of the cluster, which has the minimal local growth rate among all the other points in the cluster. The purpose of the function ξ is to indicate which clusters have high enough local growth rate at their centers for η to be as above, while the threshold for being high enough is set by the parameter $\hat{\delta}$.

Lemma 5. *Let (Z,d) be a finite metric space. Let $0 < \Delta \le \operatorname{diam}(Z)$. Let $\hat{\delta} \in (0,1/2]$, $\gamma_1 \ge 2$, $\gamma_2 \le 1/16$. There exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of (Z,d) and a collection of uniform functions $\{\xi_P : Z \to \{0,1\} \mid P \in \mathcal{P}\}$ and $\{\eta_P : Z \to (0,1] \mid P \in \mathcal{P}\}$ such that the probabilistic partition $\hat{\mathcal{P}}$ is a strong $(\eta,\hat{\delta})$-uniformly locally padded probabilistic partition; and the following conditions hold for any $P \in \operatorname{supp}(\hat{\mathcal{P}})$ and any $x \in Z$:*

- *If* $\xi_P(x) = 1$ *then:* $2^{-6}/\ln \rho(x, 2\Delta, \gamma_1, \gamma_2) \leq \eta_P(x) \leq 2^{-6}/\ln(1/\hat{\delta})$.
- *If* $\xi_P(x) = 0$ *then:* $\eta_P(x) = 2^{-6}/\ln(1/\hat{\delta})$ *and* $\bar{\rho}(x, 2\Delta, \gamma_1, \gamma_2) < 1/\hat{\delta}$.

Proof. We generate a probabilistic partition $\hat{\mathcal{P}}$ of Z by invoking the probabilistic decomposition Lemma 4 iteratively. Define the partition P of Z into clusters by generating a sequence of clusters: $C_1, C_2, \ldots C_s$, for some $s \in [n]$ whose value will be determined later . Notice that we are generating a distribution over partitions and therefore the generated clusters are random variables. First we deterministically assign centers $v_1, v_2, \ldots, v_s$ and parameters $\chi_1, \chi_2, \ldots, \chi_s$. Let $W_1 = Z$ and $j = 1$. Conduct the following iterative process:

1. Let $v_j \in W_j$ be the point minimizing $\hat{\chi}_j = \rho(x, 2\Delta, \gamma_1, \gamma_2)$ over all $x \in W_j$.
2. Set $\chi_j = \max\{2/\hat{\delta}^{1/2}, \hat{\chi}_j\}$.
3. Let $W_{j+1} = W_j \setminus B(v_j, \Delta/4)$.
4. 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 $Z_1 = Z$. For $j = 1, 2, 3 \ldots s$:

1. Let $(S_{v_j}, \bar{S}_{v_j})$ be the partition created by $S_{v_j} = B_{Z_j}(v_j, r)$ and $\bar{S}_{v_j} = Z_j \setminus S_{v_j}$ where r is distributed as in Lemma 4 with parameter $\kappa = 8\ln(\chi_j)/\Delta$.
2. Set $C_j = S_{v_j}$, $Z_{j+1} = \bar{S}_{v_j}$.
3. For all $x \in C_j$ let $\eta_P(x) = 2^{-6}/\max\{\ln \hat{\chi}_j, \ln(1/\hat{\delta})\}$. If $\hat{\chi}_j \geq 1/\hat{\delta}$ set $\xi_P(x) = 1$, otherwise set $\xi_P(x) = 0$.

Throughout the analysis fix some $\hat{\delta} \leq \delta \leq 1$. Let $\theta = \delta^{1/2}$, hence $\theta \geq 2\chi_j^{-1}$ for all $j \in [s]$. Let $\eta_j = 2^{-4}\ln(1/\theta)/\ln \chi_j = 2^{-5}\ln(1/\delta)/\ln \chi_j$. Note that for all $x \in C_j$ we have $\eta_P(x) \cdot \ln(1/\delta) = 2^{-6}\ln(1/\delta)\min\{1/\ln \hat{\chi}_j, 1/\ln(1/\hat{\delta})\} \leq 2^{-5}\ln(1/\delta)\min\{1/\ln \hat{\chi}_j, 1/\ln(2/\hat{\delta}^{1/2})\} = \eta_j$. Observe that some clusters may be empty and that it is not necessarily the case that $v_m \in C_m$. We now prove the properties in the lemma for some $x \in Z$. Consider the distribution over the clusters $C_1, C_2, \ldots C_s$ as defined above. For $1 \leq m \leq s$, define the events:

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

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

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

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

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

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 4

$$\begin{aligned}
&\Pr[\mathcal{F}_m] \\
&= \Pr[B(x, \eta_m\Delta) \bowtie (S_{v_m}, \bar{S}_{v_m}) | \mathcal{Z}_m] \\
&\le (1-\theta)(\Pr[B(x, \eta_m\Delta) \nsubseteq \bar{S}_{v_m} | \mathcal{Z}_m] + \theta\chi_m^{-1}) \\
&= (1-\theta)(\Pr[\bar{\mathcal{G}}_m] + \theta\chi_m^{-1}).
\end{aligned} \tag{4}$$

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 (4) and the induction hypothesis. Since the choice of radius is the only randomness in the process of creating P, the event of padding for $z \in Z$ is independent of all choices of radii for centers $v_j \notin T_z$. That is, for any assignment to clusters of points outside $B(z, 2\Delta)$ (this may determine radius choices for points in $Z \setminus B(z, \Delta)$), the padding probability will not be affected.

Fix some $x \in Z$, $T = T_x$. Observe that for all $v_j \in T$, $d(v_j, x) \le \Delta$, and so we get $B(v_j, 2\gamma_2\Delta) \subseteq B(x, 2\Delta)$. On other hand $B(v_j, 2\gamma_1\Delta) \supseteq B(x, 2\Delta)$. 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 $j \neq j'$, $d(v_j, v_{j'}) > \Delta/4 \ge 4\gamma_2\Delta$, so that $B(v_j, 2\gamma_2\Delta) \cap B(v_{j'}, 2\gamma_2\Delta) = \emptyset$. Hence, we get:

$$\begin{aligned}
\sum_{j \ge 1, v_j \in T} \chi_j^{-1} &\le \sum_{j \ge 1, v_j \in T} \hat{\chi}_j^{-1} \\
&\le \sum_{j \ge 1, v_j \in T} \frac{|B(v_j, 2\gamma_2\Delta)|}{|B(v_j, 2\gamma_1\Delta)|} \\
&\le \sum_{j \ge 1, v_j \in T} \frac{|B(v_j, 2\gamma_2\Delta)|}{|B(x, 2\Delta)|} \le 1.
\end{aligned}$$

Let $j \in [s]$ such that $P(x) = C_j$, then as $\eta_P(x) \cdot \ln(1/\delta) \le \eta_j$ follows $B(x, \eta_P(x) \cdot \ln(1/\delta)\Delta) \subseteq B(x, \eta_j\Delta)$. We conclude from the claim (3) for $m = 1$ that:

$$\begin{aligned}
&\Pr[B(x, \eta_P(x) \cdot \ln(1/\delta)\Delta) \nsubseteq P(x)] \le \Pr[\mathcal{E}_1] \le \\
&(1-\theta)(1 + \theta \cdot \sum_{j \ge 1, v_j \in T} \chi_j^{-1}) \le (1-\theta)(1+\theta) = 1 - \delta.
\end{aligned}$$

It follows that $\hat{\mathcal{P}}$ is strong uniformly padded. Finally, we show the properties stated in the lemma. Let $x \in Z$ and $j \in [s]$ be such that $x \in C_j$. For the first property if $\xi_P(x) = 1$ by definition $\hat{\chi}_j \ge 1/\hat{\delta}$

so $\eta_P(x) = 2^{-6}/\ln \rho(v_j, 2\Delta, \gamma_1, \gamma_2)$ and by the minimality of v_j, $\eta_P(x) \geq 2^{-6}/\ln \rho(x, 2\Delta, \gamma_1, \gamma_2)$. By definition also $\eta_P(x) \leq 2^{-6}/\ln(1/\hat{\delta})$. As for the second property, $\xi_P(x) = 0$ implies that $\hat{\chi}_j = \rho(v_j, 2\Delta, \gamma_1, \gamma_2) < 1/\hat{\delta}$ and $\bar{\rho}(x, 2\Delta, \gamma_1, \gamma_2) \leq \rho(v_j, 2\Delta, \gamma_1, \gamma_2)$, also by definition $\eta_P(x) = 2^{-6}/\ln(1/\hat{\delta})$. □

The following corollary shows that our probabilistic partitions may lead to results similar to those given in [FRT03] (which are based on [CKR01] and improved analysis of [FHRT03]).

Corollary 6. *Let (X, d) be a metric space. Let $\gamma_1 = 2$, $\gamma_2 = 1/32$. For any $\Delta > 0$ there exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$, which for any $1/2 \leq \delta \leq 1$ is (η, δ) padded, where*

$$\eta_P^{(\delta)}(x) = \min\left\{\frac{\ln(1/\delta)}{2^6 \ln(\rho(x, 2\Delta, \gamma_1, \gamma_2))}, 2^{-6}\right\} .$$

Proof. Let $\hat{\delta} = 1/2$, and let $\hat{\mathcal{P}}$ be a Δ-bounded probabilistic partition as in Lemma 5 with parameters $\hat{\delta}, \gamma_1, \gamma_2$. Let $\rho(x) = \rho(x, 2\Delta, \gamma_1, \gamma_2)$, $B(x) = B(x, \eta_P^{(\delta)}(x)\Delta)$ and let $1/2 \leq \delta \leq 1$. We distinguish between two cases:

Case 1: $\rho(x) < 2$. We will show that $\Pr[B(x) \nsubseteq P(x)] = 0$. Let j be the minimal such that v_j is a center of a cluster C_j that intersects $B(x)$. We will show that it must be the case that $d(x, v_j) \leq \Delta/8$. Assume this is the case then since $\eta_P^{(\delta)}(x) \leq 2^{-6}$, it follows that $B(x) \subseteq B(x, \Delta/2^6) \subseteq B(v_j, \Delta/4) \subseteq C_j = P(x)$. Now if we assume that $d(x, v_j) > \Delta/8$ we will reach to a contradiction: Let $A = |B(x, 4\Delta)|$, $a = |B(x, \Delta/16)|$, $B = |B(v_j, 4\Delta)|$ and $b = |B(v_j, \Delta/16)|$. Note that $\rho(x) = A/a$ and $\rho(v_j) = B/b$. By our assumption we have that $B(x, \Delta/16) \cap B(v_j, \Delta/16) = \emptyset$. As $d(x, v_j) \leq \Delta/2 + \Delta/32 \leq \Delta$ we have $a + b \leq A$, $a + b \leq B$, so that $A + B \geq 2(a + b)$. On the other hand from the minimality of $\rho(v_j)$ follows $\rho(v_j) < \rho(x) < 2$, therefore $A < 2a$ and $B < 2b$, hence $A + B < 2(a + b)$, a contradiction.

Case 2: $\rho(x) \geq 2$. In this case we simply use the argument in Lemma 5 which states that if $x \in C_j$ with center v_j then x is $(\eta'_P(x)\ln(1/\delta), \delta)$-padded for $\eta'_P(x) = 2^{-6}/\max\{\ln(\rho(v_j)), \ln 2\}$, and as v_j minimizes $\rho(v_j) \leq \rho(x)$, we have that $\eta_P^{(\delta)}(x) \leq \eta'_P(x)\ln(1/\delta)$ it follows that

$$\Pr[B(x) \subseteq P(x)] \geq \delta$$

□

Remark. In [MN06] an extension of the [FRT03] lemma was used to obtain metric Ramsey theorems ([BFM86, BBM06, BLMN05c]). We note that similar lemma follows from arguments in the proof of Lemma 5 combined with the above corollary, essentially extending the corollary to hold for all values of $0 < \delta \leq 1$.

3.2 Padding Lemma for Decomposable Metrics

In this section we extend the uniform padding lemma, and obtain an additional lower bound on the padding parameter with respect to the "decomposability" of the metric space, as given by the following definition.

Definition 18. Let (X,d) be a finite metric space. Let $\tau \in (0,1]$. We say that X is (locally) *τ-decomposable* if for any $0 < \Delta < \mathrm{diam}(X)$ there exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of X such that for all $\delta \le 1$ satisfying $\ln(1/\delta) \le 2^6\tau^{-1}$, $\hat{\mathcal{P}}$ is $(\tau \cdot \ln(1/\delta), \delta)$-(locally) padded.

It is known [LS91, Bar96] that any metric space is $\Omega(1/\log n)$-decomposable, however there are certain families of metric spaces which have a much larger decomposition parameter, such as doubling metrics and metrics derived from graphs that exclude a fixed minor. Note that we require padding for a wide range of the parameter δ and not just a fixed value (a common value used in many papers is $\delta = 1/2$).

Lemma 7 (Uniform Padding Lemma for Decomposable Metrics)**.** *Let (X,d) be a finite metric space. Assume X is (locally) τ-decomposable. 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 $\gamma_1 \ge 2$,$\gamma_2 \le 1/16$. 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 the probabilistic partition $\hat{\mathcal{P}}$ is a strong $(\eta, \hat{\delta})$-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: $2^{-7}/\ln\rho(x, 2\Delta, \gamma_1, \gamma_2) \le \eta_P(x) \le 2^{-7}/\ln(1/\hat{\delta})$.*
- *If $\xi_P(x) = 0$ then: $\eta_P(x) = 2^{-7}/\ln(1/\hat{\delta})$ and $\bar{\rho}(x, 2\Delta, \gamma_1, \gamma_2) < 1/\hat{\delta}$.*

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

Proof. We generate a probabilistic partition $\hat{\mathcal{P}}$ of X in two phases. the first phase is done by invoking the probabilistic decomposition Lemma 4 iteratively. 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 do the same partition as in Lemma 5 while 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 18. 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 $\hat{\chi}_j = \rho(x, 2\Delta, \gamma_1, \gamma_2)$ 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 4 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 $\hat{\chi}_j \ge 1/\hat{\delta}$ set $\xi_P(x) = 1$, otherwise set $\xi_P(x) = 0$.

Fix some $\hat{\delta} \le \delta \le 1$. Let $\theta = \delta^{1/4}$. Note that $\theta \ge 2\chi_j^{-1}$ 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 $P_2' = \{D_1, D_2, \dots, D_t\}$ be a Δ-bounded probabilistic partition of X, such that for all $\delta \le 1$ satisfying $\ln(1/\delta) \le 2^6\tau^{-1}$, P_2' is $(\tau \cdot \ln(1/\delta), \delta)$-padded. Let $Z = \bigcup_{i=1}^{s} C_i$ and $\bar{Z} = X \setminus Z$ (the un-assigned points), then let $P_2 = \{D_1 \cap \bar{Z}, D_2 \cap \bar{Z}, \dots, D_t \cap \bar{Z}\}$. 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**, $\tau \le 2^{-6}/\ln \hat{\chi}_j$, since by definition $\tau \le 2^{-6}/\ln(1/\hat{\delta})$ 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 $C_1, C_2, \dots C_s$ 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 $1 \le m \le s$:

$$\Pr[\mathcal{E}_m] \le (1-\theta)(1 + \theta \sum_{j \ge m, v_j \in T} \chi_j^{-1}). \tag{5}$$

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 4

$$\begin{aligned}
&\Pr[\mathcal{F}_m] \qquad\qquad (6)\\
&\quad= \Pr[B(x, \eta_m\Delta) \bowtie (S_{v_m}, \bar{S}_{v_m}) | \mathcal{Z}_m]\\
&\quad\le (1-\theta)(\Pr[B(x, \eta_m\Delta) \nsubseteq \bar{S}_{v_m} | \mathcal{Z}_m] + \theta\chi_m^{-1})\\
&\quad= (1-\theta)(\Pr[\bar{\mathcal{G}}_m] + \theta\chi_m^{-1}).
\end{aligned}$$

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:

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

The second inequality follows from (6) and the induction hypothesis. Fix some $x \in X$, $T = T_x$. Observe that for all $v_j \in T$, $d(v_j, x) \leq \Delta$, and so we get $B(v_j, 2\gamma_2\Delta) \subseteq B(x, 2\Delta)$. On the other hand $B(v_j, 2\gamma_1\Delta) \supseteq B(x, 2\Delta)$. 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 $j \neq j'$, $d(v_j, v_{j'}) > \Delta/4 \geq 4\gamma_2\Delta$, so that $B(v_j, 2\gamma_2\Delta) \cap B(v_{j'}, 2\gamma_2\Delta) = \emptyset$. Hence, we get:

$$\begin{aligned}
\sum_{j \geq 1, v_j \in T} \chi_j^{-1} &\leq \sum_{j \geq 1, v_j \in T} \hat{\chi}_j^{-1} \\
&\leq \sum_{j \geq 1, v_j \in T} \frac{|B(v_j, 2\gamma_2\Delta)|}{|B(v_j, 2\gamma_1\Delta)|} \\
&\leq \sum_{j \geq 1, v_j \in T} \frac{|B(v_j, 2\gamma_2\Delta)|}{|B(x, 2\Delta)|} \leq 1.
\end{aligned}$$

We conclude from the claim (5) for $m = 1$ that

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

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 $\eta_P(x) \cdot \ln(1/\delta) \leq \eta_j$ 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 $j \in [s]$, $\eta_P(x) \cdot \ln(1/\delta) \leq \eta_j$ 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 $\Pr[B(x, (\tau/2)\ln(1/\delta)) \subseteq P_2(x)] \geq \delta^{1/2}$, hence

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

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 $\eta_P(x) = 2^{-7}/\ln\hat{\chi}_j = 2^{-7}/\ln\rho(v_j, 2\Delta, \gamma_1, \gamma_2)$ and by the minimality of v_j, $\eta_P(x) \geq 2^{-7}/\ln\rho(x, 2\Delta, \gamma_1, \gamma_2)$. 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 $\bar{\rho}(x, 2\Delta, \gamma_1, \gamma_2) \leq \rho(v_j, 2\Delta, \gamma_1, \gamma_2) < 1/\hat{\delta}$. □

Lemma 8 (Local Padding Lemma for Doubling Metrics)**.** *Every finite metric space (X, d) is locally τ-decomposable where $\tau = 2^{-6}/\dim(X)$.*

Proof. Fix $0 < \Delta < \operatorname{diam}(X)$ and let λ denote the doubling constant of X. We generate a probabilistic partition $\hat{\mathcal{P}}$ of X by invoking the probabilistic decomposition Lemma 4 iteratively. Define the partition P of X into clusters by generating a sequence of clusters: $C_1, C_2, \dots C_s$.

First we deterministically assign centers $v_1, v_2, \dots, v_s$, by choosing an arbitrary sequence of an arbitrary $\Delta/4$-net of X. Now the algorithm for the partition is as follows: Let $Z_1 = X$. For $j = 1, 2, 3 \dots s$:

1. Let $(S_{v_j}, \bar{S}_{v_j})$ be the partition created by invoking Lemma 4 on Z_j with center $v = v_j$ and parameter $\chi = \chi_j = \lambda^4$.
2. Set $C_j = S_{v_j}$, $Z_{j+1} = \bar{S}_{v_j}$.

Throughout the analysis fix some δ and let $\theta = \delta^{1/2}$. Note that $\theta \geq \lambda^{-3} \geq 2\chi^{-1}$ as required, where we use the fact that $\lambda \geq 2$ assuming $|X| > 1$.

Recall that $\eta_j = 2^{-4}\ln(1/\theta)/\ln \chi_j = 2^{-5}\ln(1/\delta)/\ln \chi_j$, and define: $\eta_P(x) = \eta_j/\ln(1/\delta) = \tau/2$. Define the events

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

Also let $T = T_x = B(x, \Delta)$. The following inductive claim is identical to that in Lemma 5: For every $1 \leq m \leq s$:

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

Now consider a fixed choice of partition P. Let t_T be the number of center points v_j such that $v_j \in T$. Consider covering of T by balls of radius $\Delta/8$. Observe that there exists such a covering with at most λ^4 balls. Since the centers are a net for any $j \neq j', d(v_j, v_j') > \Delta/4$. It follows that each of the balls in the covering of T contains at most one v_j and therefore $t_T \leq \lambda^4$. We therefore obtain:

$$\sum_{j \geq 1, v_j \in T} \chi_j^{-1} = t_T \cdot \lambda^{-4} \leq 1.$$

For $x \in X$, if $P(x) = S_{v_j}$ then by definition $\eta_P(x)\ln(1/\delta) = \eta_j$. We conclude that:

$$\Pr[B(x, (\eta_P(x))\ln(1/\delta)\Delta) \nsubseteq P(x)] = \Pr[\mathcal{E}_1] \leq (1-\theta)(1 + \theta\mathbb{E}[\sum_{j \geq 1, v_j \in T} \chi_j^{-1}]) \leq (1-\theta)(1+\theta) = 1 - \delta.$$

□

We also have the following Lemma from [KPR93, FT03]

Lemma 9. *Let G be a weighted graph that excludes the minor K_r. Then the metric (X, d) derived from the graph is τ-decomposable for any $0 < \Delta < \operatorname{diam}(X)$ where $\tau = 2^{-6}/r^2$.*

3.3 Hierarchical Padding Lemma

Definition 19 (Hierarchical Partition)**.** Given a finite metric space (X, d) and a parameter $k > 1$, let $\Lambda = \frac{\max_{x,y\in X}\{d(x,y)\}}{\min_{x\neq y\in X}\{d(x,y)\}}$ be the aspect ratio of (X, d) and let $I = \{0 \leq i \leq \log_k \Lambda | i \in \mathbb{N}\}$. Let $\Delta_0 = \mathrm{diam}(X)$, and for each $0 < i \in I$, $\Delta_i = \Delta_{i-1}/k$. A *$k$-hierarchical partition* H of (X, d) is a hierarchical collection of partitions $\{P_i\}_{i\in I}$, each P_i is Δ_i-bounded, where P_0 consists of a single cluster equal to X and for any $0 < i \in I$ and $x \in X$, $P_i(x) \subseteq P_{i-1}(x)$.

Definition 20 (Prob. Hierarchical Partition)**.** A *probabilistic k-hierarchical partition* $\hat{\mathcal{H}}$ of a finite metric space (X, d) consists of a probability distribution over a set $\mathcal{H}$ of k-hierarchical partitions. A collection of functions defined on X, $f = \{f_{P,i} | P_i \in H, H \in \mathcal{H}, i \in I\}$ is *uniform* with respect to $\mathcal{H}$ if for every $H \in \mathcal{H}$, $i \in I$, $f_{P,i}$ is uniform with respect to P_i.

Definition 21 (Uniformly Padded PHP)**.** Let $\hat{\mathcal{H}}$ be a probabilistic k-hierarchical partition. Given collection of functions $\eta = \{\eta_{P,i} : X \to [0, 1] | i \in I, P_i \in H, H \in \mathcal{H}\}$ and $\hat{\delta} \in (0, 1]$, $\hat{\mathcal{H}}$ is called *$(\eta, \hat{\delta})$-padded* if the following condition holds for all $i \in I$ and for any $x \in X$:

$$\Pr[B(x, \eta_{P,i}(x)\Delta_i) \subseteq P_i(x)] \geq \hat{\delta}.$$

$\hat{\mathcal{H}}$ is called *strong $(\eta, \hat{\delta})$-padded* if for all $\hat{\delta} \leq \delta \leq 1$, $\hat{\mathcal{H}}$ is *$(\eta \cdot \ln(1/\delta), \delta)$-padded.* We say $\hat{\mathcal{H}}$ is *uniformly padded* if η is uniform with respect to $\mathcal{H}$.

In order to construct partitions in a hierarchical manner, one has to note that the padding in level $i \in I$ can fail because of the partition of level $j < i$. The intuition is that this probability decays exponentially with $i - j$, however in order to make this work we will use the fact that our partitions are strongly padded, and argue about padding in all the levels $1, \ldots, i-1$ with larger value of δ. The main property of the hierarchical partition is that the sum of the inverse padding parameters over all levels in which there actually was a local growth rate (this is indicated by $\xi = 1$) is bounded by a logarithm of a "global" growth rate - this is attained by a telescopic sum argument.

Lemma 10 (Hierarchical Uniform Padding Lemma for Decomposable Metrics)**.** *Let (X, d) be a τ-decomposable finite metric space, and let $\gamma_1 = 16$, $\gamma_2 = 1/16$. Let $\hat{\delta} \in (0, \frac{1}{2}]$ such that $\ln(1/\hat{\delta}) \leq 2^6\tau^{-1}$. There exists a probabilistic 2-hierarchical partition $\hat{\mathcal{H}}$ of (X, d) and uniform collections of functions $\xi = \{\xi_{P,i} : X \to \{0, 1\} | i \in I, P_i \in H, H \in \mathcal{H}\}$ and $\eta = \{\eta_{P,i} : X \to \{0, 1/\ln(1/\hat{\delta})\} | i \in I, P_i \in H, H \in \mathcal{H}\}$, such that $\hat{\mathcal{H}}$ is strong $(\eta, \hat{\delta})$-uniformly padded, and the following properties hold:*

- $$\sum_{j\leq i} \xi_{P,j}(x)\eta_{P,j}(x)^{-1} \leq 2^{14} \ln\left(\frac{n}{|B(x, \Delta_{i+4})|}\right).$$

and for any $H \in \mathcal{H}$, $0 < i \in I$, $P_i \in H$:

- $\eta_{P,i} \geq \tau/8$.
- *If $\xi_{P,i}(x) = 1$ then: $\eta_{P,i}(x) \leq 2^{-9}/\ln(1/\hat{\delta})$.*
- *If $\xi_{P,i}(x) = 0$ then: $\eta_{P,i}(x) = 2^{-9}/\ln(1/\hat{\delta})$ and $\bar{\rho}(x, \Delta_{i-1}, \gamma_1, \gamma_2) < 1/\hat{\delta}$.*

Proof. We create a probability distribution over hierarchical partitions, by showing how to sample a random $H \in \mathcal{H}$, and uniform functions ξ and η. Define P_0 as a single cluster equal to X. For all $x \in X$, set $\hat{\eta}_{P,0}(x) = 2^{-9}/\ln(1/\hat{\delta})$, $\xi_{P,0}(x) = 0$. The rest of the levels of the partition are created iteratively using Lemma 7 as follows. Let $i = 1$.

1. For each cluster $S \in P_{i-1}$, let $P[S]$ be a Δ_i-bounded probabilistic partition created by invoking Lemma 7 on S with the parameters $\hat{\delta}, \gamma_1, \gamma_2$, and let ${\xi'}_{P[S]}$, $\eta'_{P[S]}$ be the uniform functions defined in Lemma 7.

2. Let $P_i = \cup_{S \in P_{i-1}} P[S]$.

3. For each cluster $S \in P_{i-1}$ and each $x \in S$ let $\eta_{P,i}(x) = \min\{\frac{1}{4} \cdot \eta'_{P[S]}(x), \frac{3}{2} \cdot \eta_{P,i-1}(x)\}$. If it is the case that $\eta_{P,i}(x) = \frac{1}{4} \cdot \eta'_{P[S]}(x)$ and also ${\xi'}_{P[S]}(x) = 0$ then set $\xi_{P,i}(x) = 0$, otherwise $\xi_{P,i}(x) = 1$.

4. Let $i = i + 1$, if $i \in I$, return to 1.

Note, that for $i \in I$, $x, y \in X$ such that $P_i(x) = P_i(y)$, it follows by induction that $\eta_{P,i}(x) = \eta_{P,i}(y)$ and $\xi_{P,i}(x) = \xi_{P,i}(y)$, by using the fact that η' and ξ' are uniform functions with respect to $P[S]$, where $S = P_{i-1}(x) = P_{i-1}(y)$.

We prove by induction on i that P_i is strong $(\eta, \hat{\delta})$-uniformly padded, *i.e.* that it is $(\eta \cdot \ln(1/\delta), \delta)$-padded for all $\hat{\delta} \le \delta \le 1$. Assume it holds for $i-1$ and we will prove for i. Now fix some $\hat{\delta} \le \delta \le 1$. Let $B_i = B(x, \eta_{P,i}(x)\ln(1/\delta)\Delta_i)$. We have:

$$\begin{aligned} &\Pr[B_i \subseteq P_i(x)] = \\ &\qquad \Pr[B_i \subseteq P_{i-1}(x)] \cdot \Pr[B_i \subseteq P_i(x) | B_i \subseteq P_{i-1}(x)]. \end{aligned} \tag{7}$$

Let $S = P_{i-1}(x)$. Note that $\eta_{P,i}(x)\ln(1/\delta) \le \frac{1}{4} \cdot \eta'_{P[S]}(x)\ln(1/\delta) = \eta'_{P[S]}(x)\ln(1/\delta^{1/4})$. Since $\delta^{1/4} \ge \hat{\delta}$, we have by Lemma 7 on S that $\Pr[B_i \subseteq P_i(x) | B_i \subseteq P_{i-1}(x)] \ge \delta^{1/4}$.

Next observe that by definition $\eta_{P,i}(x)\ln(1/\delta) \le \frac{3}{2} \cdot \eta_{P,i-1}(x)\ln(1/\delta) = \frac{3}{2} \cdot \frac{4}{3} \eta_{P,i-1}(x)\ln(1/\delta^{3/4}) = 2\eta_{P,i-1}(x)\ln(1/\delta^{3/4})$. Since $\Delta_i = \Delta_{i-1}/2$ we get that $\eta_{P,i}(x)\ln(1/\delta)\Delta_i \le \eta_{P,i-1}(x)\ln(1/\delta^{3/4})\Delta_{i-1}$. Therefore $B_i \subseteq B(x, \eta_{P,i-1}(x)\ln(1/\delta^{3/4})\Delta_{i-1})$. Using the induction hypothesis it follows that $\Pr[B_i \subseteq P_{i-1}(x)] \ge \delta^{3/4}$. We conclude from (7) above that the inductive claim holds: $\Pr[B_i \subseteq P_i(x)] \ge \delta^{1/4} \cdot \delta^{3/4} = \delta$. This completes the proof that $\mathcal{H}$ is strong $(\eta, \hat{\delta})$-uniformly padded.

We now turn to prove the properties stated in the lemma. The second property holds by induction on i: assume $\eta_{P,i-1}(x) \ge \tau/8$ and by the first property of Lemma 7 $\eta_{P,i}(x) = \min\{\frac{1}{4} \cdot \eta'_{P[S]}(x), \frac{3}{2} \cdot \eta_{P,i-1}(x)\} \ge \min\{\frac{1}{4} \cdot \tau/2, \frac{3}{2} \cdot \tau/8\} = \tau/8$. Consider some $i \in I$, $x \in X$ and let $S = P_{i-1}(x)$. The third property holds as $\eta_{P,i}(x) \le \frac{1}{4}\eta'_{P[S]}(x) \le 2^{-9}/\ln(1/\hat{\delta})$, using Lemma 7. Let us prove the fourth property. By definition if $\xi_{P,i}(x) = 0$ then $\eta_{P,i}(x) = \frac{1}{4}\eta'_{P[S]}(x)$ and ${\xi'}_{P[S]}(x) = 0$. Using Lemma 7 we have that $\eta_{P,i}(x) = 2^{-9}/\ln(1/\hat{\delta})$ and that $\bar{\rho}(x, \Delta_{i-1}, \gamma_1, \gamma_2) < 1/\hat{\delta}$.

It remains to prove the first property of the lemma. Define $\psi_{P,i}(x) = 2^{-9} \cdot \xi_{P,i}(x)\eta_{P,i}(x)^{-1}$. Using Lemma 7 it is easy to derive the following recursion: $\psi_{P,i}(x) \le \ln\rho(x, \Delta_{i-1}, \gamma_1, \gamma_2) + (2/3)\psi_{P,i-1}(x)$. A simple induction on t shows that for any $0 \le t < i$: $\sum_{t<j\le i} \psi_{P,j}(x) \le 3\sum_{t<j\le i} \ln\rho(x, \Delta_{j-1}, \gamma_1, \gamma_2) + 2\psi_{P,t}(x)$. Now observe that as $\gamma_1 = 16$, $\gamma_2 = 1/16$ and that for

any $j \in I$:

$$\begin{aligned}\ln \rho(x, \Delta_j, \gamma_1, \gamma_2) &= \ln\left(\frac{|B(x, \Delta_j\gamma_1)|}{|B(x, \Delta_j\gamma_2)|}\right) \\ &= \sum_{h=-3}^{4} \ln\left(\frac{|B(x, 2\Delta_{j+h})|}{|B(x, \Delta_{j+h})|}\right).\end{aligned}$$

It follows that

$$\begin{aligned}\sum_{0<j\le i} \psi_{P,j}(x) &\le 3 \sum_{0<j\le i} \ln \rho(x, \Delta_{j-1}, \gamma_1, \gamma_2) \\ &= 3 \sum_{0\le j<i} \sum_{h=-3}^{4} \ln\left(\frac{|B(x, 2\Delta_{j+h})|}{|B(x, \Delta_{j+h})|}\right) \\ &= 3 \sum_{h=-3}^{4} \sum_{0\le j<i} \ln\left(\frac{|B(x, 2\Delta_{j+h})|}{|B(x, \Delta_{j+h})|}\right) \\ &= 24 \ln\left(\frac{n}{|B(x, \Delta_{i+4})|}\right).\end{aligned}$$

This completes the proof of the first property of the lemma. □

4 Embedding with Scaling Distortion

In this section we prove our main theorem on embeddings with scaling distortion. The construction is based on the following lemma which gives an embedding into the real line, which is good for all pairs in expectation. The parameter ζ determines the quality of the embedding, and as a consequence the number of coordinates needed (which is calculated in Section 4.2) for the distortion to be good for all pairs, is also a function of ζ.

4.1 Main Scaling Lemma

Lemma 11. *Let (X, d) be a finite metric space on n points and let $0 < \zeta \le 1/8$, then there exists a distribution $\mathcal{D}$ over functions $f : X \to \mathbb{R}$ such that for all $u, v \in X$:*

1. *For all $f \in \operatorname{supp}(\mathcal{D})$,*

$$|f(u) - f(v)| \le C\left\lceil \ln\left(\frac{n}{|B(u, d(u, v))|}\right)\right\rceil \cdot d(u, v).$$

2. $$\Pr_{f\sim\mathcal{D}}\left[|f(u) - f(v)| \ge \zeta^3 \cdot d(u, v)/C\right] \ge 1 - \zeta,$$

   where C is a universal positive constant.

In the remainder of this section we prove this lemma, let us begin with the construction of the distribution $\mathcal{D}$.

Let $\Delta_0 = \mathrm{diam}(X)$. For $i \in \mathbb{N}$ let $\Delta_i = (\zeta/8)^i \Delta_0$ and let P_i be a Δ_i-bounded partition.

For all $i \in \mathbb{N}$ let $\sigma_i : X \to [0,1]$, $\xi_i : X \to \{0,1\}$, $\eta_i : X \to \mathbb{R}^+$ be uniform functions with respect to P_i, the functions η_i and ξ_i will be randomly generated by the probabilistic partition. For every scale $i \in \mathbb{N}$ define $\varphi_i : X \to \mathbb{R}^+$ as

$$\varphi_i(x) = \min\left\{\frac{\xi_i(x)}{\eta_i(x)} d\left(x, X \setminus P_i(x)\right), \zeta\Delta_i/4\right\} \ , \tag{8}$$

and for $i \in \mathbb{N}$ define $\psi_i : X \to \mathbb{R}^+$ as

$$\psi_i(x) = \sigma_i(x) \cdot \varphi_i(x).$$

Finally let $f : X \to \mathbb{R}^+$ be defined as $f = \sum_{i\in\mathbb{N}} \psi_i$. Note that f is well defined because $f(x) = \sum_{i\in\mathbb{N}} \psi_i(x) \le \sum_{i\in\mathbb{N}} \Delta_i$, and this is a geometric progression and $\psi_i \ge 0$.

The distribution $\mathcal{D}$ on embeddings f is obtained by choosing each P_i from the distribution $\hat{\mathcal{P}}_i$ as in Lemma 5 with parameters $Z = X$, $\Delta = \Delta_i$, $\hat{\delta} = 1/2$, $\gamma_1 = 8/\zeta$ and $\gamma_2 = 1/16$. For each $i \in \mathbb{N}$ set $\xi_i = \xi_{P_i}$ and $\eta_i = \eta_{P_i}$ as defined in the lemma. For each $i \in \mathbb{N}$, let σ_i be a uniform function with respect to P_i defined by setting $\{\sigma_i(C) | C \in P_i, 0 < i \in I\}$ as i.i.d random variables chosen uniformly in the interval $[0,1]$.

Lemma 12. *For all $u, v \in X$, $f \in \mathrm{supp}(\mathcal{D})$,*

$$|f(u) - f(v)| \le C \left\lceil \ln\left(\frac{|X|}{|B(u, d(u,v))|}\right)\right\rceil d(u,v).$$

where C is a universal constant.

Proof. Fix some $u, v \in X$ and $f \in \mathrm{supp}(\mathcal{D})$. Hence $\{P_i\}_{i\in\mathbb{N}}$, $\{\sigma_i\}_{i\in\mathbb{N}}$ are fixed. Let $\ell \in \mathbb{N}$ be the maximum index such that $\Delta_\ell \ge 2d(u,v)$, if no such ℓ exists then let $\ell = 0$. We bound $|f(u) - f(v)|$ by separating the sum into two intervals $0 \le i < \ell$, and $i \ge \ell$:

$$|f(u) - f(v)| \le \sum_{0\le i<\ell} |\psi_i(u) - \psi_i(v)| + \sum_{i\ge\ell} |\psi_i(u)| + \sum_{i\ge\ell} |\psi_i(v)|. \tag{9}$$

Proposition 13. *For any $x, y \in X$, a set $U \subset X$ and $r \in \mathbb{R}^+$, $\min\{d(u,U), r\} - \min\{d(v,U), r\} \le d(u,v)$.*

Proof. If it is the case that $r = \min\{d(v,U), r\}$ then $\min\{d(u,U), r\} - \min\{d(v,U), r\} = \min\{d(u,U), r\} - r \le r - r = 0$. Otherwise $\min\{d(v,U), r\} = d(v,U)$ and $\min\{d(u,U), r\} - \min\{d(v,U), r\} \le d(u,U) - d(v,U) \le d(u,v)$ by the triangle inequality. ☐

Each term of (9) is bounded as follows:

Claim 14. *For any $u, v \in X$, $\psi_i(u) - \psi_i(v) \le \frac{\xi_i(u)}{\eta_i(u)} d(u,v)$.*

Proof. The fact that σ_i, ξ_i, η_i are uniform implies that for each $0 \le i < \ell$: if it is the case that $P_i(u) = P_i(v)$ then by Proposition 13 with $U = P_i(x)$ and $r = \zeta\Delta_i/4$, $\psi_i(u) - \psi_i(v) \le \frac{\xi_i(u)}{\eta_i(u)} d(u,v)$. Otherwise, if $P_i(u) \ne P_i(v)$, then $d(u, X \setminus P_i(u)) \le d(u,v)$ and hence $\psi_i(u) - \psi_i(v) \le \psi_i(u) \le \frac{\xi_i(u)}{\eta_i(u)} d(u,v)$. ☐

By symmetry we have that

$$|\psi_i(u) - \psi_i(v)| \le \frac{\xi_i(u)}{\eta_i(u)} d(u,v) + \frac{\xi_i(v)}{\eta_i(v)} d(u,v) \ .$$

For any $x \in X$

$$\begin{aligned} \sum_{0\le i<\ell} \frac{\xi_i(x)}{\eta_i(x)} &= \sum_{0\le i<\ell : \xi_i(x)=1} \eta_i(x)^{-1} \\ &\le \sum_{0\le i<\ell : \xi_i(x)=1} 2^6 \ln \rho(x, 2\Delta_i, \gamma_1, \gamma_2) \\ &\le 2^6 \sum_{0\le i<\ell} \ln\left(\frac{|B(x, 2\gamma_1\Delta_i)|}{|B(x, 2\gamma_2\Delta_i)|}\right) \\ &\le 2^6 \cdot 3 \ln\left(\frac{|X|}{|B(x, \Delta_{\ell-1}/8)|}\right) \\ &\le 2^9 \ln\left(\frac{|X|}{|B(x, \Delta_\ell)|}\right) . \end{aligned} \tag{10}$$

The first inequality follows from the first property of Lemma 5, and the third inequality holds as $2\gamma_1\Delta_i = 16\Delta_{i-1} = 2\gamma_2 \cdot 8^2\Delta_{i-1} \le 2\gamma_2\Delta_{i-3}$ (since $\zeta \le 1/8$), this suggests that the sum is telescopic and is bounded accordingly. And now, noticing that $|\psi_i(u)| \le \zeta\Delta_i/4$ for all $u \in X$ and $i \in \mathbb{N}$,

$$\begin{aligned} |f(u) - f(v)| &\le \sum_{0\le i<\ell} |\psi_i(u) - \psi_i(v)| + \sum_{i\ge\ell} |\psi_i(u)| + \sum_{i\ge\ell} |\psi_i(v)| \\ &\le \sum_{0\le i<\ell} \left(\frac{\xi_i(u)}{\eta_i(u)} + \frac{\xi_i(v)}{\eta_i(v)}\right) d(u,v) + (\zeta/4)\sum_{i\ge\ell}\Delta_i + (\zeta/4)\sum_{i\ge\ell}\Delta_i \\ &\le 2^9 \left(\ln\left(\frac{|X|}{|B(u,\Delta_\ell)|}\right) + \ln\left(\frac{|X|}{|B(v,\Delta_\ell)|}\right)\right) d(u,v) + \zeta\Delta_\ell \\ &\le C\left\lceil \ln\left(\frac{|X|}{|B(u, d(u,v))|}\right)\right\rceil d(u,v). \end{aligned}$$

The third inequality uses (10). The last inequality uses the fact that $B(u, d(u,v)) \subseteq B(u,\Delta_\ell) \cap B(v,\Delta_\ell)$ and that the maximality of ℓ suggests that $\Delta_\ell \le 16d(u,v)/\zeta$. □

Lemma 15. *For each $u, v \in X$,* $\Pr\left[|f(u) - f(v)| \ge \zeta^3 \cdot d(u,v)/C\right] \ge 1 - \zeta$.

Let $s : \binom{X}{2} \to \mathbb{N}$ by $s(u,v) = k$ for the unique k satisfying $8\Delta_k \le d(u,v) < 8\Delta_{k-1}$. We will use the following claims:

Claim 16. *For each $u, v \in X$, let $k = s(u,v)$, then $\xi_k(u) + \xi_k(v) > 0$.*

Proof. Using Claim 3 with parameters $r = 2\Delta_k$, γ_1, γ_2, we have that indeed $2(1+\gamma_2)r < 8\Delta_k \le d(u,v)$ and $(\gamma_1 - \gamma_2 - 2)r \ge 8\Delta_{k-1} > d(u,v)$ so $\max\{\bar{\rho}(u, 2\Delta_k, \gamma_1, \gamma_2), \bar{\rho}(v, 2\Delta_k, \gamma_1, \gamma_2)\} \ge 2$. By the second property of Lemma 5 it follows that $\xi_k(u) + \xi_k(v) > 0$, using that $1/\hat{\delta} = 2$. □

Claim 17. *Let $A, B \in \mathbb{R}^+$ and let α, β be i.i.d random variables uniformly distributed in $[0,1]$. Then for any $C \in \mathbb{R}$ and $\gamma > 0$:*

$$\Pr[|C + A\alpha - B\beta| < \gamma \cdot \max\{A, B\}] < 2\gamma.$$

Proof. Assume w.l.o.g $A \geq B$. Consider the condition $|C + A\alpha - B\beta| < \gamma \cdot \max\{A, B\} = \gamma A$. If $C - B\beta \geq 0$ then it implies $\alpha < \gamma$. Otherwise $|\alpha - \frac{B\beta - C}{A}| < \gamma$. $\square$

Proof of the lemma. Fix $u, v \in X$ and let $k = s(u,v)$. Since $\xi_k(u) + \xi_k(v) > 0$ then assume without loss of generality that $\xi_k(u) = 1$. Recall that $\hat{\mathcal{P}}_k$ is a strong $(\eta_k, 1/2)$ locally padded probabilistic partition, hence it is $(\eta \cdot \ln(1/\delta), \delta)$-padded for all $1/2 \leq \delta \leq 1$. We take $\delta = 1 - \zeta/2$. Note that as $0 < \zeta \leq 1/8$, $\frac{1}{1-\zeta/2} = 1 + \frac{\zeta/2}{1-\zeta/2} \geq e^{\frac{\zeta/2}{2(1-\zeta/2)}}$ hence $\ln\left(\frac{1}{1-\zeta/2}\right) \geq \zeta/4$

Let $\mathcal{E}_{u-\mathsf{pad}}$ be the event $\{B(u, \eta_k(u) \cdot \zeta\Delta_k/4) \subseteq P_k(u)\}$. From the properties of Lemma 5 we have $\Pr[\mathcal{E}_{u-\mathsf{pad}}] \geq 1 - \zeta/2$. In this case, given $\mathcal{E}_{u-\mathsf{pad}}$,

$$\varphi_k(u) = \min\left\{\frac{d(u, X \setminus P_k(u))}{\eta_k(u)}, \zeta\Delta_k/4\right\} \geq \zeta\Delta_k/4.$$

Let $\mathcal{E}_{u-\mathsf{color}}$ be the event that $|\sum_{0<j\leq k}(\psi_j(u) - \psi_j(v))| \geq (\zeta/4)^2\Delta_k$ and $\mathcal{E}_{uv-\mathsf{good}}$ be the event that both events $\mathcal{E}_{u-\mathsf{pad}}$, $\mathcal{E}_{u-\mathsf{color}}$ hold. We will show that

$$\Pr[\mathcal{E}_{u-\mathsf{color}} \mid \mathcal{E}_{u-\mathsf{pad}}] \geq 1 - \zeta/2 \;, \tag{11}$$

therefore

$$\Pr[\mathcal{E}_{uv-\mathsf{good}}] = \Pr[\mathcal{E}_{u-\mathsf{pad}} \wedge \mathcal{E}_{u-\mathsf{color}}] = \Pr[\mathcal{E}_{u-\mathsf{pad}}] \cdot \Pr[\mathcal{E}_{u-\mathsf{color}} \mid \mathcal{E}_{u-\mathsf{pad}}] \geq (1-\zeta/2)^2 \geq 1 - \zeta.$$

Now to prove (11). Define $A = \varphi_k(u)$, $B = \varphi_k(v)$, $\alpha = \sigma_k(u)$, $\beta = \sigma_k(v)$ and $C = \sum_{j<k}(\psi_j(u) - \psi_j(v))$. Since $\mathrm{diam}(P_k(u)) \leq \Delta_k < d(u,v)$ we have that $P_k(v) \neq P_k(u)$. Thus α and β are independent random variables uniformly distributed in $[0,1]$, hence we can apply Claim 17 with $\gamma = \zeta/4$, noticing that given $\mathcal{E}_{u-\mathsf{pad}}$, $\max\{A, B\} \geq \zeta\Delta_k/4$

$$\begin{aligned}
\Pr[\neg\mathcal{E}_{u-\mathsf{color}} \mid \mathcal{E}_{u-\mathsf{pad}}] &\leq \Pr[|C + A\alpha - B\beta| < \gamma \cdot \max\{A, B\} \mid \mathcal{E}_{u-\mathsf{pad}}] \\
&\leq \Pr[|C + A\alpha - B\beta| < (\zeta/4)^2\Delta_k \mid \mathcal{E}_{u-\mathsf{pad}}] \\
&< \zeta/2.
\end{aligned}$$

Note that $|\psi_j(u) - \psi_j(v)| \leq \zeta\Delta_j/4$, hence

$$|\sum_{j>k}(\psi_j(u) - \psi_j(v))| \leq (\zeta/4)\sum_{j>k}\Delta_j \leq (\zeta/4) \cdot \zeta\Delta_k/6 = (2/3) \cdot (\zeta/4)^2\Delta_k \;.$$

We conclude that with probability at least $1 - \zeta$ event $\mathcal{E}_{uv-\mathsf{good}}$ occur and then

$$|f(u) - f(v)| \geq |\sum_{0<j\leq k}(\psi_j(u) - \psi_j(v))| - |\sum_{j>k}(\psi_j(u) - \psi_j(v))| \geq (1/3) \cdot (\zeta/4)^2\Delta_k \geq \zeta^3 d(u,v)/C \;,$$

for $C \geq 384$.

$\square$

Lemma 18. *The embedding of Lemma 15 can actually give a stronger local result. For any pair u, v with $s(u,v) = k$ define $Q = Q(u,v) \subseteq \binom{X}{2}$ by*

$$Q = \left\{ (u', v') \in \binom{X}{2} \mid s(u', v') < k \vee (s(u', v') = k \wedge d(\{u, v\}, \{u', v'\}) \geq 4\Delta_k) \right\} ,$$

then

$$\Pr \left[\neg \mathcal{E}_{uv-good} \mid \bigwedge_{(u',v') \in Q} \mathcal{E}_{u'v'-good} \right] \leq \zeta.$$

Proof. The observation is that the bound on the probability of event $\mathcal{E}_{uv-\mathsf{good}}$ depends only on random variables $\sigma_k(u), \sigma_k(v)$ and w.l.o.g the event $\mathcal{E}_{u-\mathsf{pad}}$, given any outcome for scales $1, 2, \dots k-1$, and is oblivious to all events that happen in scales $k+1, k+2, \dots$. The events $\{\mathcal{E}_{u'v'-\mathsf{good}}\}_{(u',v') \in Q}$ either depend on scale $< k$, in this case $\mathcal{E}_{uv-\mathsf{good}}$ holds with probability at least $1 - \zeta$ given any outcome for those events. If $s(u', v') = k$ then it must be that $d(\{u, v\}, \{u', v'\}) \geq 4\Delta_k$, now the locality of the partition suggests that the event $\mathcal{E}_{u-\mathsf{pad}}$ has probability at least $1 - \zeta/2$ given any outcome for $\mathcal{E}_{u'v'-\mathsf{good}}$. Since any partition $P_k \in$ *supp*$(\hat{\mathcal{P}}_k)$ is Δ_k-bounded it follows that $\{P_k(u), P_k(v)\} \cap \{P_k(u'), P_k(v')\} = \emptyset$, *i.e.* the random variables σ_k for each pair are independent. □

4.2 Scaling Distortion with Low Dimension

Now we prove the following corollary of the embedding into the line

Corollary 19. *For any $1 \leq p \leq \infty$, any finite metric space (X, d) on n points and any $\theta \geq (12/\log\log n)$ there is an embedding $F : X \to l_p^D$ with coarse scaling distortion $O(\log(2/\epsilon) \cdot \log^\theta n)$ where the dimension $D = O\left(\frac{\log n}{\theta \log\log n}\right)$.*

This implies Theorems 4, 5 and Theorem 9 when taking $\theta = 1/(12 \log\log n)$.

Proof. Let $D = c \cdot \log n/(\theta \log\log n)$ for some constant c to be determined later. Let $\zeta = \frac{1}{\ln^{\theta/3} n}$. We sample for any $t \in [D]$ an embedding $f^{(t)} : X \to \mathbb{R}_+$ as in Lemma 11 with parameter ζ and let $F = D^{-1/p} \bigoplus_t f^{(t)}$. Fix any $\varepsilon > 0$ and let $u, v \in \hat{G}_\varepsilon$. Let $\mathcal{Z}_t = \mathcal{Z}_t(u,v)$ be the indicator for the event $\neg\mathcal{E}_{uv-\mathsf{good}}$, *i.e.* we failed in the t-th coordinate. Let $\mathcal{Z} = \mathcal{Z}(u,v) = \sum_{t \in [D]} \mathcal{Z}_t$. We are interested to bound the probability of the bad event, that $\mathcal{Z} \geq D/2$. Note that $\mathbb{E}[\mathcal{Z}] \leq \zeta D$, so let $a \geq 1$ such that $\mathbb{E}[\mathcal{Z}] = \zeta D/a$. Using Chernoff bound:

$$\Pr[\mathcal{Z} \geq D/2] = \Pr\left[Z > a\mathbb{E}[\mathcal{Z}]/(2\zeta)\right] \leq \left(\frac{e^{a/(2\zeta)-1}}{(a/(2\zeta))^{a/(2\zeta)}} \right)^{\mathbb{E}[\mathcal{Z}]} \leq (2e\zeta)^{D/2} . \tag{12}$$

As $\sqrt{\zeta} = \frac{1}{\ln^{\theta/6} n} \leq \frac{1}{2e}$ it follows that

$$\Pr[\mathcal{Z} \geq D/2] \leq \sqrt{\zeta}^{D/2} = \left(\frac{1}{\ln^{\theta/6} n} \right)^{c \cdot \log n/(\theta \log\log n)} \leq 1/n^3,$$

for large enough constant c.

As there are $\binom{n}{2}$ pairs, by the union bound there is probability at least $1-1/n$ that none of the bad events $\mathcal{Z}(u,v)$ occur, in such a case, using the first property of Lemma 11

$$\begin{aligned}
\|F(u)-F(v)\|_p^p &= D^{-1}\sum_{t\in[D]}|f^{(t)}(u)-f^{(t)}(v)|^p \\
&\leq D^{-1}\cdot D\left(C\left\lceil \ln\left(\frac{n}{|B(u,d(u,v))|}\right)\right\rceil d(u,v)\right)^p \\
&= O((\ln(2/\varepsilon)\cdot d(u,v))^p),
\end{aligned} \tag{13}$$

since by definition of $\hat{G}_\varepsilon$, $|B(u,d(u,v))|\geq \varepsilon n/2$.

Let $S=S(u,v)\subseteq[D]$ be the subset of coordinates in which event $\mathcal{E}_{uv-\mathsf{good}}$ holds, then as $|S|\geq D/2$ and by the second property of Lemma 11

$$\begin{aligned}
\|F(u)-F(v)\|_p^p &= D^{-1}\sum_{t\in[D]}|f^{(t)}(u)-f^{(t)}(v)|^p \\
&\geq D^{-1}\sum_{t\in S}|f^{(t)}(u)-f^{(t)}(v)|^p \\
&\geq D^{-1}|S|\left(\zeta^3 d(u,v)/C\right)^p \\
&= \Omega\left(\frac{d(u,v)}{\ln^\theta n}\right)^p.
\end{aligned}$$

□

4.3 Embedding into l_p

In this section we show the proof of Theorem 11, which gives an improved scaling distortion bound of $O(\lceil\log(2/\epsilon)/p\rceil)$, when embedding into l_p, with the price of higher dimension. As in the previous section, the bulk of the proof is showing an embedding into the line with the desired properties, described in the following lemma.

Lemma 20. *Let (X,d) be a finite metric space on n points and let $\kappa\geq 1$, then there exists a distribution $\mathcal{D}$ over functions $f:X\to\mathbb{R}$ such that for all $\epsilon\in(0,1]$ and all $x,y\in\hat{G}(\epsilon)$:*

1. *For all $f\in\operatorname{supp}(\mathcal{D})$,*

$$|f(x)-f(y)|\leq C\left\lceil \ln\left(\frac{2}{\epsilon}\right)/\kappa+1\right\rceil\cdot d(x,y).$$

2. $$\Pr_{f\sim\mathcal{D}}\left[|f(x)-f(y)|\geq d(x,y)/C\right]\geq\frac{1}{4e^{5\kappa}},$$

where C is a universal positive constant.

The proof of this lemma is in the spirit of Lemma 11, the main difference is that we choose a partition with very small probability of padding, *i.e.* the parameter $\hat{\delta}\approx e^{-\kappa}$. This will improve the distortion by a factor of $\ln(1/\hat{\delta})=\kappa$, but choosing $\hat{\delta}$ in such a way Claim 16 does not hold anymore. There may be pairs x,y such that $\xi_i(x)=\xi_i(y)=0$. For such cases we need to modify f

by adding additional terms that are essentially distances to random subsets of the space, similarly to Bourgain's original embedding, and show that if indeed $\xi_i(x) = \xi_i(y) = 0$ then we can get the contribution from these additional terms.

Let $s = e^{\kappa}$. Let I and Δ_i for $i \in I$ be as in Definition 19. We will define functions $\psi, \mu : X \to \mathbb{R}^+$ and let $f = \psi + \mu$. In what follows we define ψ. We construct a uniformly $(\eta, 1/s)$-padded probabilistic 2-hierarchical partition $\hat{\mathcal{H}}$ as in Lemma 10, and let ξ be as defined in the lemma. Now fix a hierarchical partition $H = \{P_i\}_{i\in I} \in \mathcal{H}$. We define the embedding by defining the coordinates for each $x \in X$. For each $0 < i \in I$ we define a function $\psi_i : X \to \mathbb{R}^+$ and for $x \in X$, let $\psi(x) = \sum_{i\in I} \psi_i(x)$.

Let $\sigma_i : X \to \{0, 1\}$ be a uniform function with respect to P_i define by letting $\{\sigma_i(C)|C \in P_i, 0 < i \in I\}$ be i.i.d symmetric $\{0, 1\}$-valued Bernoulli random variables. The embedding is defined as follows: for each $x \in X$ and $0 < i \in I$ let

$$\psi_i(x) = \sigma_i(x) \cdot \min\left\{\frac{\xi_i(x)}{\kappa \cdot \eta_i(x)} \cdot d(x, X \setminus P_i(x)), \Delta_i\right\}.$$

Next, we define the function μ, based on the embedding technique of Bourgain [Bou85] and its generalization by Matoušek [Mat90]. 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 $A_k \subseteq X$, with each point of X included in A_k independently with probability s^{-k}. For each $k \in K$ and $x \in X$, define:

$$I_k(x) = \{i \in I|\forall u \in P_i(x),\ s^{k-2} < |B(u, 4\Delta_i)| \le s^k\}.$$

We make the following simple observations:

Claim 21. *The following hold for every $i \in I$:*

- *For any $x \in X$: $|\{k|i \in I_k(x)\}| \le 2$.*
- *For every $k \in K$: the function $i \in I_k(x)$ is uniform with respect to P_i.*

We define $i_k : X \to I$, where $i_k(x) = \min\{i|i \in I_k(x)\}$.

For each $k \in K$ we define a function $\mu_k : X \to \mathbb{R}^+$ and let $\mu(x) = \sum_{k\in K} \mu_k(x)$. The function μ_k is defined as follows: for each $x \in X$ and $k \in K$, let $i = i_k(x)$ and

$$\mu_k(x) = \min\left\{\frac{1}{8}d(x, A_k), 2^9 d(x, X \setminus P_i(x)), \Delta_i\right\}.$$

Upper Bound Proof

Claim 22. *For any $i \in I$ and $x, y \in X$,*

$$\psi_i(x) - \psi_i(y) \le \min\left\{\frac{\xi_i(x)}{\kappa \cdot \eta_i(x)} \cdot d(x, y), \Delta_i\right\}.$$

The proof is essentially the same as the proof of Claim 14.

Claim 23. *For any $k \in K$ and $x, y \in X$,*

$$\mu_k(x) - \mu_k(y) \le \min\{2^9 d(x, y), \Delta_{i_k(x)}\}.$$

Proof. Let $i = i_k(x)$ and $i' = i_k(y)$. There are two cases. In Case 1, assume $P_i(x) = P_i(y)$, and first we show that $i = i'$. By Claim 21 we have that $i \in I_k(y)$, implying $i' \le i$. Since $H = \{P_i\}_{i \in I}$ is a hierarchical partition we have that $P_{i'}(x) = P_{i'}(y)$. Hence Claim 21 implies that $i' \in I_k(x)$, so that $i \le i'$, which implies $i' = i$.

Since $\mu_k(x) \le \Delta_i$ we have that $\mu_k(x) - \mu_k(y) \le \mu_k(x) \le \Delta_i$. To prove $\mu_k(x) - \mu_k(y) \le 2^9 d(x,y)$ consider the value of $\mu_k(y)$. If $\mu_k(y) = \frac{1}{8}d(y, A_k)$ then $\mu_k(x) - \mu_k(y) \le \frac{1}{8}(d(x, A_k) - d(y, A_k)) \le \frac{1}{8}d(x,y)$. Otherwise, if $\mu_k(y) = 2^9 d(y, X \setminus P_i(x))$ then

$$\mu_k(x) - \mu_k(y) \le 2^9(d(x, X \setminus P_i(x)) - d(y, X \setminus P_i(x))) \le 2^9 d(x,y).$$

Finally, if $\mu_k(y) = \Delta_i$ then $\mu_k(x) - \mu_k(y) \le \Delta_i - \Delta_i = 0$.

Next, consider Case 2 where $P_i(x) \ne P_i(y)$. In this case we have that $d(x, X \setminus P_i(x)) \le d(x,y)$ which implies that

$$\mu_k(x) - \mu_k(y) \le \mu_k(x) \le \min\{2^9 d(x,y), \Delta_i\} \ .$$

□

Let ℓ be largest such that $\Delta_{\ell+4} \ge d(x,y) \ge \max\{r_{\epsilon/2}(x), r_{\epsilon/2}(y)\}$. If no such ℓ exists then let $\ell = 0$.

By Claim 22 and Lemma 10 we have

$$\begin{aligned}
\sum_{0<i\le\ell} (\psi_i(x) - \psi_i(y)) &\le \sum_{0<i\le\ell} \frac{\xi_i(x)}{\kappa \cdot \eta_i(x)} \cdot d(x,y) \\
&\le 2^{14} \cdot \ln\left(\frac{n}{|B(x, \Delta_{\ell+4})|}\right) \cdot d(x,y)/\kappa \le (2^{14}\ln(2/\epsilon)) \cdot d(x,y)/\kappa.
\end{aligned}$$

We also have that

$$\sum_{\ell<i\in I} (\psi_i(x) - \psi_i(y)) \le \sum_{\ell<i\in I} \Delta_i \le \Delta_\ell \le 2^5 d(x,y).$$

It follows that

$$|\psi(x) - \psi(y)| = |\sum_{0<i\in I} (\psi_i(x) - \psi_i(y))| \le \left(2^{14}\ln(2/\epsilon)/\kappa + 2^5\right) \cdot d(x,y).$$

Let k' be the largest such that $s^{k'} \le \epsilon n/2$. Note that $|\{k \in K \mid k > k'\}| \le \lceil \log_s n \rceil - \lfloor \log_s(\epsilon n/2) \rfloor \le \ln(2/\epsilon)/\kappa + 2$, hence

$$\sum_{k'<k\in K} (\mu_k(x) - \mu_k(y)) \le \sum_{k'<k\in K} 2^9 d(x,y) \le 2^9 \cdot (\ln(2/\epsilon)/\kappa + 2) d(x,y).$$

Now, if $k \le k'$ and $i \in I_k(x)$ then for any $u \in P_i(x)$ we have $|B(x, 2\Delta_i)| \le |B(u, 4\Delta_i)| \le s^k \le \epsilon n/2$. It follows that $d(x,y) \ge r_{\epsilon/2}(x) \ge 2\Delta_i$. Let $\ell' = \min\{i \in I \mid d(x,y) \ge 2\Delta_i\}$. Using Claim 23 and the first property of Claim 21 we get

$$\sum_{k'\ge k\in K} (\mu_k(x) - \mu_k(y)) \le \sum_{k'\ge k\in K} \Delta_{i_k(x)} \le \sum_{\ell'\le i\in I} \sum_{k\in K \mid i\in I_k(x)} \Delta_i \le \sum_{\ell'\le i\in I} 2\Delta_i \le 4\Delta_{\ell'} \le 2d(x,y).$$

It follows that

$$|\mu(x)-\mu(y)| = |\sum_{k\in K}(\mu_k(x)-\mu_k(y))| \le 2^9\left(\ln(2/\epsilon)/\kappa+3\right)\cdot d(x,y).$$

It follows that

$$|f(x)-f(y)| = |\psi(x)+\mu(x)-\psi(y)-\mu(y)| \le 2^{15}\left(\ln(2/\epsilon)/\kappa+1\right)\cdot d(x,y).$$

Lower Bound Proof

Let $0<\ell\in I$ be such that $8\Delta_\ell < d(x,y) \le 16\Delta_\ell$. We distinguish between the following two cases:

- **Case 1:** Either $\xi_\ell(x)=1$ or $\xi_\ell(y)=1$.

  Assume w.l.o.g that $\xi_\ell(x)=1$. Let $\mathcal{E}_{u-\text{pad}}$ be the event that

  $$B(x,\eta_\ell(x)\ln s\cdot\Delta_\ell)\subseteq P_\ell(x).$$

  As $\hat{\mathcal{H}}$ is $(\eta,1/s)$-padded, $\Pr[\mathcal{E}_{u-\text{pad}}]\ge 1/s$, recalling that $\kappa=\ln s$, if this event occurs

  $$\psi_\ell(x)\ge\sigma_\ell(x)\cdot\min\left\{\frac{\xi_\ell(x)}{\kappa\cdot\eta_\ell(x)}\cdot\eta_\ell(x)\kappa\cdot\Delta_\ell,\Delta_\ell\right\}=\sigma_\ell(x)\cdot\Delta_\ell.$$

  Assume that $\mathcal{E}_{u-\text{pad}}$ occurs. Since $\operatorname{diam}(P_\ell(x))\le\Delta_\ell<d(x,y)$ we have that $P_\ell(y)\neq P_\ell(x)$, so the value of $\sigma_\ell(x)$ is independent of the value of $f(y)$. We distinguish between two cases:

  - $|f(x)-f(y)-\psi_\ell(x)|\ge\frac{1}{2}\Delta_\ell$. In this case there is probability $1/2$ that $\sigma_\ell(x)=0$, so that $\psi_\ell(x)=0$.
  - $|f(x)-f(y)-\psi_\ell(x)|\le\frac{1}{2}\Delta_\ell$. In this case there is probability $1/2$ that $\sigma_\ell(x)=1$, so that $\psi_\ell(x)\ge\Delta_\ell$.

  We conclude that with probability at least $1/(2s)$: $|f(x)-f(y)|\ge\frac{1}{2}\Delta_\ell$.

- **Case 2:** $\xi_{P,\ell}(x)=\xi_{P,\ell}(y)=0$

  It follows from Lemma 10 that $\max\{\bar{\rho}(x,2\Delta_\ell,\gamma_1,\gamma_2),\bar{\rho}(y,2\Delta_\ell,\gamma_1,\gamma_2)\}<s$. Let $x'\in B(x,2\Delta_\ell)$ and $y'\in B(y,2\Delta_\ell)$ such that $\rho(x',2\Delta_\ell,\gamma_1,\gamma_2)=\bar{\rho}(x,2\Delta_\ell,\gamma_1,\gamma_2)$ and $\rho(y',2\Delta_\ell,\gamma_1,\gamma_2)=\bar{\rho}(y,2\Delta_\ell,\gamma_1,\gamma_2)$.

  Recall that $\gamma_1=16$, $\gamma_2=1/16$. For $z\in\{x',y'\}$ we have:

  $$s>\rho(z,2\Delta_\ell,\gamma_1,\gamma_2)=\frac{|B(z,32\Delta_\ell)|}{|B(z,2\Delta_\ell/16)|}\ge\frac{|B(x,14\Delta_\ell)|}{|B(z,\Delta_\ell/8)|},$$

  using that $d(x,x')\le 2\Delta_\ell$ and $d(x,y')\le d(x,y)+d(y,y')\le 18\Delta_\ell$, so that $B(x,14\Delta_\ell)\subseteq B(z,32\Delta_\ell)$.

  Let $k\in K$ be such that $s^{k-1}<|B(x,14\Delta_\ell)|\le s^k$. We deduce that for $z\in\{x',y'\}$, $|B(z,\Delta_\ell/8)|>s^{k-2}$. Consider an arbitrary point $u\in P_\ell(x)$, as $d(u,x')\le 3\Delta_\ell$ it follows that $s^{k-2} < |B(u, 4\Delta_\ell)| \le s^k$. This implies that $\ell \in I_k(x)$ and therefore $i_k(x) \le \ell$. As $\hat{\mathcal{H}}$ is $(\eta, 1/s)$-padded we have the following bound

$$\Pr[B(x, \eta_\ell(x) \cdot \kappa\Delta_\ell) \subseteq P_\ell(x)] \ge 1/s.$$

Assume that this event occurs. Since H is hierarchical we get that for every $i \le \ell$, $B(x, \eta_\ell(x) \cdot \kappa\Delta_\ell) \subseteq P_\ell(x) \subseteq P_i(x)$ and in particular this holds for $i = i_k(x)$. As $\xi_\ell(x) = 0$ we have that $\eta_\ell(x) = 2^{-9}/\kappa$. Hence,

$$2^9 \cdot d(x, X \setminus P_i(x)) \ge 2^9 \cdot \eta_\ell(x)\kappa\Delta_\ell = \Delta_\ell.$$

Implying:

$$\mu_k(x) = \min\left\{\frac{1}{8}d(x, A_k), 2^9 \cdot d(x, X \setminus P_i(x)), \Delta_i\right\} \ge \min\left\{\frac{1}{8}d(x, A_k), \Delta_\ell\right\}.$$

The following is a variant on the original argument in [Bou85, Mat90]. Define the events: $\mathcal{A}_1 = B(y', \Delta_\ell/8) \cap A_k \ne \emptyset$, $\mathcal{A}_2 = B(x', \Delta_\ell/8) \cap A_k \ne \emptyset$ and $\mathcal{A}_3 = [B(x, 14\Delta_\ell) \setminus B(y', \Delta_\ell/8)] \cap A_k = \emptyset$. Then for $m \in \{1, 2\}$:

$$\begin{aligned} \Pr[\mathcal{A}_m] &\ge 1 - \left(1 - s^{-k}\right)^{s^{k-2}} \ge 1 - e^{-s^{-k} \cdot s^{k-2}} = 1 - e^{-s^{-2}} \ge s^{-2}/2, \\ \Pr[\mathcal{A}_3] &\ge \left(1 - s^{-k}\right)^{s^k} \ge 1/4, \end{aligned}$$

using $s \ge 2$. Observe that $d(x', y') \ge d(x, y) - d(x, x') - d(y, y') \ge d(x, y) - 4\Delta_\ell \ge 4\Delta_\ell$, implying $B(y', \Delta_\ell/8) \cap B(x', \Delta_\ell/8) = \emptyset$. It follows that event $\mathcal{A}_1$ is independent of either event $\mathcal{A}_2$ or $\mathcal{A}_3$.

Assume event $\mathcal{A}_1$ occurs. It follows that $d(y, A_k) \le d(y, y') + \Delta_\ell/8 \le \frac{17}{8}\Delta_\ell$. We distinguish between two cases:

- $|f(x) - f(y) - (\mu_k(x) - \mu_k(y))| \ge \frac{3}{8}\Delta_\ell$. In this case there is probability at least $s^{-2}/2$ that event $\mathcal{A}_2$ occurs, in such a case $d(x, A_k) \le d(x, x') + \Delta_\ell/8 \le \frac{17}{8}\Delta_\ell$ so that $|\mu_k(x) - \mu_k(y)| \le \frac{1}{8}\max\{d(x, A_k), d(y, A_k)\} \le \frac{17}{64}\Delta_\ell$. We therefore get with probability at least $s^{-2}/2$ that $|f(x) - f(y) \ge \frac{24}{64}\Delta_\ell - \frac{17}{64}\Delta_\ell \ge \Delta_\ell/10$.
- $|f(x) - f(y) - (\mu_\ell(x) - \mu_\ell(y))| < \frac{3}{8}\Delta_\ell$. In this case there is probability at least $1/4$ that event $\mathcal{A}_3$ occurs. Observe that:

$$\begin{aligned} d(x, B(y', \Delta_\ell/8)) &\ge d(x, y) - d(y, y') - \Delta_\ell/8 \\ &\ge 8\Delta_\ell - 2\Delta_\ell - \Delta_\ell/8 = \frac{47}{8}\Delta_\ell, \end{aligned}$$

  implying that $d(x, A_k) \ge \min\{14\Delta_\ell, d(x, B(y', \Delta_\ell/8))\} \ge \frac{47}{8}\Delta_\ell$ and therefore $\mu_k(x) \ge \min\{\frac{1}{8} \cdot \frac{47}{8}\Delta_\ell, \Delta_\ell\} = \frac{47}{64}\Delta_\ell$. Since $\mu_k(y) \le \frac{1}{8}d(y, A_k) \le \frac{17}{64}\Delta_\ell$ we obtain that: $\mu_k(x) - \mu_k(y) \ge \frac{30}{64}\Delta_\ell$. We therefore get with probability at least $1/4$ that $|f(x) - f(y)| \ge \frac{30}{64}\Delta_\ell - \frac{3}{8}\Delta_\ell \ge \Delta_\ell/10$.

We conclude that given events $\mathcal{E}_{u-\text{pad}}$ and $\mathcal{A}_1$, with probability at least $s^{-2}/2$: $|f(x) - f(y)| \ge \Delta_\ell/10$.

It follows that with probability at least $s^{-5}/4$:

$$|f(x) - f(y)| \geq \Delta_\ell/10 \geq d(x,y)/160.$$

This concludes the proof of Lemma 20.

Proof of Theorem 11. Fix some $1 \leq p < \infty$. [21] Let $D = e^{cp}\ln n$ for a universal constant c and define $F : X \to l_p^D$ by $F(x) = D^{-1/p} \bigoplus_{t=1}^{D} f^{(t)}(x)$ where each $f^{(t)}$ is sampled as in Lemma 20. Let $x, y \in \hat{G}(\epsilon)$, then by the first property of the lemma

$$\|F(x) - F(y)\|_p^p = D^{-1} \sum_{t=1}^{D} |f^{(t)}(x) - f^{(t)}(y)|^p \leq (C\ln(2/\epsilon)/\kappa + 1)^p\, d(x,y)^p$$

Let $Z_t(x,y)$ be an indicator random variable for the event that $|f^{(t)}(x) - f^{(t)}(y)| \geq d(x,y)/C$, and $Z = Z(x,y) = \sum_{t\in[D]} Z_t(x,y)$. By the second property of the lemma, for any $t \in [D]$, $\Pr[Z_t(x,y)] \geq \frac{1}{4e^{5\kappa}}$, thus $\mathbb{E}[Z] \geq \frac{D}{4e^{5\kappa}} \geq 16\ln n$ for constant $c \geq 11$ (and using that $\kappa \leq p$). By Chernoff bound

$$\Pr[Z < \mathbb{E}[Z]/2] \leq e^{-E[Z]/8} \leq 1/n^2\ .$$

Observe that if $Z \geq \mathbb{E}[Z]/2$ then if we write $G(x,y) = \{t \in [D] \mid Z_t(x,y)\}$, it holds that $|G(x,y)| \geq \frac{D}{8e^{5\kappa}}$ and then

$$\|F(x) - F(y)\|_p^p \geq \frac{1}{D} \sum_{t\in G(x,y)} |f^{(t)}(x) - f^{(t)}(y)|^p \geq \frac{1}{8e^{5\kappa}}\,(d(x,y)/C)^p \geq \left(\frac{d(x,y)}{8Ce^5}\right)^p.$$

The proof is concluded by applying the union bound over the $\binom{n}{2}$ pairs. $\square$

5 Extending to Infinite Compact Spaces

In this section we extend our main result to infinite compact spaces. In what follows (X,d) is a compact metric space equipped with a probability measure σ. Our aim is to bound the ℓ_q-distortion of embedding X into l_p spaces by $O(q)$, and as before the initial step is to bound the scaling distortion.

Theorem 12. *Let $1 \leq p \leq \infty$ and let (X,d) be a compact metric space. There exists an embedding $f : X \to l_p$ having coarsely scaling distortion $O(\lceil(\log\frac{2}{\epsilon})\rceil)$. For any $1 \leq q < \infty$, the ℓ_q-distortion of this embedding is:* $\mathrm{dist}_q(f) = O(q)$.

5.1 Uniform Probabilistic Partitions for Infinite Spaces

For infinite metric spaces case we require an extension of the definition of local growth rate, which can also be infinite.

[21] For $p = \infty$ there is an isometric embedding using $n-1$ or less dimensions, in particular the Fréchet embedding.

Definition 22. The local growth rate of $x \in X$ at radius $r > 0$ for given scales $\gamma_1, \gamma_2 > 0$ is defined as

$$\rho(x, r, \gamma_1, \gamma_2) = \begin{cases} \frac{\sigma(B(x, r\gamma_1))}{\sigma(B(x, r\gamma_2))} & \sigma(B(x, r\gamma_2)) > 0 \\ \infty & \sigma(B(x, r\gamma_2)) = 0 \end{cases}$$

$\bar{\rho}$ is defined as before.

The definitions for padded partitions remain the same, as the proof of Lemma 4. Now the partition lemma will be the following

Lemma 24. *Let (X, d) be a compact metric space. Let $Z \subseteq X$. Let $0 < \Delta \leq \operatorname{diam}(Z)$. Let $\hat{\delta} \in (0, 1/2]$, $\gamma_1 \geq 2$, $\gamma_2 \leq 1/16$. There exists a Δ-bounded probabilistic partition $\hat{\mathcal{P}}$ of (Z, d) and a collection of uniform functions $\{\xi_P : Z \to \{0, 1\} \mid P \in \mathcal{P}\}$ and $\{\eta_P : Z \to (0, 1] \mid P \in \mathcal{P}\}$ such that the probabilistic partition $\hat{\mathcal{P}}$ is a strong $(\eta, \hat{\delta})$-uniformly locally padded probabilistic partition; and the following conditions hold for any $P \in$ supp$(\hat{\mathcal{P}})$ and any $x \in Z$:*

- *If $\xi_P(x) = 1$ then:*
  - *If $\rho(x, 2\Delta, \gamma_1, \gamma_2) < \infty$ then $2^{-6}/\ln \rho(x, 2\Delta, \gamma_1, \gamma_2) \leq \eta_P(x) \leq 2^{-6}/\ln(1/\hat{\delta})$.*
  - *Otherwise, when $\rho(x, 2\Delta, \gamma_1, \gamma_2) = \infty$ then $\eta_P(x) = 0$.*
- *If $\xi_P(x) = 0$ then: $\eta_P(x) = 2^{-6}/\ln(1/\hat{\delta})$ and $\bar{\rho}(x, 2\Delta, \gamma_1, \gamma_2) < 1/\hat{\delta}$.*

Our partition algorithm will be similar to the one of Lemma 5. First we deterministically assign a set of centers $C = \{v_1, v_2, \ldots, v_s\} \subseteq Z$ and parameters $\chi_1, \chi_2, \ldots, \chi_s \in \mathbb{R}_+ \cup \{\infty\}$. Let $W_1 = Z$ and $j = 1$. Conduct the following iterative process:

1. Let $v_j \in W_j$ be the point minimizing $\hat{\chi_j} = \rho(x, 2\Delta, \gamma_1, \gamma_2)$ over all $x \in W_j$.
2. Set $\chi_j = \max\{2/\hat{\delta}^{1/2}, \hat{\chi_j}\}$.
3. Let $W_{j+1} = W_j \setminus B(v_j, \Delta/4)$.
4. Set $j = j + 1$. If $W_j \neq \emptyset$ return to 1.

One observation we require is that the number s of cluster centers in every partition is indeed finite, using the following claim:

Claim 25. *For any $\Delta > 0$ and the algorithm described above, there exists some $s \in \mathbb{N}$ such that $W_s = \emptyset$.*

Proof. Since the metric is compact by definition it is also totally bounded (*i.e.* for every $r > 0$ there exists a finite cover of X with balls of radius at most r). The algorithm starts by assigning a set of centers C that are actually a $\Delta/4$-net, and we can show that this net is finite. Take $r = \Delta/8$ and consider the finite cover with balls of radius at most r. Every net point c must be covered by this cover, so there is a ball B_c in the cover with center x such that $d(x, c) < r$, which implies that these balls B_c are distinct for every $c \in C$, so as the cover is finite also C is finite. □

Let $t \leq s$ be the minimal index such that $\chi_t = \infty$. Now the algorithm for the partition and functions ξ, η is as follows: Let $Z_1 = Z$. For $j = 1, 2, \ldots, t - 1$:

1. Let $(S_{v_j}, \bar{S}_{v_j})$ be the partition created by $S_{v_j} = B_{Z_j}(v_j, r)$ and $\bar{S}_{v_j} = Z_j \setminus S_{v_j}$ where r is distributed as in Lemma 4 with parameter $\lambda = 8\ln(\chi_j)/\Delta$.

2. Set $C_j = S_{v_j}$, $Z_{j+1} = \bar{S}_{v_j}$.

3. For all $x \in C_j$ let $\eta_P(x) = 2^{-6}/\max\{\ln \hat{\chi}_j, \ln(1/\hat{\delta})\}$. If $\hat{\chi}_j \geq 1/\hat{\delta}$ set $\xi_P(x) = 1$, otherwise set $\xi_P(x) = 0$.

For $j = t, t+1 \ldots s$:

1. Let $C_j = B_{Z_j}(v_j, \Delta/4)$, $Z_{j+1} = Z_j \setminus C_j$.

2. For all $x \in C_j$ let $\eta_P(x) = 0$, $\xi_P(x) = 1$.

The proof remains essentially the same, replacing every $|B(x,r)|$ by $\sigma(B(x,r))$ in the part that bounds $\sum_{j\geq 1, v_j \in T} \chi_j^{-1}$. It is easy to see that the padding analysis of Lemma 5 still holds for all points $x \in C_j$ where $j < t$, and it will hold for $j \geq t$ since for such points $\eta_P(x) = 0$, which means that we need to pad a ball of radius 0, so the padding probability is 1, and the other properties are easily checked.

5.2 Embedding Infinite Spaces into l_p

As in the finite case, we first construct an embedding into the real line, that is good in expectation.

Lemma 26. *Let (X, d) be a compact metric space with diameter Δ and let $0 < \zeta \leq 1$, then there exists a distribution $\mathcal{D}$ over functions $f : X \to \mathbb{R}$ such that for all $u, v \in X$:*

1. *For all $f \in \operatorname{supp}(\mathcal{D})$,*

   - *If there exists $\epsilon > 0$ such that $u, v \in \hat{G}(\epsilon)$*

   $$|f(u) - f(v)| \leq C \left\lceil \ln\left(\frac{2}{\sigma(B(u, d(u,v)))}\right)\right\rceil \cdot d(u,v).$$

   - *Otherwise*

   $$|f(u) - f(v)| \leq \Delta \ .$$

2. $$\Pr_{f\sim\mathcal{D}}\left[|f(u) - f(v)| \geq \zeta^3 \cdot d(u,v)/C\right] \geq 1 - \zeta,$$

   where C is a universal positive constant.

Proof. The proof of the lemma is very similar to the proof of Lemma 11, we highlight the main differences.

The embedding f is defined as in Lemma 11, where $\varphi_i : X \to \mathbb{R}^+$ is defined as

$$\varphi_i(x) = \begin{cases} \min\left\{\frac{\xi_i(x)}{\eta_i(x)} d\left(x, X \setminus P_i(x)\right), \zeta\Delta_i/4\right\} & \eta_i(x) > 0 \\ \zeta\Delta_i/4 & \eta_i(x) = 0 \end{cases} \tag{14}$$

For the upper bound proof, fix a pair $u, v \in X$ such that $u, v \in \hat{G}(\epsilon)$ for $\epsilon > 0$. Then both $\sigma(B(u, d(u,v))), \sigma(B(v, d(u,v))) > 0$. The proof of Lemma 12 will still hold for such u, v by the same argument shown there, just replacing the size of a ball by its measure. This is true because the choice of scale ℓ was such that the growth rate is indeed finite $\rho(u, 2\Delta_i, \gamma_1, \gamma_2) < \infty$ for all $i < \ell$.

For any pair $u, v \in X$, we have that $\psi_i(u) - \psi_i(u) \le \zeta\Delta_i/4$, hence

$$|f(u) - f(v)| = |\sum_{i>0} \psi_i(u) - \psi_i(v)| \le \zeta/4 \sum_{i>0} \Delta_i < \Delta_0.$$

The proof of the lower bound is essentially the same as in Lemma 15.

□

Proof of Theorem 12. Define the embedding $F : X \to L_p(\mathcal{D})$ as a convex direct sum of all $f \in$ *supp*$(\mathcal{D})$, each f is naturally weighted by $\Pr(f)$. It can be seen that $\|F(x)-F(y)\|_p^p = \mathbb{E}_{f\sim\mathcal{D}}[|f(x) - f(y)|^p]$, hence applying Lemma 26 with $\zeta = 1/2$ we get that for any $\epsilon > 0$ and all $(u, v) \in \hat{G}(\epsilon)$,

$$\text{dist}_F(u, v) \le O(\log(2/\epsilon)) \ .$$

□

5.3 Scaling Distortion Vs. ℓ_q-distortion for Infinite Spaces

The main difference from the proof of Lemma 1 is that not all pairs $u, v \in X$ have an $\epsilon > 0$ such that $(u, v) \in \hat{G}(\epsilon)$. This means in particular that having scaling distortion gives no guarantees on the distortion of such pairs. Luckily, the measure of the set of such pairs is zero, hence it is enough to obtain for every pair some finite bound on the distortion.

Let C be the universal constant of the distortion. Let $G_i = \hat{G}(2^{-i}) \setminus \hat{G}(2^{-(i-1)})$ and $G_\infty = \binom{X}{2} \setminus \left(\bigcup_{\epsilon>0} \hat{G}(\epsilon)\right)$, and note that for all $x \in X$ if $G_i(x) = \{y \in X \mid (x, y) \in G_i\}$ then $\sigma(G_i(x)) \le 2^{-(i-1)}$, hence $\Pi(G_i) = \int_x \int_y 1_{y\in G_i(x)} d\sigma d\sigma \le 2^{-(i-1)}$. Also note that as $\Pi(\hat{G}(\epsilon)) \ge 1 - \epsilon/2$, we have that $\Pi(G_\infty) = 0$. We can now bound the ℓ_q-distortion as follows:

$$\begin{aligned}
\mathbb{E}_{(x,y)\sim\Pi}[\text{dist}_F(x, y)^q]^{1/q} &= \left(\int_x \int_y \text{dist}_F(x, y)^q d\sigma d\sigma\right)^{1/q} \\
&= \left(\int_x \left(\sum_{i=1}^{\infty} \int_{y\in G_i(x)} \text{dist}_F(x, y)^q d\sigma + \int_{y\in G_\infty(x)} \text{dist}_F(x, y)^q d\sigma\right) d\sigma\right)^{1/q} \\
&\le 2C \left(\int_x \left(\sum_{i=1}^{\infty} \int_{y\in G_i(x)} (\log(2^i))^q d\sigma + 0\right) d\sigma\right)^{1/q} \\
&\le 2C \left(\sum_{i=1}^{\infty} \frac{i^q}{2^i}\right)^{1/q} \\
&= O(q).
\end{aligned}$$

Given weights $w : X \times X \to \mathbb{R}_+$ on the pairs such that $\int_x \int_y w(x, y)\Pi(x, y)d\sigma d\sigma = 1$, an analogous calculation to the finite case also bound the weighted ℓ_q-distortion by $O(q + \log \hat{\Phi}(w))$ (above was shown the case that for all $x, y \in X$, $w(x, y) = 1$).

6 Embedding of Doubling Metrics

In this section we focus on metrics with bounded doubling constant λ (recall Definition 11). The main result of this section is a low distortion embedding of metric spaces into l_p of dimension $O(\log\lambda)$. Other results shown here are an extension to scaling distortion, which implies constant average distortion with low dimension $\tilde{O}(\log\lambda)$, a distortion-dimension tradeoff for doubling metrics and "snow-flake" embedding in the spirit of Assouad.

6.1 Low Dimensional Embedding for Doubling Metrics

Theorem 8. *There exists a universal constant C such that for any n-point metric space (X,d) and any $C/\log\log n < \theta \le 1$, there exists an embedding $f : X \to l_p^D$ with distortion $O\left(\log^{1+\theta} n\right)$ where $D = O\left(\frac{\dim(X)}{\theta}\right)$.*

One can take θ to be any small positive constant and obtain low distortion in the (asymptotically) optimal dimension. Another interesting choice is to take $\theta = O(1/\log\log n)$, and get the standard $O(\log n)$ distortion with only $O(\log\lambda \cdot \log\log n)$ dimensions.

The proof is also based on the embedding into the line of Lemma 11, with the parameter ζ being much smaller. The analysis uses nets of the space for each scale, which is standard technique for doubling metrics, then argues that it is enough to have a successful embedding only for certain pairs of points in the net in order to have a successful embedding for all pairs. The low dimension is then obtained by arguing that there are few dependencies between the relevant pairs of points in the nets, and then using Lovasz local lemma in order to show that small number of dimensions is sufficient to obtain a positive success probability for all relevant pairs in the nets. W.l.o.g we may assume that n is larger than some absolute constant. Now for the formal proof:

Let $\lambda = 2^{\dim(X)}$ and $D = \lceil (c\log\lambda)/\theta \rceil$ for some constant c to be determined later. Let $\zeta = \frac{1}{\ln^{\theta/3} n}$ and let C be the constant from Lemma 11. For any $t \in [D]$ let $f^{(t)} : X \to \mathbb{R}_+$ be an embedding as in Lemma 11 with parameter ζ (the exact choice of $f^{(t)}$ will be determined later), and let $F = D^{-1/p} \bigoplus_{t=1}^{D} f^{(t)}$. Fix any $\varepsilon > 0$ and let $x, y \in \hat{G}_\varepsilon$. Recall that $\Delta_0 = \mathrm{diam}(X)$ and for $i > 0$, $\Delta_i = (\zeta/8)^i \Delta_0$. By the same calculation as in (13) we have that

$$\|F(x) - F(y)\|_p = O(\ln(2/\varepsilon) \cdot d(u,v)).$$

The proof on the contraction of the embedding uses a set of nets of the space. For any $i \in \mathbb{N}$, let N_i be a $\frac{\zeta^3\Delta_i}{C^2 \ln n}$-net of X. Let $M \subseteq \binom{X}{2}$ be the set of net pairs for which we would like the embedding to give the distortion bound, formally $M = \left\{(u,v) \in \binom{X}{2} \mid \exists i \in \mathbb{N} :\ u,v \in N_i, 7\Delta_i \le d(u,v) \le 9\Delta_{i-1}\right\}$. Recall from Lemma 11 that $\mathcal{E}^{(t)}_{uv-\mathsf{good}}$ stands for the event that there is sufficient contribution for the pair u,v in coordinate $t \in [D]$ (see proof of Lemma 15 for precise definition). For all $(u,v) \in M$, let $\mathcal{E}_{(u,v)}$ be the event that $\mathcal{E}^{(t)}_{uv-\mathsf{good}}$ holds for at least $D/2$ of the coordinates $t \in [D]$. Define the event $\mathcal{E} = \bigcap_{(u,v)\in M} \mathcal{E}_{(u,v)}$ that captures the case that all pairs in M have the desired property. The main technical lemma is that $\mathcal{E}$ occurs with non-zero probability:

Lemma 27. $\Pr[\mathcal{E}] > 0$.

Let us first show that if the event $\mathcal{E}$ took place, then the contraction of *every* pair $x, y \in X$ is bounded. Let $i = s(x,y)$ (recall that $i = s(x,y)$ uniquely satisfy $8\Delta_i \le d(x,y) < 8\Delta_{i-1}$). Consider

$u,v \in N_i$ satisfying $d(x,u) = d(x,N_i)$ and $d(y,v) = d(y,N_i)$, then $d(u,v) \le d(x,y) + d(u,x) + d(y,v) \le 8\Delta_{i-1} + 2\Delta_i \le 9\Delta_{i-1}$ and $d(u,v) \ge d(x,y) - d(x,u) - d(y,v) \ge 8\Delta_i - 2\frac{\Delta_i}{C^2} \ge 7\Delta_i$, so by the definition of M follows that $(u,v) \in M$. The next claim shows that since x,y are very close to u,v respectively, then by the expansion upper bound $F(x)$ and $F(y)$ will be close to $F(u)$ and $F(v)$ respectively, therefore a lower bound is obtained.

Claim 28. *Let $x,y,u,v \in X$ be as above, then given $\mathcal{E}$:*

$$\|F(x) - F(y)\|_p \ge \zeta^3 d(x,y)/(12C) \ .$$

Proof. First note that if event $\mathcal{E}_{(u,v)}$ holds then letting $S \subseteq [D]$ be the subset of good coordinates for u,v, by Lemma 11 in each good coordinate there is contribution of at least $\zeta^3 d(u,v)/C$, and since there are at least $D/2$ good coordinates,

$$\|F(u) - F(v)\|_p^p \ge D^{-1} \sum_{t \in S} |f^{(t)}(u) - f^{(t)}(v)|^p \ge (\zeta^3 d(u,v)/(2C))^p. \tag{15}$$

Since N_i is $\frac{\zeta^3 \Delta_i}{C^2 \ln n}$-net, then $d(x,u) \le \frac{\zeta^3 \Delta_i}{C^2 \ln n}$. By the first property of Lemma 11,

$$\|F(x) - F(u)\|_p^p = D^{-1} \sum_{t=1}^{D} |f^{(t)}(x) - f^{(t)}(u)|^p \le (C \ln n \cdot d(x,u))^p \le (\zeta^3 \Delta_i / C)^p \le (\zeta^3 d(x,y)/(8C))^p$$

using that $\Delta_i \le d(x,y)/8$. Similarly $\|F(y) - F(v)\|_p \le \zeta^3 d(x,y)/(8C)$, then

$$\begin{aligned}
\|F(x) - F(y)\|_p &= \|F(x) - F(u) + F(u) - F(v) + F(v) - F(y)\|_p \\
&\ge \|F(u) - F(v)\|_p - \|F(x) - F(u)\|_p - \|F(y) - F(v)\|_p \\
&\ge \zeta^3 d(u,v)/(2C) - 2 \cdot \zeta^3 d(x,y)/(8C) \\
&\ge \zeta^3 d(x,y)/(12C),
\end{aligned}$$

where the second inequality follows using (15) and the last inequality follows using that $d(u,v) \ge 2d(x,y)/3$. ☐

6.1.1 Proof of Lemma 27

We begin with a variation of Lovasz local lemma in which the bad events have rating, and events may only depend on other events with equal or larger rating. See the general case of this lemma Lemma 40 for a proof.

Lemma 29 (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 directed dependency graph $G = (V, E)$, where $V = \{\mathcal{E}_{(u,v)} \mid (u, v) \in M\}$, and the rating of a vertex $c(\mathcal{E}_{(u,v)}) = s(u, v)$. Define that $(\mathcal{E}_{(u,v)}, \mathcal{E}_{(u',v')}) \in E$ iff $i = s(u, v) = s(u', v')$ and $d(\{u, v\}, \{u', v'\}) \leq 4\Delta_i$.

Claim 30. *The out-degree of G is bounded by $\lambda^{15 \ln \ln n}$.*

Proof. Fix $\mathcal{E}_{(u,v)} \in V$, we bound the number of pairs $u', v' \in M$ such that $(\mathcal{E}_{(u,v)}, \mathcal{E}_{(u',v')}) \in E$.

Since $i = s(u, v) = s(u', v')$ we have that $8\Delta_i \leq d(u, v), d(u', v') < 8\Delta_{i-1}$, hence if $(u, v) \in N_{i'}$ or $(u', v') \in N_{i'}$ then i' satisfies $i-1 \leq i' \leq i+1$ by the definition of M, so let $N = N_{i-1} \cup N_i \cup N_{i+1}$. Assume w.l.o.g $d(u, u') \leq 4\Delta_i$, hence $d(u, v') \leq d(u, u') + d(u', v') \leq 4\Delta_i + 8\Delta_{i-1} \leq \Delta_{i-2}$ and it follows that $u, v, u', v' \in B = B(u, \Delta_{i-2})$. The number of pairs can be bounded by $|N \cap B|^2$. Since (X, d) is λ-doubling, the ball B of radius $r_1 = (8/\zeta)^2 \Delta_i$ can be covered by $A = \lambda^{\lceil \log(r_1/r_2) \rceil}$ balls of radius $r_2 = \frac{\zeta^4 \Delta_i}{16C^2 \ln n}$, and $A \leq \lambda^{8+2\log C + \log \ln n + \log(1/\zeta^6)}$. Each of these small balls of radius r_2 contains at most one point in the net N_{i+1}. Recall that $\zeta = \frac{1}{\ln^{\theta/3} n}$, so assuming n is large enough it follows that $|N \cap B|^2 \leq |N_{i-1} \cap B|^2 + |N_i \cap B|^2 + |N_{i+1} \cap B|^2 \leq \lambda^{15 \ln \ln n}$. □

The construction of the graph is based on the proposition that pairs of net points that do not have an edge connecting them in G, either have different critical scales or they have the same scale i but are farther than $\approx \Delta_i$ apart, and hence do not change each other's bound on their success probability. Indeed by Lemma 18 the bound on the probability of some event $\mathcal{E}(u, v)$ still holds given any outcome for events $\mathcal{E}(u', v')$ of smaller or equal rating such that $(\mathcal{E}_{(u,v)}, \mathcal{E}_{(u',v')}) \notin E$.

Claim 31.

$$\Pr\left[\neg\mathcal{E}_{(u,v)} \mid \bigwedge_{(u',v') \in Q} \mathcal{E}_{(u',v')}\right] \leq \lambda^{-16 \ln \ln n},$$

for all $Q \subseteq \left\{(u', v') \mid s(u, v) \geq s(u', v') \wedge \left(\mathcal{E}_{(u,v)}, \mathcal{E}_{(u',v')}\right) \notin E\right\}$.

Proof. By Lemma 18 for all $t \in [D]$

$$\Pr\left[\neg\mathcal{E}^{(t)}_{uv-\mathsf{good}} \mid \bigwedge_{(u',v') \in Q} \mathcal{E}_{(u',v')}\right] \leq \zeta.$$

It follows from Chernoff bound (similarly to (12)) that the probability that more than $D/2$ coordinates fail is bounded above by:

$$\Pr\left[\neg\mathcal{E}_{(u,v)} \mid \bigwedge_{(u',v') \in Q} \mathcal{E}_{(u',v')}\right] \leq \sqrt{\zeta}^{D/2} \leq \lambda^{-16 \ln \ln n}, \tag{16}$$

where the last inequality hold as $D = \lceil (c \log \lambda)/\theta \rceil$, and c is a sufficiently large constant. □

Apply Lemma 29 to the graph G we defined, by Claim 30 let $d = \lambda^{15 \ln \ln n}$ and by Claim 31 we can let $p = \lambda^{-16 \ln \ln n}$ satisfying the first condition of Lemma 29. It is easy to see that the second condition also holds (since $\lambda \geq 2$ and assuming $\ln \ln n \geq 2$). Therefore $\Pr[\mathcal{E}] = \Pr[\bigwedge_{(u,v) \in M} \mathcal{E}_{(u,v)}] > 0$, which concludes the proof of Lemma 27.

6.2 Low Dimensional Embedding of Doubling Metrics with Scaling Distortion

In this section we show an extension of the previous result to embedding with the scaling distortion property.

Theorem 16. *For any λ-doubling metric space (X, d) there exists an embedding $f : X \to l_p^D$ with coarse scaling distortion $O\left(\log^{26}(\frac{1}{\epsilon})\right)$ where $D = O(\log \lambda \log\log \lambda)$.*

6.2.1 Proof overview

We highlight the differences between the proof of Theorem 8 and Theorem 16. We assume the reader is familiar with the proof of Theorem 8.

1. The main difference is that in the analysis of the lower bound, a contribution for a pair is "looked for" in one of many scales, instead of examining a single critical scale.

2. We partition the possible $\epsilon \in (0, 1]$ values into $\approx \log\log\log n$ buckets (see equation 17 and definition of ϵ_k). For each scale Δ_i and each of the $\approx \log\log\log n$ possible values of ϵ we build a $\approx \Delta_i/\mathrm{polylog}(\lambda, 1/\epsilon)$-net.

   A naive approach would be to assign separate coordinates for each ϵ_k and increase the dimension and hence the distortion by a factor of $\log\log\log n$. To avoid paying this $\log\log\log n$ factor we sieve the nets $\bar{N}_k^i$ in a subtle manner (see definition of N_k^i for details).

3. The local growth rate of each node is defined with respect to some ϵ value in non standard manner - this is done so that for sufficiently many levels (as a function of ϵ) there will be a density change. This is defined by $\gamma_1(x, i)$.

4. A pair with distance $\approx \Delta_i$ and epsilon that falls into bucket k (hence $k \approx \log\log(2/\epsilon)$) "looks" for a contribution in the levels $i + k/2, \ldots, i + k$, see the definition of $\hat{\mathcal{E}}_{(i,k,u,v)}$ for details. This is necessary to avoid collisions between contributing scales of pairs with different ϵ values.

5. Showing independence of lower bound successes between two pairs is technical and relies on the sieving process. For a pair u, v related to a net N_i^k the scales examined are $\approx i + k/2, \ldots i + k$. Claim 44 shows that examining only these scales ensures that u, v are independent of a pair u', v' if one of the following occurs (1) u', v' belong to a different scale than that of u, v; (2) u', v' are far enough from u, v in the metric space; (3) u', v' has a different ϵ_k value from that of u, v.

6. Proving that all pairs have the desired scaling distortion given that the sieved net points N_k^i have this property is more involved now since it depends on the ϵ, see Lemma 37.

7. The application of the local lemma is complicated due to two issues - (1) we use the general case (2) we do not proceed simply from scale i to scale $i + 1$, but rather use the ranking function in a non-trivial manner, see proof of Lemma 34.

6.2.2 The proof

Let C be a constant to be defined later, and $D = C \log\lambda \log\log\lambda$. Let $\Delta_0 = \mathrm{diam}(X)$, $I = \{i \in \mathbb{Z} \mid 1 \le i \le (\log\Delta_0 + \log\log n)/3\}$. For $i \in I$ let $\Delta_i = \Delta_0/8^i$. By Lemma 8 we have that (X, d) is locally τ-decomposable for $\tau = 2^{-6}/\log\lambda$.

Define an ϵ-value for every point in every scale $i \in I$. The idea is that the number of scales we seek contribution from depends on the density around the point in scale i, so the growth rate ratio must be defined beforehand with respect to this density. Let $c = 12$. For any $i \in I$, $x \in X$ let $\epsilon_i(x) = |B(x, 2\Delta_i)|/n$, and let $\gamma_1(x, i) = 8^{2c+4}\log^{2c}(64/\epsilon_i(x))$. Fix $\gamma_2 = 1/16$. We shall define the embedding f by defining for each $1 \le t \le D$, a function $f^{(t)} : X \to \mathbb{R}^+$ and let $f = D^{-1/p} \bigoplus_{1\le t\le D} f^{(t)}$.

Fix t, $1 \le t \le D$. In what follows we define $f^{(t)}$. For each $0 < i \in I$ construct a Δ_i-bounded $(\eta_i, 1/2)$-padded probabilistic partition $\hat{\mathcal{P}}_i$, as in Lemma 7 with parameters τ, $\gamma_1(\cdot, i), \gamma_2$ and $\hat{\delta} = 1/2$. Fix some $P_i \in \mathcal{P}_i$ for all $i \in I$.

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 $\phi_i^{(t)}(x) = \xi_{P,i}(x)\eta_{P,i}(x)^{-1}$.

Claim 32. *For any $x \in X$, $1 \le t \le D$ and $i \in I$ we have*

$$\sum_{j\le i} \phi_j^{(t)}(x) \le c2^9 \log^2\left(\frac{n}{|B(x, \Delta_{i+1})|}\right).$$

Proof.

$$\begin{aligned}
\sum_{j\le i}\phi_j(x) &= \sum_{j\le i:\xi_j(x)=1} \eta_j^{-1}(x) \le \sum_{j\le i:\xi_j(x)=1} 2^7 \log\rho(x, 2\Delta_j, \gamma_1(x, j), \gamma_2) \\
&\le 2^7 \sum_{j\le i} \sum_{h=\lfloor -\log_8(\gamma_1(x,j))\rfloor}^{1} \log\left(\frac{|B(x, 8\Delta_{j+h})|}{|B(x, \Delta_{j+h})|}\right) \\
&\le 2^7 \sum_{h=\lfloor -2c-4-2c\log\log(2/\epsilon_i(x))\rfloor}^{1} \sum_{j\le i} \log\left(\frac{|B(x, 8\Delta_{j+h})|}{|B(x, \Delta_{j+h})|}\right) \\
&\le 2^7\left(4c\left(1 + \log\log\left(\frac{n}{|B(x, \Delta_{i+1})|}\right)\right)\right)\log\left(\frac{n}{|B(x, \Delta_{i+1})|}\right) \\
&\le c2^9\log^2\left(\frac{n}{|B(x, \Delta_{i+1})|}\right).
\end{aligned}$$

$\square$

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 $f_i^{(t)}(x) = \sigma_i^{(t)}(P_i(x)) \cdot \min\{\phi_i^{(t)}(x) \cdot d(x, X \setminus P_i(x)), \Delta_i\}$.

Lemma 33. *There exists a universal constant $C_1 > 0$ such that for any $(x, y) \in \hat{G}(\epsilon)$:*

$$|f^{(t)}(x) - f^{(t)}(y)| \le C_1 \log^2(2/\epsilon) \cdot d(x, y).$$

Proof. Define ℓ to be largest such that $\Delta_{\ell+1} \geq d(x,y) \geq \max\{r_{\epsilon/2}(x), r_{\epsilon/2}(y)\}$. If no such ℓ exists then let $\ell = 0$.

By Claim 32 we have

$$\begin{aligned}\sum_{0<i\leq\ell} (f_i^{(t)}(x) - f_i^{(t)}(y)) &\leq \sum_{0<i\leq\ell} \phi_i^{(t)}(x) \cdot d(x,y) \\ &\leq c2^9 \log^2\left(\frac{n}{|B(x,\Delta_{\ell+1})|}\right) \cdot d(x,y) \\ &\leq c2^9 \log^2(2/\epsilon) \cdot d(x,y).\end{aligned}$$

We also have that

$$\sum_{\ell<i\in I} (f_i^{(t)}(x) - f_i^{(t)}(y)) \leq \sum_{\ell<i\in I} \Delta_i \leq \Delta_\ell \leq 8^2 \cdot d(x,y).$$

It follows that

$$\begin{aligned}|f^{(t)}(x) - f^{(t)}(y)| &= |\sum_{0<i\in I} (f_i^{(t)}(x) - f_i^{(t)}(y))| \\ &\leq \left(c2^{10} \log^2(2/\epsilon) + 8^2\right) \cdot d(x,y).\end{aligned}$$

□

6.2.3 Scaling Lower Bound Analysis

For any $x,y \in X$ let $\epsilon_{x,y} = \max\left\{\frac{|B^\circ(x,d(x,y))|}{n}, \frac{|B^\circ(y,d(x,y))|}{n}\right\}$. Let

$$K = \left\{k \in [\lceil \log\log n \rceil] \mid k = c^j, j \in \mathbb{N}\right\} \tag{17}$$

For any $k \in K$ let $\epsilon_k = 2^{-8^k}$ and define $\epsilon_1 = 1$. Define $I_k = \{i \in I \mid i = jk, j \in \mathbb{N}\}$. For any $i \in I_k$ let $\bar{N}_k^i$ be a $\frac{\Delta_i}{2^{20}\log\lambda\log^3(2/\epsilon_k)}$-net.

We now wish to sieve the nets: for any $k \in K$ and $i \in I_k$ remove all the points u from the net $\bar{N}_k^i$ if one of these conditions apply:

- $|B(u,\Delta_{i-1})| \geq \epsilon_{k/c} n$, or
- $|B(u,\Delta_{i-k-4})| < \epsilon_k n$,

and call the resulting set N_k^i. The intuition is that the nets we created contain "too many" points, in a sense that the degree of the dependency graph of the Lovasz Local Lemma will be large, so we ignore those net points that play no role in the embedding analysis.

Let $M = \left\{(i,k,u,v) \mid k \in K,\ i \in I_k,\ u,v \in N_k^i, 7\Delta_{i-1} \leq d(u,v) \leq 65\Delta_{i-k-1}\right\}$. Define a function $T : M \to 2^{[D]}$ such that for $t \in [D]$:

$$t \in T(i,k,u,v) \Leftrightarrow \left|f^{(t)}(u) - f^{(t)}(v)\right| \geq \frac{\Delta_i}{4\log(2/\epsilon_k)}.$$

For all $(i,k,u,v) \in M$, let $\mathcal{E}_{(i,k,u,v)}$ be the event that $|T(i,k,u,v)| \geq 15D/16$.

Then we define the event $\mathcal{E} = \bigcap_{(i,k,u,v)\in M} \mathcal{E}_{(i,k,u,v)}$. The main Lemma to prove is:

Lemma 34.

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

we defer the proof for later. In what follows we show that using this lemma we can prove the main theorem.

Let $x, y \in X$, $\epsilon = \epsilon_{x,y}$ (note that $1/n \le \epsilon < 1$). Let $c \le k = k_{x,y} \in K$ be such that $\epsilon_k \le \epsilon < \epsilon_{k/c}$. Let $i' \in I$ be such that $\Delta_{i'-2} \le d(x,y) < \Delta_{i'-3}$, and let $i = i_{x,y} \in I_k$ be the minimal such that $i \ge i'$. Let $u = u(x) \in \bar{N}_k^i$ and $v = v(y) \in \bar{N}_k^i$ such that $d(x,u) = d(x, \bar{N}_k^i)$ and $d(y,v) = d(y, \bar{N}_k^i)$.

The following claim show that indeed we did not remove points from the nets, that were needed for the embedding.

Claim 35. *For any $x, y \in X$, $u(x), v(y) \in N_{k_{x,y}}^{i_{x,y}}$.*

Proof. Let $k = k_{x,y}$, $i = i_{x,y}$, $u = u(x)$, $v = v(y)$. Since $\bar{N}_k^i$ is a $\frac{\Delta_i}{2^{20}\log\lambda\log^3(2/\epsilon_k)}$-net, $|B(u, \Delta_{i-1})| \le |B(x, \Delta_{i-2}/2)| \le \epsilon_{x,y} n < \epsilon_{k/c} n$. On the other hand
$|B(u, \Delta_{i-k-4})| \ge |B(u, \Delta_{i'-4})| \ge \max\{|B(x, d(x,y))|, |B(y, d(x,y))|\} \ge \epsilon_{x,y} n \ge \epsilon_k n$.

The argument for v is similar. $\square$

We will use the following claim

Claim 36. *For any $t \in [D]$ and $(i,k,u,v) \in M$, let $m \in I$ be the minimal such that $\Delta_m \le \frac{\Delta_i}{32\log(2/\epsilon_k)}$. Then for $w \in \{u, v\}$:*

$$\sum_{j \le m} \phi_j^{(t)}(w) \le 2^{13} \log^2(2/\epsilon_k) \log\lambda.$$

Proof. By definition of Δ_m we have that $m \le i + 2\log_8\log(2/\epsilon_k) + \log_8(32) + 1$. From the proof of Claim 35 we have that $|B(u, \Delta_{i-k-4})|, |B(v, \Delta_{i-k-4})| \ge \epsilon_k n$.

By Lemma 7 for any $i \in I$, $\eta_{P,i}(w) \ge 1/(2^7 \log\lambda)$. Using Claim 32 we get

$$\begin{aligned}
\sum_{j \le m} \phi_j^{(t)}(w) &= \sum_{j \le i-k-5} \phi_j^{(t)}(w) + \sum_{j=i-k-4}^{m} \phi_j^{(t)}(w) \\
&\le 2^7 \log^2\left(\frac{n}{|B(w, \Delta_{i-k-4})|}\right) + (m - (i-k-4) + 1) 2^7 \log\lambda \\
&\le 2^7 \log^2(2/\epsilon_k) + (2\log_8\log(2/\epsilon_k) + 8) 2^7 \log\lambda \\
&\le 2^{13} \log^2(2/\epsilon_k) \log\lambda.
\end{aligned}$$

$\square$

We now show the analogue of Claim 28 for the scaling case, in this case a more delicate argument is needed, as there is no sufficiently small universal upper bound on the distortion, but one that depends on ϵ, hence we consider the contribution of different scales: small, medium and large, separately.

Lemma 37. *For any $t \in T(i,k,u,v)$*

$$\left| f^{(t)}(x) - f^{(t)}(y) \right| \ge \frac{\Delta_i}{16\log(2/\epsilon_k)}.$$

Proof. Let $m \in I$ be the minimal such that $\Delta_m \le \frac{\Delta_i}{32\log(2/\epsilon_k)}$.

By Claim 35, we have that $\max\{d(x,u), d(y,v)\} \le \frac{\Delta_i}{2^{20}\log\lambda\log^3(2/\epsilon_k)}$. We define for any $u, x \in X$

$$J_{u,x} = \{j \in I \mid P_j(u) = P_j(x)\}.$$

$$\begin{aligned}
\left|\sum_{j\in I}\left(f_j^{(t)}(u) - f_j^{(t)}(v)\right)\right| &\le \left|\sum_{j\le m}\left(f_j^{(t)}(u) - f_j^{(t)}(v)\right)\right| + \left|\sum_{j>m}\left(f_j^{(t)}(u) - f_j^{(t)}(v)\right)\right| \qquad (18)\\
&\le \left|\sum_{j\in J_{u,x};j\le m} f_j^{(t)}(u) - \sum_{j\in J_{v,y};j\le m} f_j^{(t)}(v)\right| + \left|\sum_{j\notin J_{u,x};j\le m} f_j^{(t)}(u) - \sum_{j\notin J_{v,y};j\le m} f_j^{(t)}(v)\right| \\
&\quad + \sum_{j>m}\Delta_j.
\end{aligned}$$

First we bound the contribution of coordinates in which the points x, y fall in different clusters than u, v respectively, using Claim 36

$$\begin{aligned}
&\left|\sum_{j\notin J_{u,x};j\le m} f_j^{(t)}(u) - \sum_{j\notin J_{v,y};j\le m} f_j^{(t)}(v)\right| \qquad (19)\\
&\le \sum_{j\notin J_{u,x};j\le m} f_j^{(t)}(u) + \sum_{j\notin J_{v,y};j\le m} f_j^{(t)}(v)\\
&\le \sum_{j\notin J_{u,x};j\le m} \phi_j^{(t)}(u)d(u,x) + \sum_{j\notin J_{v,y};j\le m} \phi_j^{(t)}(v)d(v,y)\\
&\le \frac{\Delta_i}{2^{20}\log\lambda\log^3(2/\epsilon_k)}2^{14}\log^2(2/\epsilon_k)\log\lambda\\
&\le \frac{\Delta_i}{2^5\log(2/\epsilon_k)}.
\end{aligned}$$

However we know that

$$\left|\sum_{j\in I} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \ge \frac{\Delta_i}{4\log(2/\epsilon_k)},$$

and since $\sum_{j>m}\Delta_j \le \Delta_m \le \frac{\Delta_i}{32\log(2/\epsilon_k)}$, by plugging this and (19) into (18) we get

$$\begin{aligned}
\left|\sum_{j\in J_{u,x};j\le m} f_j^{(t)}(u) - \sum_{j\in J_{v,y};j\le m} f_j^{(t)}(v)\right| &\ge \frac{\Delta_i}{4\log(2/\epsilon_k)} - \frac{\Delta_i}{2^5\log(2/\epsilon_k)} - \frac{\Delta_i}{32\log(2/\epsilon_k)}\\
&\ge \frac{3\Delta_i}{16\log(2/\epsilon_k)}.
\end{aligned}$$

Assume w.l.o.g that $\sum_{j\in J_{u,x}} f_j^{(t)}(x) - \sum_{j\in J_{v,y}} f_j^{(t)}(y) > 0$, then notice that for any $j \in J_{u,x}$, $t \in D$: $d(u, X \setminus P_j(u)) \le d(u,x) + d(x, X\setminus P_j(u))$, and since the partition is uniform we get that

$$f_j^{(t)}(x) \ge f_j^{(t)}(u) - \phi_j^{(t)}(u)\cdot\frac{\Delta_i}{2^{20}\log\lambda\log^3(2/\epsilon_k)},$$

and similarly

$$f_j^{(t)}(y) \le f_j^{(t)}(v) + \phi_j^{(t)}(v) \cdot \frac{\Delta_i}{2^{20} \log\lambda \log^3(2/\epsilon_k)}.$$

Then by Claim 36

$$\begin{aligned}
&\left| \sum_{j\in J_{u,x};j\le m} f_j^{(t)}(x) - \sum_{j\in J_{v,y};j\le m} f_j^{(t)}(y) \right| \\
&\ge \left| \sum_{j\in J_{u,x};j\le m} \left(f_j^{(t)}(u) - \frac{\phi_j^{(t)}(u)\cdot\Delta_i}{2^{20}\log\lambda\log^3(2/\epsilon_k)} \right) - \sum_{j\in J_{v,y};j\le m} \left(f_j^{(t)}(v) + \frac{\phi_j^{(t)}(v)\cdot\Delta_i}{2^{20}\log\lambda\log^3(2/\epsilon_k)} \right) \right| \\
&\ge \left| \sum_{j\in J_{u,x};j\le m} f_j^{(t)}(u) - \sum_{j\in J_{v,y};j\le m} f_j^{(t)}(v) \right| - \left| \sum_{j\le m} \frac{\phi_j^{(t)}(u)\cdot\Delta_i}{2^{20}\log\lambda\log^3(2/\epsilon_k)} + \frac{\phi_j^{(t)}(v)\cdot\Delta_i}{2^{20}\log\lambda\log^3(2/\epsilon_k)} \right| \\
&\ge \frac{3\Delta_i}{16\log(2/\epsilon_k)} - 2\frac{\Delta_i}{2^6\log(2/\epsilon_k)} \\
&= \frac{5\Delta_i}{32\log(2/\epsilon_k)},
\end{aligned}$$

Using the same argument as in (19) we get that

$$\left| \sum_{j\notin J_{u,x};j\le m} f_j^{(t)}(x) - \sum_{j\notin J_{v,y};j\le m} f_j^{(t)}(y) \right| \le \frac{\Delta_i}{2^5\log(2/\epsilon_k)},$$

as well. and finally

$$\begin{aligned}
\left| \sum_{j\le m} \left(f_j^{(t)}(x) - f_j^{(t)}(y) \right) \right| &\ge \left| \sum_{j\in J_{u,x};j\le m} f_j^{(t)}(x) - \sum_{j\in J_{v,y};j\le m} f_j^{(t)}(y) \right| - \left| \sum_{j\notin J_{u,x};j\le m} f_j^{(t)}(x) - \sum_{j\notin J_{v,y};j\le m} f_j^{(t)}(y) \right| \\
&\ge \frac{5\Delta_i}{32\log(2/\epsilon_k)} - \frac{\Delta_i}{2^5\log(2/\epsilon_k)} \\
&\ge \frac{\Delta_i}{8\log(2/\epsilon_k)}.
\end{aligned}$$

Notice that $\left|\sum_{j>m}\left(f_j^{(t)}(x) - f_j^{(t)}(y)\right)\right| \le \frac{\Delta_i}{32\log(2/\epsilon_k)}$, hence

$$\left| \sum_{j\in I} \left(f_j^{(t)}(x) - f_j^{(t)}(y) \right) \right| \ge \frac{\Delta_i}{16\log(2/\epsilon_k)}.$$

□

As in the previous section, we have

Lemma 38. *If event $\mathcal{E}$ took place then there exists a universal constant $C_2 > 0$ such that for any $\epsilon' > 0$ and any $x, y \in \hat{G}_{\epsilon'}$*

$$\|f(x) - f(y)\|_p \ge C_2 \frac{d(x,y)}{\log^{2c}(2/\epsilon')}.$$

Proof. Any ϵ' such that $d(x,y) > \max\{r_{\epsilon'/2}(x), r_{\epsilon'/2}(y)\}$ satisfies $\epsilon' \le 2\epsilon = 2\epsilon_{x,y}$, hence it is enough to lower bound the contribution by $\Omega\left(\frac{d(x,y)}{\log^{2c}(2/\epsilon)}\right)$. Let $i = i_{x,y}$, $k = k_{x,y}$ and $u = u(x)$, $v = v(y)$. Noticing that $\Delta_{i-k-3} \ge d(x,y)$, $|T(i,k,u,v)| \ge D/16$ and that $\log(2/\epsilon_k) \le \log^c(2/\epsilon)$ for all $\epsilon \le 1/2^8$, we get from Lemma 37 that

$$\begin{aligned}
\|f(x) - f(y)\|_p^p &= D^{-1}\sum_{t\in D} |f^{(t)}(x) - f^{(t)}(y)|^p \\
&\ge D^{-1}\sum_{t\in T(i,k,u,v)} \left(\frac{\Delta_i}{16\log(2/\epsilon_k)}\right)^p \\
&\ge D^{-1}|T(i,k,u,v)|\left(\frac{d(x,y)}{2^{13}\log^2(2/\epsilon_k)}\right)^p \\
&\ge \left(\frac{d(x,y)}{2^{17}\log^{2c}(2/\epsilon)}\right)^p.
\end{aligned}$$

So set $C_2 = 2^{18}$. (If it is the case that $\epsilon \ge 1/2^8$ then $\log(2/\epsilon_k) = 8^c$, so we show $\|f(x) - f(y)\|_p^p \ge C_2' d(x,y)$). ◻

6.2.4 Proof of Lemma 34

Define for every $(i,k,u,v) \in M$, $i + k/2 \le \ell < i + k$ and $t \in [D]$ the event $\mathcal{F}_{(i,k,u,v,t,\ell)}$ as

$$\left(\left|f_\ell^{(t)}(u) - f_\ell^{(t)}(v)\right| > \Delta_\ell \wedge \left|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \le \frac{\Delta_\ell}{2}\right) \bigvee$$
$$\left(\left(f_\ell^{(t)}(u) = f_\ell^{(t)}(v) = 0\right) \wedge \left|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \ge \frac{\Delta_\ell}{2}\right).$$

Now define event $\hat{\mathcal{E}}_{(i,k,u,v)}$ as

$$\exists S \subseteq [D], |S| \ge 15D/16, \forall t \in S, \exists \ell \text{ s.t. } i + k/2 \le \ell < i + k \text{ and } \mathcal{F}_{(i,k,u,v,t,l)} \text{ holds.}$$

Claim 39. *For all $(i,k,u,v) \in M$, $\hat{\mathcal{E}}_{(i,k,u,v)}$ implies $\mathcal{E}_{(i,k,u,v)}$*

Proof. Let $S \subseteq [D]$ be the subset of good coordinates from the definition of $\hat{\mathcal{E}}_{(i,k,u,v)}$. For any $t \in S$, let $i + k/2 \le \ell(t) < i + k$ be such that $\mathcal{F}_{(i,k,u,v,t,\ell(t))}$ holds. Then for such $t \in S$:

$$\left|\sum_{j\le\ell(t)} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \ge \frac{\Delta_{\ell(t)}}{2}.$$

We also have that

$$\left|\sum_{j>\ell(t)} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \le \sum_{j>\ell(t)} \Delta_j \le \frac{\Delta_{\ell(t)}}{4}.$$

Which implies that

$$\left|\sum_{j\in I} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \ge \frac{\Delta_{\ell(t)}}{4} \ge \frac{\Delta_i 8^{-(k-1)}}{4} \ge \frac{\Delta_i}{4\log(2/\epsilon_k)},$$

as required.

$\square$

Now we shall use a variation the general case of the local Lemma, the reason being that in the graph we shall soon define the degree of the vertices will depend on k, and cannot be uniformly bounded.

Lemma 40 (Lovasz Local Lemma - General Case)**.** *Let $\mathcal{A}_1, \mathcal{A}_2, \dots \mathcal{A}_n$ be events in some probability space. Let $G(V, E)$ be a directed graph on n vertices, each vertex corresponds 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 $c(\mathcal{A}_i) \leq c(\mathcal{A}_j)$. Assume that for all $i = 1, \dots, n$ there exists $x_i \in [0, 1)$ such that*

$$\Pr\left[\mathcal{A}_i \mid \bigwedge_{j \in Q} \neg\mathcal{A}_j\right] \leq x_i \prod_{j:(i,j)\in E} (1 - x_j),$$

for all $Q \subseteq \{j : (\mathcal{A}_i, \mathcal{A}_j) \notin E \wedge c(\mathcal{A}_i) \geq c(\mathcal{A}_j)\}$, then

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

Proof. We iteratively apply the Lovasz 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 $\mathcal{A}_i \in V_1$,

$$\Pr\left[\mathcal{A}_i \mid \bigwedge_{j \in Q} \neg\mathcal{A}_j\right] \leq x_i \prod_{j:(i,j)\in E_1} (1 - x_j),$$

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

$$\Pr\left[\mathcal{A}_i \mid \bigwedge_{j \in Q} \neg\mathcal{A}_j\right] \leq x_i \prod_{j:(i,j)\in E_k} (1 - x_j),$$

for any Q satisfying $Q \subseteq \{j : (\mathcal{A}_i, \mathcal{A}_j) \notin E_k \wedge c(\mathcal{A}_j) = k\} \cup \{j : \mathcal{A}_j \in V_1 \cup \dots \cup V_{k-1}\}$. 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 $G = (V, E)$, where $V = \left\{\hat{\mathcal{E}}_{(i,k,u,v)} \mid (i, k, u, v) \in M\right\}$. Define $c : V \to I$ by $c(\hat{\mathcal{E}}_{(i,k,u,v)}) = i + k$.

We say that a pair of vertices $\left(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}\right) \in E$ if all of these conditions apply:

- $d(\{u,v\},\{u',v'\}) \le 4\Delta_i$.
- $i = i'$.
- $k = k'$.

Claim 41. *The out-degree of $\hat{\mathcal{E}}_{(i,k,u,v)} \in G$ is bounded by $\lambda^{30k \log\log(2\lambda)}$.*

Proof. Fix some $\hat{\mathcal{E}}_{(i,k,u,v)} \in V$, we will see how many pairs $u', v' \in N_k^i$ can exists such that $(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i,k,u',v')}) \in E$.

For any such u', v' assume w.l.o.g that $d(u,u') \le 4\Delta_i$, hence as $d(u,v), d(u',v') \le 65\Delta_{i-k-1}$ we get $u,v,u',v' \in B = B(u, \Delta_{i-k-4})$. The number of pairs can be bounded by $|N_k^i \cap B|^2$. Since (X,d) is λ-doubling the ball B can be covered by $\lambda^{33+12k+\log\log\lambda}$ balls of radius $\frac{\Delta_i}{8^{7+3k}\log\lambda}$, each of these contains at most one point of the set N_k^i. As $k \ge c = 12$, $|N_k^i \cap B|^2 \le \lambda^{30k\log\log(2\lambda)}$. □

Lemma 42.

$$\Pr\left[\neg\mathcal{E}_{(i,k,u,v)} \mid \bigwedge_{(i',k',u',v')\in Q} \mathcal{E}_{(i',k',u',v')}\right] \le \lambda^{-32k\log\log(2\lambda)},$$

for all $Q \subseteq \{(i',k',u',v') \mid i+k \ge i'+k' \wedge \left(\mathcal{E}_{(i,k,u,v)}, \mathcal{E}_{(i',k',u',v')}\right) \notin E\}$

Before we prove this lemma, let us see that it implies Lemma 34. Apply Lemma 40 to the graph G we defined. For any $(i,k,u,v) \in M$ assign the number $x_k = \lambda^{-30k\log\log(2\lambda)}$ for the vertex $\hat{\mathcal{E}}_{(i,k,u,v)}$. From the definition of G it can be seen that if $\left(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}\right) \in E$ then $x_{k'} = x_k$.

By Claim 41 there at most $\lambda^{30k\log\log(2\lambda)}$ neighbors to the vertex $\hat{\mathcal{E}}_{(i,k,u,v)}$, so for any such vertex:

$$x_k \prod_{(i',k',u',v'):\left(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}\right)\in E} (1 - x_{k'}) \ge x_k \left(1 - x_k\right)^{\lambda^{30k\log\log(2\lambda)}} \ge 1/4 \cdot x_k \ge \lambda^{-32k\log\log(2\lambda)}.$$

By Lemma 42 we get that indeed

$$\Pr\left[\neg\hat{\mathcal{E}}_{(i,k,u,v)} \mid \bigwedge_{(i',k',u',v'):\left(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}\right)\notin E} \mathcal{E}_{(i',k',u',v')}\right] \le \lambda^{-32k\log\log(2\lambda)},$$

as required by Lemma 40, hence

$$\Pr[\bigwedge_{(i,k,u,v)\in M} \hat{\mathcal{E}}_{(i,k,u,v)}] > 0.$$

By Claim 39 we have

$$\Pr[\mathcal{E}] = \Pr[\bigwedge_{(i,k,u,v)\in M} \mathcal{E}_{(i,k,u,v)}] > 0.$$

6.2.5 Proof of Lemma 42

Claim 43. *Let* $(i,k,u,v) \in M$*,* $t \in [D]$ *and* $i + k/2 \leq \ell < i + k$*, then*

$$\Pr\left[\mathcal{F}_{(i,k,u,v,t,\ell)}\right] \geq 1/8.$$

Proof. We begin by showing that $\xi_{P,\ell}(u) = 1$ which will imply that $\phi_\ell^{(t)}(u) = \eta_{P,\ell}(u)^{-1}$. In order to show that we will prove that $\max\{\bar{\rho}(u, 2\Delta_\ell, \gamma_1(\cdot,\ell), \gamma_2), \bar{\rho}(v, 2\Delta_\ell, \gamma_1(\cdot,\ell), \gamma_2)\} \geq 2$, and then assume w.l.o.g that $\bar{\rho}(u, 2\Delta_\ell, \gamma_1(\cdot,\ell), \gamma_2) \geq 2$. It follows from Lemma 7 that $\xi_{P,\ell}(u) = 1$. Now to prove that $\max\{\bar{\rho}(u, 2\Delta_\ell, \gamma_1(\cdot,\ell), \gamma_2), \bar{\rho}(v, 2\Delta_\ell, \gamma_1(\cdot,\ell), \gamma_2)\} \geq 2$:

Consider any $a \in B(u, 2\Delta_\ell)$ (a is a potential center to the cluster containing u in scale ℓ). As $k > 2$ we have that $\ell - 1 > i$, then since $|B(a, 2\Delta_\ell)| \leq |B(u, \Delta_{i-1})| < \epsilon_{k/c} n$ we have that $\epsilon_\ell(a) \leq \epsilon_{k/c}$ which implies that $\gamma_1(a,\ell) \geq 8^4 \log^{2c}(64/\epsilon_{k/c}) \geq 8^{4+2c(k/c)} = 8^{4+2k}$. Since $\Delta_\ell \geq 8\Delta_i/8^k$ we get that $\gamma_1(a,\ell) 2\Delta_\ell \geq 8^{4+2k} \cdot \frac{16\Delta_i}{8^k} = 8^{4+2k} \cdot \frac{16\Delta_{i-k-1}}{8^{k+k+1}} \geq 2 \cdot 65\Delta_{i-k-1} \geq 2d(u,v)$, where the last inequality is by the definition of M.

The same argument shows that for any $a \in B(v, 2\Delta_\ell)$, $\gamma_1(a,\ell) 2\Delta_\ell \geq 2d(u,v)$ as well. Therefore by Claim 3 we have $\max\{\bar{\rho}(u, 2\Delta_\ell, \gamma_1(\cdot,\ell), \gamma_2), \bar{\rho}(v, 2\Delta_\ell, \gamma_1(\cdot,\ell), \gamma_2)\} \geq 2$ as required.

We now consider the 2 cases in $\mathcal{F}_{(i,u,v,t,\ell)}$: If it is the case that

$$\left|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \leq \frac{\Delta_\ell}{2}$$

then we wish that the following will hold

- $B(u, \eta_{P,\ell}^{(t)}(u)\Delta_\ell) \subseteq P_\ell(u)$.
- $\sigma_\ell^{(t)}(P_\ell(u)) = 1$.
- $\sigma_\ell^{(t)}(P_\ell(v)) = 0$.

Each of these happens independently with probability at least 1/2, the first since P_ℓ is $(\eta_\ell, 1/2)$-padded and the other two follow from $d(u,v) \geq 3\Delta_\ell \Rightarrow P_\ell(u) \neq P_\ell(v)$.

Similarly if it is the case that

$$\left|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\right| > \frac{\Delta_\ell}{2}$$

then we wish that the following will hold

- $\sigma_\ell^{(t)}(P_\ell(u)) = \sigma_\ell^{(t)}(P_\ell(v)) = 0$.

And again there is probability 1/2 for each of these.

So we have probability at least 1/8 for event $\mathcal{F}_{(i,u,v,t,\ell)}$.

□

The main independence claim is the following:

Claim 44. *Let $(i,k,u,v) \in M$, $t \in [D]$ and $i + k/2 \le \ell < i + k$. Then*

$$\Pr\left[\neg\mathcal{F}_{(i,k,u,v,t,\ell)} \mid \bigwedge_{(i',k',u',v')\in Q} \mathcal{E}_{(i',k',u',v')}\right] \le 7/8,$$

for all $Q \subseteq \{(i',k',u',v') \mid i+k \ge i'+k' \wedge \left(\mathcal{E}_{(i,k,u,v)}, \mathcal{E}_{(i',k',u',v')}\right) \notin E\}$

Proof. Fix some $\mathcal{E}_{(i',k',u',v')}$ such that $\left(\mathcal{E}_{(i,k,u,v)}, \mathcal{E}_{(i',k',u',v')}\right) \notin E$ and $i + k \ge i' + k'$.

First consider the case that $d(\{u,v\},\{u',v'\}) > 4\Delta_i$. Then since the partition is local, for any $\ell \in [i+k/2, i+k)$ the probability of the padding event and choice of σ for scale ℓ are not affected by of the outcome of events such as $\mathcal{E}_{(i',k',u',v')}$.

From now on assume that $d(\{u,v\},\{u',v'\}) \le 4\Delta_i$, and w.l.o.g $d(u,u') \le 4\Delta_i$. The idea is to show that $i'+k' \le i+k/2$, and hence as event $\mathcal{E}_{(i',k',u',v')}$ is concerned with scales at most $i'+k'-1$ the padding and choice of σ for scales $i+k/2,\dots,i+k-1$ will be independent of the outcome of events such as $\mathcal{E}_{(i',k',u',v')}$.

Case 1: $k' < k$. By the definition of K follows that $k' \le k/c$. If it is the case that $i' \le i$ then $i'+k' \le i+k/c < i+k/2$. If $i' > i$, then assume by contradiction that $i'+k' \ge i+k/2$. By the nets sieving process we have $\epsilon_{k'}n < |B(u', \Delta_{i'-k'-4})|$ and also $\epsilon_{k'}n \ge \epsilon_{k/c}n \ge |B(u, \Delta_{i-1})|$. Now $i' - k' - 4 \ge i + k/2 - k' - k' - 4 \ge i + k(1/2 - 2/c) - 4 \ge i$, as $c = 12$ and $k \ge c$. Since $d(u,u') \le 4\Delta_i$ follows that $|B(u', \Delta_{i'-k'-4})| \le |B(u', \Delta_i)| \le |B(u, \Delta_{i-1})| \le \epsilon_{k'}n$. Contradiction.

Case 2: $k' > k$. Then it must be that $i' < i$. We will show that this cannot be. Note that since $i + k \ge i' + k'$ and $k \le k'/c$ then $i \ge i' + k' - k \ge i' + k'(1 - 1/c)$. Now similarly to the previous case we have $\epsilon_k n < |B(u, \Delta_{i-k-4})| \le |B(u, \Delta_{i'+k'(1-1/c)-k'/c-4})| \le |B(u, \Delta_{i'})| \le |B(u', \Delta_{i'-1})| \le \epsilon_{k'/c}n \le \epsilon_k n$. Contradiction.

Case 3: If $k = k'$ then by the construction of G $i \ne i'$, therefore $i' < i$. By the definition of I_k, $i' + k' \le i < i + k/2$.

We conclude that if indeed $\left(\hat{\mathcal{E}}_{(i,k,u,v)}, \hat{\mathcal{E}}_{(i',k',u',v')}\right) \notin E$ then Claim 43 suggests that there is probability at least $1/8$ for event $\mathcal{F}_{(i,k,u,v,t,\ell)}$ to hold, independently of $\hat{\mathcal{E}}_{(i',k',u',v')}$. $\square$

Now we are ready to prove Lemma 42. First consider the case where $k < 60$, then fix some $\ell \in [i+k/2, i+k)$, and let $\hat{Z}_t$ be the indicator event for $\mathcal{F}_{(i,k,u,v,t,\ell)}$, $\Pr[\hat{Z}_t] \ge 1/8$ and let $\hat{Z} = \sum_{t=1}^{D} \hat{Z}_t$. As each coordinate is independent of the others, and $\mathbb{E}[\hat{Z}] \ge D/8$, using Chernoff's bound:

$$\Pr[\hat{Z} < D/16] \le \Pr[\hat{Z} < \mathbb{E}[\hat{Z}]/2] \le e^{-D/64} \le \lambda^{-32\cdot 60 \log\log(2\lambda)} \le \lambda^{-32k \log\log(2\lambda)},$$

for large enough constant C.

On the other hand if $k \ge 60$, then for every coordinate $t \in [D]$, we have $k/2$ possible values of ℓ. In each scale ℓ, by Claim 44 there is probability at most $(7/8)$ to fail, this probability is unaffected by of all other scales $\ell' < \ell$. Let $\mathcal{Y}_\ell$ be the indicator event for $\neg\mathcal{F}_{(i,k,u,v,t,\ell)}$. The probability that we failed for all scales $\ell \in [i+k/2, i+k)$ can be bounded by:

$$\Pr\left[\bigwedge_{\ell=i+k/2}^{i+k-1} \mathcal{Y}_\ell\right] = \prod_{\ell=i+k/2}^{i+k-1} \left(\Pr\left[\mathcal{Y}_\ell \mid \bigwedge_{j=i+k/2}^{\ell-1} \mathcal{Y}_j\right]\right) \le (7/8)^{k/2} = z.$$

Let Z_t be the event that we failed in the t-th coordinate, $\Pr[Z_t] \le z$, and $Z = \sum_{t \in D} Z_t$. We have that $\mathbb{E}[Z] \le zD$, let $\alpha \ge 1$ such that $\mathbb{E}[Z] = \frac{zD}{\alpha}$. Using Chernoff bound:

$$\begin{aligned}
\Pr[Z > D/2] &\le \left(\frac{e^{\alpha/(2z)-1}}{(\alpha/(2z))^{\alpha/(2z)}}\right)^{zD/\alpha} \\
&\le (2ez)^{D/2} \le \lambda^{(\log(2e)+(k/2)\log(7/8))(C/2)\log\log(2\lambda)} \\
&\le \lambda^{(k/4)\log(7/8)(C/2)\log\log(2\lambda)} \\
&\le \lambda^{-32k\log\log(2\lambda)},
\end{aligned}$$

since for $k \ge 60$ we have $\log(2e) < -(k/4)\log(7/8)$, and for large enough constant C. This concludes the proof of Lemma 42 and hence the proof of Theorem 16.

6.3 Snowflake Results

In this section we prove a stronger version of the original theorem of Assouad [Ass83], in which for a given metric space (X, d) one embeds a "snowflake" version (X, d^α) of the metric for some constant $0 < \alpha < 1$.

Theorem 17. *For any n point λ-doubling metric space (X, d), any $0 < \alpha < 1$, any $p \ge 1$, any $\theta \le 1$ and any $2^{192/\theta} \le k \le \log\lambda$, there exists an embedding of (X, d^α) into l_p with distortion $O(k^{1+\theta}\lambda^{1/(pk)}/(1-\alpha))$ and dimension $O\left(\frac{\lambda^{1/k}\ln\lambda}{\alpha\theta}\cdot\left(1-\frac{\log(1-\alpha)}{\log k}\right)\right)$.*

One can put $\alpha = 1/2$ which translate the distortion to $O(k^{1+\theta}\lambda^{1/(pk)})$ and the dimension to $O\left(\frac{\lambda^{1/k}\ln\lambda}{\theta}\right)$. Now taking $k = \log\lambda$ yields for any $p \ge 1$ distortion $O(\log^{1+\theta}\lambda)$ and dimension $O((\log\lambda)/\theta)$, which is very similar to the results of Theorem 8 where in the distortion instead of being a function of n is replaced by being a function of λ. The special case when $k = \log\lambda$ and $\theta = 1/(192\log\log\lambda)$ was shown by [GKL03]. Another interesting choice of parameters is when we embed into l_p for $p \ge \log\lambda$, then one can choose $\theta = 1$, a constant k, and obtain an embedding with constant distortion.

6.3.1 Proof overview

The high level approach is similar to that of Theorem 8. However here it is sufficient to use Lemma 8 instead of Lemma 7. In each term for scale i of the embedding (i.e. $f_i(x)$) we follow Assouad's technique [Ass83] and introduce a factor of $\Delta_i^{\alpha-1}$. Hence the upper bound of Lemma 46 is independent of the number of scales or the number of points in the metric. We exploit the higher norm l_p in the lower bound, Lemma 50. The main technical lemma is Lemma 53 which requires a subtle use of Chernoff bounds.

6.3.2 The proof

Let $\Delta_0 = \mathrm{diam}(X)$ and $I = \{i \in \mathbb{Z} \mid 1 \le i \le (\log\Delta_0 + \theta\log\log\lambda)/3\}$. For $i \in I$ let $\Delta_i = \Delta_0/8^{i/\alpha}$. Set $D = \frac{c\lambda^{1/k}\ln\lambda}{\alpha\theta}\left(1 - \frac{\log(1-\alpha)}{\log k}\right)$ for some constant c to be determined later.

Let $\delta = \lambda^{-1/k}$, $\tau = 2^{-7}\ln(1/\delta)/\ln(\lambda) = 2^{-7}/k$. We shall define the embedding f by defining for each $t \in [D]$ a function $f^{(t)} : X \to \mathbb{R}^+$ and let $f = D^{-1/p}\bigoplus_{t\in[D]} f^{(t)}$.

Fix $t \in [D]$. In what follows we define $f^{(t)}$. For each $0 < i \in I$ construct a Δ_i-bounded (τ, δ)-padded probabilistic partition $\hat{\mathcal{P}}_i$, as in Lemma 8. Fix some $P_i \in \mathcal{P}_i$ for all $i \in I$.

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 $f_i^{(t)}(x) = \sigma_i^{(t)}(P_i(x)) \cdot \Delta_i^{\alpha-1} \min\{\tau^{-1} \cdot d(x, X \setminus P_i(x)), \Delta_i\}$.

Claim 45. *For any $0 < i \in I$ and $x, y \in X$: $f_i^{(t)}(x) - f_i^{(t)}(y) \leq \Delta_i^{\alpha-1} \cdot \min\{\tau^{-1} \cdot d(x,y), \Delta_i\}$.*

The proof of this claim is essentially the same as of Claim 14.

Lemma 46. *For any $x, y \in X$ and $t \in [D]$:*

$$|f^{(t)}(x) - f^{(t)}(y)| \leq 2^9 k \frac{d(x,y)^\alpha}{1-\alpha}.$$

Proof. We will divide the sum to two parts. Let $\ell \in I$ be the minimal such that $\Delta_\ell \leq d(x,y)$ (so that $\Delta_\ell > d(x,y)/8^{1/\alpha}$). By Claim 45

$$\sum_{0<i<\ell} |f_i^{(t)}(x) - f_i^{(t)}(y)| \leq \tau^{-1} \cdot d(x,y) \sum_{0<i<\ell} \Delta_i^{\alpha-1} \leq 2^8 k \cdot d(x,y)^\alpha/(1-\alpha) \ ,$$

where we used that $\tau^{-1} = 2^7 k$ and

$$\sum_{0<i<\ell} \Delta_i^{\alpha-1} = \Delta_0^{\alpha-1} \sum_{0\leq i<\ell} (8^{(1-\alpha)/\alpha})^i \leq \frac{\Delta_\ell^{\alpha-1}}{8^{(1-\alpha)/\alpha} - 1} \leq \frac{(8^{-1/\alpha} d(x,y))^{\alpha-1}}{8^{(1-\alpha)/\alpha} - 1} \leq \frac{2d(x,y)^{\alpha-1}}{1-\alpha} \ .$$

Also

$$\sum_{i \geq \ell} |f_i^{(t)}(x) - f_i^{(t)}(y)| \leq \sum_{i\geq \ell} \Delta_i^\alpha \leq \Delta_\ell^\alpha \sum_{i \geq 0} 8^{-i} \leq 2d(x,y)^\alpha \ .$$

To conclude, for any $t \in [D]$,

$$|f^{(t)}(x) - f^{(t)}(y)| = \sum_{i \in I} |f_i^{(t)}(x) - f_i^{(t)}(y)| \leq 2^9 k \cdot d(x,y)^\alpha/(1-\alpha) \ , \tag{20}$$

$\square$

Lemma 47. *For any $p \geq 1$ and $x, y \in X$,*

$$\|f(x) - f(y)\|_p \leq 2^9 k d(x,y)^\alpha/(1-\alpha) \ .$$

Proof. By Lemma 46

$$\begin{aligned} \|f(x) - f(y)\|_p^p &= \frac{1}{D} \sum_{t \in [D]} |f^{(t)}(x) - f^{(t)}(y)|^p \\ &\leq \frac{1}{D} \sum_{t\in[D]} \left(2^9 k \cdot d(x,y)^\alpha/(1-\alpha)\right)^p \\ &= \left(2^9 k \cdot d(x,y)^\alpha/(1-\alpha)\right)^p \ . \end{aligned}$$

$\square$

6.3.3 Lower Bound Analysis

The lower bound analysis uses a set of nets. First we define a set of scales in which we hope to succeed with high probability. Let $r = \lceil (\theta/3)\log k \rceil$, let $R = \{i \in I : r|i\}$. For any $0 < i \in R$ let N_i be a $\Delta_i \left(\frac{1-\alpha}{2^{13}k^{1+\theta}}\right)^{1/\alpha}$-net of X. The purpose of considering only one of every $\approx \theta \log k$ scales is to avoid dependencies when using the local lemma.

Let $M = \{(i,u,v) \mid i \in R,\ u,v \in N_i, 7\Delta_{i-1} \le d(u,v) \le 9\Delta_{i-r-2}\}$. Given an embedding f define a function $T : M \to 2^{[D]}$ such that for $t \in [D]$:

$$t \in T(i,u,v) \Leftrightarrow \left| f^{(t)}(u) - f^{(t)}(v) \right| \ge \frac{\Delta_i^\alpha}{4k^\theta} \ .$$

For all $(i,u,v) \in M$, let $\mathcal{E}_{(i,u,v)}$ be the event $|T(i,u,v)| \ge \lambda^{-1/k} D/4$. Then we define the event $\mathcal{E} = \bigcap_{(i,u,v)\in M} \mathcal{E}_{(i,u,v)}$ that captures the case that all triplets in M have the desired property.

The main technical lemma is the following:

Lemma 48.

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

We defer the proof for later, and now show that if the event $\mathcal{E}$ took place, then we can show the lower bound. Let $x, y \in X$, and let $0 < i' \in I$ be such that $8\Delta_{i'-1} \le d(x,y) \le 8\Delta_{i'-2}$. Let $i \in R$ be the minimal such that $i \ge i'$, note that $\Delta_i^\alpha \ge \frac{\Delta_{i'}^\alpha}{k^\theta}$. Consider $u, v \in N_i$ satisfying $d(x,u) = d(x,N_i)$ and $d(y,v) = d(y,N_i)$, as $d(u,v) \le d(u,x) + d(x,y) + d(y,v) \le 8\Delta_{i'-2} + \Delta_i \le 8\Delta_{i-r-2} + \Delta_i \le 9\Delta_{i-r-2}$ and $d(u,v) \ge d(x,y) - d(u,x) - d(v,y) \ge 8\Delta_{i'-1} - \Delta_i \ge 7\Delta_{i-1}$, so by the definition of M follows that $(i,u,v) \in M$. The next lemma shows that since x, y are very close to u, v respectively, then by the triangle inequality the embedding f of x, y cannot differ by much from that of u, v (respectively).

Lemma 49. *Let $x, y \in X$, let i' such that $8\Delta_{i'-1} \le d(x,y) \le 8\Delta_{i'-2}$, let $i \in R$ be the minimal such that $i \ge i'$ and let $u, v \in N_i$ satisfying $d(x,u) = d(x,N_i)$ and $d(y,v) = d(y,N_i)$.*

Given $\mathcal{E}$, for any $t \in T(i,u,v)$:

$$\left| f^{(t)}(x) - f^{(t)}(y) \right| \ge \frac{\Delta_i^\alpha}{8k^\theta}.$$

Proof. Since N_i is $\Delta_i \left(\frac{1-\alpha}{2^{13}k^{1+\theta}}\right)^{1/\alpha}$-net, then $d(x,u)^\alpha \le \Delta_i^\alpha \frac{1-\alpha}{2^{13}k^{1+\theta}}$. By Lemma 46 $|f^{(t)}(x) - f^{(t)}(u)| \le 2^9 k \cdot d(x,u)^\alpha/(1-\alpha) \le \frac{\Delta_i^\alpha}{16k^\theta}$, and similarly $|f^{(t)}(y) - f^{(t)}(v)| \le \frac{\Delta_i^\alpha}{16k^\theta}$. By the triangle inequality we get that

$$\begin{aligned}
|f^{(t)}(x) - f^{(t)}(y)| &= |f^{(t)}(x) - f^{(t)}(u) + f^{(t)}(u) - f^{(t)}(v) + f^{(t)}(v) - f^{(t)}(y)| \\
&\ge |f^{(t)}(u) - f^{(t)}(v)| - |f^{(t)}(x) - f^{(t)}(u)| - |f^{(t)}(y) - f^{(t)}(v)| \\
&\ge \frac{\Delta_i^\alpha}{4k^\theta} - \frac{2\Delta_i^\alpha}{16k^\theta} \\
&= \frac{\Delta_i^\alpha}{8k^\theta}.
\end{aligned}$$

□

This Lemma and Lemma 48 implies the following:

Lemma 50. *There exists a universal constant $C_2 > 0$ and an embedding f such that for any $x, y \in X$*

$$\|f(x) - f(y)\|_p \geq C_2 \frac{d(x,y)^\alpha}{k^{2\theta}\lambda^{1/(pk)}}.$$

Proof. Let f be an embedding such that event $\mathcal{E}$ took place. Let $i' \in I$ such that $8\Delta_{i'-1} \leq d(x,y) < 8\Delta_{i'-2}$, $i \in R$ the minimal such that $i \geq i'$ and u, v be the nearest points to x, y respectively in the net N_i. Noticing that $\Delta_i^\alpha \geq \frac{d(x,y)^\alpha}{2^9 k^\theta}$ and that $|T(i,u,v)| \geq \lambda^{-1/k} D/4$ we get from Lemma 49 that

$$\begin{aligned}
\|f(x) - f(y)\|_p^p &= D^{-1} \sum_{t \in [D]} |f^{(t)}(x) - f^{(t)}(y)|^p \\
&\geq D^{-1} \sum_{t \in T(i,u,v)} \left(\frac{\Delta_i^\alpha}{8k^\theta}\right)^p \\
&\geq D^{-1} |T(i,u,v)| \left(\frac{d(x,y)^\alpha}{2^{12}k^{2\theta}}\right)^p \\
&\geq \lambda^{-1/k} \left(\frac{d(x,y)^\alpha}{2^{14}k^{2\theta}}\right)^p.
\end{aligned}$$

□

6.3.4 Proof of Lemma 48

Define for every $(i,u,v) \in M$, $i \leq \ell < i + r$ and $t \in [D]$ the event $\mathcal{F}_{(i,u,v,t,\ell)}$ as

$$\left(\left|f_\ell^{(t)}(u) - f_\ell^{(t)}(v)\right| \geq \Delta_\ell^\alpha \wedge \Big|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\Big| \leq \frac{\Delta_\ell^\alpha}{2}\right) \bigvee$$
$$\left(\left(f_\ell^{(t)}(u) = f_\ell^{(t)}(v) = 0\right) \wedge \Big|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\Big| > \frac{\Delta_\ell^\alpha}{2}\right).$$

Also define event $\hat{\mathcal{E}}_{(i,u,v)}$ as

$$\exists S \subseteq [D], |S| \geq \lambda^{-1/k} D/4, \forall t \in S, \exists i \leq \ell < i + r \ : \ \mathcal{F}_{(i,u,v,t,\ell)} \text{ holds.}$$

Claim 51. *For all $(i,u,v) \in M$, $\hat{\mathcal{E}}_{(i,u,v)}$ implies $\mathcal{E}_{(i,u,v)}$.*

Proof. Let $S \subseteq [D]$ be the subset of coordinates from the definition of $\hat{\mathcal{E}}_{(i,u,v)}$. For any $t \in S$, let $i \leq \ell(t) < i + r$ be such that $\mathcal{F}_{(i,u,v,t,\ell(t))}$ holds. Then for such $t \in S$:

$$\Big|\sum_{j \leq \ell(t)} f_j^{(t)}(u) - f_j^{(t)}(v)\Big| \geq \frac{\Delta_{\ell(t)}^\alpha}{2}.$$

From Claim 45 it follows that

$$\Big|\sum_{j > \ell(t)} f_j^{(t)}(u) - f_j^{(t)}(v)\Big| \leq \sum_{j > \ell(t)} \Delta_j^\alpha = \Delta_{\ell(t)}^\alpha \sum_{j>0} 8^{-j} = \Delta_{\ell(t)}^\alpha / 7 \ .$$

Which implies that

$$\left|\sum_{j\in I} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \geq \frac{\Delta_{\ell(t)}^{\alpha}}{4} \geq \frac{\Delta_i^{\alpha}}{4k^{\theta}},$$

as required. $\square$

Define a graph $G=(V,E)$, where $V=\{\hat{\mathcal{E}}_{(i,u,v)} \mid (i,u,v)\in M\}$, and the rating of a vertex $c(\hat{\mathcal{E}}_{(i,u,v)})=i$. We say that a pair of vertices $(\hat{\mathcal{E}}_{(i,u,v)},\hat{\mathcal{E}}_{(i',u',v')})\in E$ if

- $d(\{u,v\},\{u',v'\})\leq 4\Delta_i$, and
- $i=i'$.

Claim 52. *The out-degree of G is bounded by $\lambda^{(52+6\log k-2\log(1-\alpha))/\alpha}$.*

Proof. Fix some $\hat{\mathcal{E}}_{(i,u,v)}\in V$, we will see how many pairs $u',v'\in N_i$ can exists such that $(\hat{\mathcal{E}}_{(i,u,v)},\hat{\mathcal{E}}_{(i,u',v')})\in E$.

Assume w.l.o.g $d(u,u')\leq 4\Delta_i$, since $d(u,v),d(u',v')\leq 9\Delta_{i-r-2}$ follows $u,v,u',v'\in B=B(u,\Delta_{i-r-4})$. The number of pairs can be bounded by $|N_i\cap B|^2$. Since (X,d) is λ-doubling, the ball B of radius $r_1=(2^{12}k^{\theta})^{1/\alpha}\Delta_i$ can be covered by $b=\lambda^{\lceil\log(r_1/r_2)\rceil}$ balls of radius $r_2=\Delta_i\left(\frac{1-\alpha}{2^{14}k^{1+\theta}}\right)^{1/\alpha}$, where $b\leq\lambda^{(26+3\log k-\log(1-\alpha))/\alpha}$, and each of these balls contains at most one point in the net N_i. It follows that $|N_i\cap B|^2\leq b^2\leq\lambda^{(52+6\log k-2\log(1-\alpha))/\alpha}$. $\square$

Notice that events $\hat{\mathcal{E}}_{(i,u,v)}$ do not depend on the choice of partitions for scales greater than $i+r$.

Lemma 53.

$$\Pr\left[\neg\hat{\mathcal{E}}_{(i,u,v)} \mid \bigwedge_{(i',u',v')\in Q}\hat{\mathcal{E}}_{(i',u',v')}\right]\leq\lambda^{-(53+6\log k-2\log(1-\alpha))/\alpha},$$

for all $Q\subseteq\left\{(i',u',v')\mid i\geq i'\wedge\left(\hat{\mathcal{E}}_{(i,u,v)},\hat{\mathcal{E}}_{(i',u',v')}\right)\notin E\right\}$.

Before we prove this lemma, let us see that it implies Lemma 48.

Apply Lemma 29 to the graph G we defined, by Claim 52 let $d=\lambda^{(52+6\log k-2\log(1-\alpha))/\alpha}$ and by Lemma 53 we can let $p=\lambda^{-(53+6\log k-2\log(1-\alpha))/\alpha}$ satisfying the first condition of Lemma 29. It is easy to see that the second condition also holds (since $\lambda\geq 2$), hence

$$\Pr[\bigwedge_{(i,u,v)\in M}\hat{\mathcal{E}}_{(i,u,v)}]>0.$$

By Claim 51 we have

$$\Pr[\mathcal{E}]=\Pr[\bigwedge_{(i,u,v)\in M}\mathcal{E}_{(i,u,v)}]>0,$$

which concludes the proof of Lemma 48.

6.3.5 Proof of Lemma 53

In order to prove this lemma, we first show the following claim, a slight variation of a claim shown in [ABN06].

Claim 54. *Let $(i,u,v) \in M$, $t \in [D]$ and $i \le \ell < i+r$ then* $\Pr\left[\mathcal{F}_{(i,u,v,t,\ell)}\right] \ge \lambda^{-1/k}/4$.

Proof. Let $i \le \ell < i+r$ and consider the two cases in $\mathcal{F}_{(i,u,v,t,\ell)}$:

If it is the case that $\left|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\right| \le \frac{\Delta_\ell^\alpha}{2}$ then it is enough for the following to hold

- $B(u, \tau\Delta_\ell) \subseteq P_\ell(u)$.
- $\sigma_\ell^{(t)}(P_\ell(u)) = 1$.
- $\sigma_\ell^{(t)}(P_\ell(v)) = 0$.

The second and third events happen independently with probability at least 1/2, the first happens with probability at least $\delta = \lambda^{-1/k}$, since P_ℓ is (τ,δ)-padded. If all these events occur then $|f_\ell^{(t)}(u) - f_\ell^{(t)}(v)| \ge \Delta_\ell^{\alpha-1}\min\{\tau^{-1}\cdot d(u, X\setminus P_\ell(u)), \Delta_\ell\} \ge \Delta_\ell^\alpha$.

Similarly, if it is the case that $\left|\sum_{j<\ell} f_j^{(t)}(u) - f_j^{(t)}(v)\right| > \frac{\Delta_\ell^\alpha}{2}$ then it is enough that

- $\sigma_\ell^{(t)}(P_\ell(u)) = \sigma_\ell^{(t)}(P_\ell(v)) = 0$.

Again there is probability 1/2 for each of these. So we have probability at least $\lambda^{-1/k}/4$ for event $\mathcal{F}_{(i,u,v,t,\ell)}$. □

Claim 55. *Let $(i,u,v) \in M$, $t \in [D]$ and $i \le \ell < i+k$. Then*

$$\Pr\left[\neg\mathcal{F}_{(i,u,v,t,\ell)} \mid \bigwedge_{(i',u',v')\in Q} \hat{\mathcal{E}}_{(i',u',v')}\right] \le 1 - \lambda^{-1/k}/4,$$

for all $Q \subseteq \{(i',u',v') \mid i \ge i' \wedge \left(\hat{\mathcal{E}}_{(i,u,v)}, \hat{\mathcal{E}}_{(i',u',v')}\right) \notin E\}$.

Proof. First note that if $i' < i$, then event $\hat{\mathcal{E}}_{(i',u',v')}$ depend on events $\mathcal{F}_{(i',u',v',t',\ell')}$, where by definition $\ell' < i' + r \le i$ (recall that R contains only integers that divide by r), and these events depend only on the choice of partition for scales at most ℓ'. Hence the padding probability for u, v in scale ℓ and the choice of σ_ℓ is independent of these events.

If it is the case that $i' = i$, let $(i,u',v') \in M$ such that $\left(\hat{\mathcal{E}}_{(i,u,v)}, \hat{\mathcal{E}}_{(i,u',v')}\right) \notin E$. We know by the construction of G that $u', v' \notin B(u, 4\Delta_i)$ and $u', v' \notin B(v, 4\Delta_i)$. Hence u', v' are far from u, v and they fall into different clusters in every possible partition of scale ℓ. Moreover, the locality of our partition suggests that the padding of u, v in scale ℓ, for all $\ell \in [i, i+k)$, depends only on the partition of their local neighborhoods, $B(u, 2\Delta_\ell) \cup B(v, 2\Delta_\ell)$, which is disjoint from that of u', v'.

Note that even though event $\mathcal{F}_{(i,u,v,t,\ell)}$ is defined with respect to scales $\ell' \ge \ell$, since the padding probability and coloring by σ for u, v in scale ℓ will be as in Claim 54, no matter what happened in scales $\ell' < \ell$ or "far away" in scale ℓ. □

Now we are ready to prove the Lemma. For every coordinate $t \in [D]$, we have $r = \lceil(\theta/3)\log k\rceil$ possible values of ℓ. In each scale ℓ, by Claim 55 there is probability at most $q = 1 - \lambda^{-1/k}/4$ to fail, this probability is unaffected by of all other scales $\ell' < \ell$. Let $\mathcal{Y}_\ell$ be the indicator event for $\neg\mathcal{F}_{(i,u,v,t,\ell)}$. The probability that we failed for all scales $\ell \in [i, i+r)$ can be bounded by:

$$\Pr\left[\bigwedge_{\ell=i}^{i+r-1} \mathcal{Y}_\ell\right] = \prod_{\ell=i}^{i+r-1}\left(\Pr\left[\mathcal{Y}_\ell \mid \bigwedge_{j=i}^{\ell-1} \mathcal{Y}_j\right]\right) \le q^{\lceil(\theta/3)\log k\rceil}.$$

Let $z = q^{\lceil(\theta/3)\log k\rceil}$.

Case 1: Assume first that $(\theta/48)\lambda^{-1/k}\log k \ge 1$, then let Z_t be the event that we failed in the t-th coordinate (*i.e.* , $\mathcal{F}_{(i,u,v,t,\ell)}$ does not hold for all $\ell \in [i, i+r)$). Then $\Pr[Z_t] \le z$, and $Z = \sum_{t\in D} Z_t$. We know that $\mathbb{E}[Z] \le zD$, let $\beta \ge 1$ be such that $\mathbb{E}[Z] = \frac{zD}{\beta}$. Using Chernoff's bound implies that

$$\begin{aligned}\Pr[Z > qD] &= \Pr\left[Z > \left(\frac{q\beta}{z}\right)\mathbb{E}[Z]\right] \\ &\le \left(\frac{e^{q\beta/z-1}}{(q\beta/z)^{q\beta/z}}\right)^{zD/\beta} \\ &\le (ez/q)^{qD}.\end{aligned}$$

Note that $q \ge q^{(\theta/6)\log k}$ hence $z/q \le z^{1/2}$. By the assumption we have that $e \le e^{(\theta/48)\lambda^{-1/k}\log k} \le z^{-1/4}$. Since $q > 1/2$, and $q \le e^{-\lambda^{-1/k}/4}$ as well, follows that

$$\begin{aligned}(ez/q)^{qD} &\le z^{D/8} \\ &\le q^{(\theta/24)\log k\cdot D} \\ &\le e^{-\lambda^{-1/k}/4\cdot(\theta/24)\log k\cdot c\cdot\lambda^{1/k}\ln\lambda/(\alpha\theta)\cdot(1-\log(1-\alpha)/\log k)} \\ &= \lambda^{-c/96\cdot(\log k-\log(1-\alpha))/\alpha}.\end{aligned}$$

Taking $c = 96\cdot 59$ implies that $\Pr[Z > qD] \le \lambda^{-(53+6\log k-2\log(1-\alpha))/\alpha}$, as required.

Case 2: $(\theta/48)\lambda^{-1/k}\log k < 1$ we consider $\hat{Z}_t$ the event that for some $\ell \in [i, i+r)$, event $\mathcal{F}_{(i,u,v,t,\ell)}$ holds, we have that

$$\Pr[\hat{Z}_t] \ge 1 - \left(1 - \lambda^{-1/k}/4\right)^{(\theta/3)\log k} \ge 1 - e^{-\lambda^{-1/k}(\theta/48)\log k} \ge \lambda^{-1/k}(\theta/96)\log k,$$

the last inequality holds since $1 - e^{-x} \ge x/2$ when $0 \le x \le 1$. Let $q' = \lambda^{-1/k}(\theta/96)\log k$, and let $\hat{Z} = \sum_{t\in D}\hat{Z}_t$. Obviously $\mathbb{E}[\hat{Z}] \ge q'D$, using Chernoff bound implies that

$$\begin{aligned}\Pr[\hat{Z} \le \lambda^{-1/k}D] &\le \Pr[\hat{Z} \le \lambda^{-1/k}\mathbb{E}[\hat{Z}]/q'] \\ &= \Pr[\hat{Z} \le 96\mathbb{E}[\hat{Z}]/(\theta\log k)] \\ &\le e^{-\mathbb{E}[\hat{Z}](1-96/(\theta\log k))^2/2}.\end{aligned}$$

Since $\theta\log k \ge 192$ we have that $(1 - 96/(\theta\log k))^2 \ge 1/4$ hence

$$\Pr[\hat{Z} \le \lambda^{-1/k}D] \le e^{-q'D/4} \le e^{-\lambda^{-1/k}(\theta/96)\log k\cdot D}.$$

Again taking $c = 96\cdot 59$ implies that $\Pr[\hat{Z} \le \lambda^{-1/k}D] \le \lambda^{-(53+6\log k-2\log(1-\alpha))/\alpha}$, as required.

7 Scaling Distortion for Decomposable Metric

In this section we extend the theorem of [KLMN04] stating that any τ-decomposable metric embeds into l_p with distortion $O(\tau^{1/p-1} \cdot \log^{1/p} n)$, and give a version with scaling distortion to it.

Theorem 18. *Let $1 \le p \le \infty$. For any n-point τ-decomposable metric space (X,d) there exists an embedding $f : X \to l_p$ with coarse scaling distortion $O(\min\{(1/\tau)^{1-1/p}(\log \frac{2}{\epsilon})^{1/p}, \log \frac{2}{\epsilon}\})$ and dimension $O(\log^2 n)$.*

Proof overview: The embedding is similar to the one obtaining the $O(\log(2/\epsilon))$ scaling distortion result, with few important differences: In order to obtain improved distortion, which is done by using the properties of the l_p norm, we let the padding parameter appear in the embedding definition with power $1/p$. This definition implies that unlike the previous analysis, the padding parameters of all the scales can not be summed up, hence a different coordinate should be assigned for every scale (and hence we get the weaker dimension of $O(\log^2 n)$). In order to have only $O(\log n)$ different scales for every point (and not $O(\log(\mathrm{diam}(X)))$, we ignore those scales for which $\xi = 0$, *i.e.* those that do not have sufficient local growth rate. For this reason we must use our hierarchical probabilistic partitions, as those guarantee that points who share a cluster, also shared a cluster in all previous scales, thus by the uniformity of the function ξ these points have the same "coordinates arrangement" up to the current scale.

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 $f : X \to l_p^{D'D}$, by defining for each $1 \le t \le D$, an embedding $f^{(t)} : X \to l_p^{D'}$ and let $f = D^{-1/p} \bigoplus_{1 \le t \le D} f^{(t)}$.

Fix t, $1 \le t \le D$. In what follows we define $f^{(t)}$. We construct a strong $(\eta, 1/2)$-uniformly padded probabilistic 2-hierarchical partition $\hat{\mathcal{H}}$ as in Lemma 10, and let ξ be as defined in the lemma. Now fix a hierarchical partition $H = \{P_i\}_{i \in I} \in \mathcal{H}$. Let $D(x) = \sum_{0<i\in I} \xi_{P,i}(x)$. Another consequence of Lemma 10 is:

Claim 56. *For any $x \in X$: $D(x) \le D'$.*

Proof. Note that $\eta_{P,i}(x) \le 2^{-9}$, it follows that

$$D(x) = \sum_{0<i\in I} \xi_{P,i}(x) \le \sum_{0<i\in I} 2^{-9} \xi_{P,i}(x) \eta_{P,i}(x)^{-1} \le 32 \log n \le D'.$$

□

Let $J = \{1 \le j \le D' | j \in \mathbb{Z}\}$ the set of indexes of the coordinates, and for $x \in X$, let $J(x) = \{1 \le j \le D(x) | j \in \mathbb{Z}\}$ and let $\bar{J}(x) = J \setminus J(x)$. For each $x \in X$ and $i \in I$, let $\hat{j}_i(x) = \sum_{0<i'\le i} \xi_{P,i'}(x)$. For $j \in J(x)$, let $\hat{i}_j(x)$ be the smallest i such that $\hat{j}_i(x) = j$.

We have following important property:

Claim 57. *If for some $0 < i \in I$, we have that $P_i(x) = P_i(y)$ then for all $1 \le j \le \hat{j}_i(x)$, $\hat{i}_j(x) = \hat{i}_j(y)$.*

Proof. Since the partition is hierarchical we have that $P_\ell(x) = P_\ell(y)$ for all $0 < \ell \le i$. Since ξ is uniform with respect to H we have that $\xi_{P,\ell}(x) = \xi_{P,\ell}(y)$. This implies that $\hat{j}_\ell(x) = \hat{j}_\ell(y)$ for all $\ell \le i$. Let $1 \le j \le \hat{j}_i(x)$ and ℓ the smallest such that $\hat{j}_\ell(x) = \hat{j}_\ell(y) = j$, it follows that $\hat{i}_j(x) = \hat{i}_j(y) = \ell$. □

We define the embedding $f^{(t)}$ by defining the coordinates for each $x \in X$. For every $i \in I$ let $\sigma_i^{(t)} : X \to \{0,1\}$ be a uniform function with respect to P_i defined by letting $\{\sigma_i^{(t)}(C) | C \in P_i, 0 < i \in I\}$ be i.i.d symmetric $\{0,1\}$-valued Bernoulli random variables. Let $f^{(t)} : X \to l_p^{D'}$ be defined as $f^{(t)} = \bigoplus_{j \in [D']} \psi_j^{(t)}$. For each $j \in [D']$ define $\psi_j^{(t)} : X \to \mathbb{R}^+$ as

$$\psi_j^{(t)}(x) = \sigma_j^{(t)}(x) \cdot \varphi_j^{(t)}(x),$$

where $\varphi_j^{(t)} : X \to \mathbb{R}^+$ is defined as

$$\varphi_j^{(t)}(x) = \begin{cases} \min\left\{\frac{\xi_{P,i}(x)}{\eta_{P,i}(x)^{1/p}} d\left(x, X \setminus P_i(x)\right), \Delta_i\right\} & j \in J(x), i = \hat{i}_j(x) \\ 0 & j \in \bar{J}(x) \end{cases} \tag{21}$$

Define $g_i^{(t)} : X \times X \to \mathbb{R}^+$ as follows: $g_i^{(t)}(x,y) = \min\left\{\frac{\xi_{P,i}(x)}{\eta_{P,i}(x)^{1/p}} \cdot d(x,y), \Delta_i\right\}$ (Note that $g_i^{(t)}$ is nonsymmetric).

Claim 58. *For any $x, y \in X$ such that $D(x) \geq D(y)$:*

- *For any $j \in J(x) \cap J(y)$, let $i = \hat{i}_j(x)$ and $i' = \hat{i}_j(y)$, then*

$$|\psi_j^{(t)}(x) - \psi_j^{(t)}(y)| \leq \max\{g_i^{(t)}(x,y), g_{i'}^{(t)}(y,x)\}.$$

- *For any $j \in J(x) \setminus J(y)$, let $i = \hat{i}_j(x)$, then $|\psi_j^{(t)}(x) - \psi_j^{(t)}(y)| \leq g_i^{(t)}(x,y)$.*

Proof. Assume w.l.o.g $j \in J(x)$, and first we prove the first bullet. We have two cases. In Case 1, assume $P_i(x) = P_i(y)$ then by Claim 57 we get that $i' = \hat{i}_j(y) = \hat{i}_j(x) = i$. It follows that

$$|\psi_j^{(t)}(x) - \psi_j^{(t)}(y)| = \sigma_i^{(t)}(P_i(x)) \cdot |\varphi_i^{(t)}(x) - \varphi_i^{(t)}(y)|.$$

We will show that $\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y) \leq g_i^{(t)}(x,y)$. The bound $\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y) \leq \Delta_i$ is immediate. To prove $\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y) \leq \frac{\xi_{P,i}(x)}{\eta_{P,i}(x)^{1/p}} \cdot d(x,y)$ consider the value of $\varphi_j^{(t)}(y)$. Assume first $\varphi_j^{(t)}(y) = \frac{\xi_{P,i}(y)}{\eta_{P,i}(y)^{1/p}} \cdot d(y, X \setminus P_i(y))$. From the uniform padding property of H we get that $\xi_{P,i}(y) = \xi_{P,i}(x)$ and $\eta_{P,i}(y) = \eta_{P,i}(x)$ therefore

$$\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y) \leq \frac{\xi_{P,i}(x)}{\eta_{P,i}(x)^{1/p}} \cdot \left(d(x, X \setminus P_i(x)) - d(y, X \setminus P_i(x))\right) \leq \frac{\xi_{P,i}(x)}{\eta_{P,i}(x)^{1/p}} \cdot d(x,y).$$

In the second case $\varphi_j^{(t)}(y) = \Delta_i$ and therefore $\varphi_i^{(t)}(x) - \varphi_i^{(t)}(y) \leq \Delta_i - \Delta_i = 0$. Thus proving the claim in this case.

Next, consider Case 2 where $P_i(x) \neq P_i(y)$. In this case we have that $d(x, X \setminus P_i(x)) \leq d(x,y)$ which implies that

$$\psi_j^{(t)}(x) - \psi_j^{(t)}(y) \leq \varphi_j^{(t)}(x) \leq g_i(x,y). \tag{22}$$

The bound $g_{i'}^{(t)}(y,x)$ is obtained by considering $\varphi_j^{(t)}(y) - \varphi_j^{(t)}(x)$.

For the second bullet it must be that $P_i(x) \neq P_i(y)$ (otherwise we would get $i' = i$ which would be a contradiction). Since $j \notin J(y)$ then $\psi_j^{(t)}(y) = 0$ and we are done by (22). □

Lemma 59. *There exists a universal constant $C_1 > 0$ such that for any $\epsilon > 0$ and any $(x, y) \in \hat{G}(\epsilon)$:*

$$\|f^{(t)}(x) - f^{(t)}(y)\|_p^p \leq \ln(2/\epsilon) \cdot (C_1 \cdot d(x, y))^p.$$

Proof. Assume w.l.o.g $D(x) \geq D(y)$. Claim 58 implies that

$$\begin{aligned}
\|f^{(t)}(x) - f^{(t)}(y)\|_p^p &= \sum_{j \in J} |\psi_j^{(t)}(x) - \psi_j^{(t)}(y)|^p \\
&\leq \sum_{j \in J(x) \cap J(y)} \max\{g^{(t)}_{\hat{i}_j(x)}(x, y), g^{(t)}_{\hat{i}_j(y)}(y, x)\}^p + \sum_{j \in J(x) \setminus J(y)} g^{(t)}_{\hat{i}_j(x)}(x, y)^p \\
&\leq \sum_{0 < i \in I} \left(g_i^{(t)}(x, y)^p + g_i^{(t)}(y, x)^p \right). \qquad (23)
\end{aligned}$$

Now, define ℓ to be largest such that $\Delta_{\ell+4} \geq d(x, y) \geq \max\{r_{\epsilon/2}(x), r_{\epsilon/2}(y)\}$. If no such ℓ exists then let $\ell = 0$.

By Lemma 10 we have

$$\begin{aligned}
\sum_{0 < i \leq \ell} g_i^{(t)}(x, y)^p &\leq \sum_{0 < i \leq \ell} \frac{\xi_{P,i}(x)}{\eta_{P,i}(x)} \cdot d(x, y)^p \\
&\leq 2^{14} \cdot \ln\left(\frac{n}{|B(x, \Delta_{\ell+4})|} \right) \cdot d(x, y)^p \leq (2^{14} \ln(2/\epsilon)) \cdot d(x, y)^p.
\end{aligned}$$

We also have that

$$\sum_{\ell < i \in I} g_i^{(t)}(x, y)^p \leq \sum_{\ell < i \in I} \Delta_i^p \leq \Delta_\ell^p \leq 2^{5p} d(x, y)^p.$$

Therefore, using (23) we get

$$\|f^{(t)}(x) - f^{(t)}(y)\|_p^p = \sum_{0 < i \in I} \left(g_i^{(t)}(x, y)^p + g_i^{(t)}(y, x)^p \right) \leq 2 \left(2^{14} \ln(2/\epsilon) + 2^{5p} \right) \cdot d(x, y)^p.$$

□

Lemma 60. *There exists a universal constant $C_2 > 0$ such that for any $x, y \in X$, with probability at least $1/8$:*

$$\|f^{(t)}(x) - f^{(t)}(y)\|_p^p \geq \tau^{p-1} \cdot (C_2 \cdot d(x, y))^p.$$

Proof. Let $0 < \ell \in I$ be such that $8\Delta_\ell \leq d(x, y) \leq 16\Delta_\ell$. By Claim 3 we have that $\max\{\bar{\rho}(x, 2\Delta_\ell, \gamma_1, \gamma_2), \bar{\rho}(y, 2\Delta_\ell, \gamma_1, \gamma_2)\} \geq 2$. Assume w.l.o.g that $\bar{\rho}(x, 2\Delta_\ell, \gamma_1, \gamma_2) \geq 2$. It follows from Lemma 10 that $\xi_{P,\ell}(x) = 1$. As $\hat{\mathcal{H}}$ is $(\eta, 1/2)$-padded we have the following bound

$$\Pr[B(x, \eta_{P,\ell}(x)\Delta_\ell) \subseteq P_\ell(x)] \geq 1/2.$$

Therefore with probability at least $1/2$:

$$\left(\frac{\xi_{P,\ell}(x)}{\eta_{P,\ell}(x)^{1/p}} \cdot d(x, X \setminus P_\ell(x)) \right)^p \geq \frac{1}{\eta_{P,\ell}(x)} \cdot (\eta_{P,\ell}(x)\Delta_\ell)^p = \eta_{P,\ell}(x)^{p-1} \Delta_\ell^p \geq (\tau/8)^{p-1} \Delta_\ell^p, \qquad (24)$$

where the last inequality follows from the second property of Lemma 10.

Let $j = \hat{j}_\ell(x)$. Note that since $\xi_{P,\ell}(x) = 1$ we have that $\ell = \hat{i}_j(x)$. Since $\mathrm{diam}(P_\ell(x)) \le \Delta_\ell < d(x,y)$ we have that $P_\ell(y) \neq P_\ell(x)$. Now, if $j \notin J(y)$ then $\psi_j^{(t)}(y) = 0$ and with probability $1/2$ we have $\sigma_\ell(P_\ell(x)) = 1$ so that by (24) $|\psi_j^{(t)}(x) - \psi_j^{(t)}(y)|^p = \min\left\{\left(\frac{\xi_{P,\ell}(x)}{\eta_{P,\ell}(x)^{1/p}} \cdot d(x, X \setminus P_\ell(x))\right)^p, \Delta_i^p\right\} \ge (\tau/8)^{p-1}\Delta_\ell^p$. Otherwise, if $j \in J(y)$, then for $\ell' = \hat{i}_j(y)$ we have $P_\ell(x) \neq P_{\ell'}(y)$. We get that there is probability $1/4$ that $\sigma_\ell(P_\ell(x)) = 1$ and $\sigma_{\ell'}(P_{\ell'}(y)) = 0$ so that $|\psi_j^{(t)}(x) - \psi_j^{(t)}(y)|^p \ge (\tau/8)^{p-1}\Delta_\ell^p$.

We conclude that with probability at least $1/2 \cdot 1/4 = 1/8$:

$$\|f^{(t)}(x) - f^{(t)}(y)\|_p^p \ge |(\psi_j^{(t)}(x) - \psi_j^{(t)}(y))|^p \ge (\tau/8)^{p-1}\Delta_\ell^p \ge (\tau/8)^{p-1}2^{-4p}d(x,y)^p.$$

□

Lemma 61. *There exists a universal constants $C_1', C_2' > 0$ such that w.h.p for any $\epsilon > 0$ and any $(x,y) \in \hat{G}(\epsilon)$:*

$$C_2' \cdot \tau^{1-1/p} \cdot d(x,y) \le \|f(x) - f(y)\|_p \le C_1' \left(\ln(1/\epsilon)\right)^{1/p} \cdot d(x,y).$$

Proof. By definition

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

Lemma 59 implies that

$$\|f(x) - f(y)\|_p^p \le \ln(1/\epsilon) \left(C_1 \cdot d(x,y)\right)^p.$$

For $t \in [D]$ let $Z_t(x,y)$ be an indicator random variable for the event $\|f^{(t)}(x) - f^{(t)}(y)\|_p^p \ge \left((\tau/8)^{1-1/p}C_2 d(x,y)\right)^p$, and $Z = Z(x,y) = \sum_{t \in [D]} Z_t(x,y)$. By Lemma 60 we have that $\Pr[Z_t(x,y)] \ge 1/8$ thus $\mathbb{E}[Z] \ge D/8 \ge 16 \ln n$ for a constant $c \ge 2^7$. Applying Chernoff bounds

$$\Pr[Z < E[Z]/2] \le e^{-E[Z]/8} \le 1/n^2 \ .$$

Note that if $Z \ge \mathbb{E}[Z]/2$ then letting $G(x,y) = \{t \in [D] \mid Z_t(x,y)\}$, then $|G(x,y)| \ge D/16$ and then

$$\|f(x) - f(y)\|_p^p \ge \frac{1}{D} \sum_{t \in G(x,y)} \|f^{(t)}(x) - f^{(t)}(y)\|_p^p \ge (\tau^{1-1/p}C_2 \cdot d(x,y)/2^7)^p.$$

The proof is complete by applying a union bound on all pairs. □

8 Partial Embedding, Scaling Distortion and the ℓ_q-Distortion

In this section we show the relation between scaling distortion and the ℓ_q-distortion. The idea is to consider the values of ϵ which are some exponentially decreasing series (like all powers of $1/2$), then in the formula for the ℓ_q-distortion, partition the pairs according to which $\hat{G}(\epsilon)$ they belong to. We show that this analysis is tight in Lemma 2. Recall the definition of $\Phi, \hat{\Phi}$ is as in Definition 5.

Lemma 1. *Given an n-point metric space (X, d_X) and a metric space (Y, d_Y). If there exists an embedding $f : X \to Y$ with scaling distortion α then for any distribution Π over $\binom{X}{2}$:*[22]

$$\mathrm{dist}_q^{(\Pi)}(f) \le \left(2\int_{\frac{1}{2}\binom{n}{2}^{-1}\hat{\Phi}(\Pi)}^{1} \alpha(x\hat{\Phi}(\Pi)^{-1})^q dx\right)^{1/q} + \alpha(\hat{\Phi}(\Pi)^{-1}).$$

Proof. We may restrict to the case $\Phi(\Pi) \le \binom{n}{2}$. Otherwise $\hat{\Phi}(\Pi) > \binom{n}{2}$ and therefore $\mathrm{dist}_q^{(\Pi)}(f) \le \mathrm{dist}(f) \le \alpha(\hat{\Phi}(\Pi)^{-1})$. Recall that

$$\mathrm{dist}_q^{(\Pi)}(f) = \mathbb{E}_\Pi[\mathrm{dist}_f(u,v)^q]^{1/q}.$$

Define for each $\epsilon \in (0,1)$ the set $G(\epsilon)$ of the $(1-\epsilon)\binom{n}{2}$ pairs u, v of smallest distortion $\mathrm{dist}_f(u,v)$ over all pairs in $\binom{X}{2}$. Since f is a $(1-\epsilon)$-partial embedding for any $\epsilon \in (0,1)$ we have that for each $\{u,v\} \in G(\epsilon)$, $\mathrm{dist}_f(u,v) \le \alpha(\epsilon)$. Let $G_i = G(2^{-i}\hat{\Phi}(\Pi)^{-1}) \setminus G(2^{-(i-1)}\hat{\Phi}(\Pi)^{-1})$. Since α is a monotonic non-increasing function, it follows that

$$\begin{aligned}
\mathbb{E}_\Pi[\mathrm{dist}_f(u,v)^q] &= \sum_{u\neq v\in X} \pi(u,v)\mathrm{dist}_f(u,v)^q \\
&\le \sum_{\{u,v\}\in G(\hat{\Phi}(\Pi)^{-1})} \pi(u,v)\alpha(\hat{\Phi}(\Pi)^{-1})^q + \\
&\quad \sum_{i=1}^{\lfloor \log(\binom{n}{2}\hat{\Phi}(\Pi)^{-1})\rfloor} \sum_{\{u,v\}\in G_i} \pi(u,v)\alpha(2^{-i}\hat{\Phi}(\Pi)^{-1})^q \\
&\le \sum_{u\neq v\in X} \pi(u,v)\cdot\alpha(\hat{\Phi}(\Pi)^{-1})^q + \\
&\quad \sum_{i=1}^{\lfloor \log(\binom{n}{2}\hat{\Phi}(\Pi)^{-1})\rfloor} |G_i|\cdot\left(\frac{\hat{\Phi}(\Pi)}{\binom{n}{2}}\sum_{u\neq v\in X}\pi(u,v)\right)\cdot\alpha(2^{-i}\hat{\Phi}(\Pi)^{-1})^q \\
&\le \alpha(\hat{\Phi}(\Pi)^{-1})^q + \sum_{i=1}^{\lfloor \log(\binom{n}{2}\hat{\Phi}(\Pi)^{-1})\rfloor} 2^{-i}\cdot\alpha(2^{-i}\hat{\Phi}(\Pi)^{-1})^q \\
&\le \alpha(\hat{\Phi}(\Pi)^{-1})^q + \left(2\int_{\frac{1}{2}\binom{n}{2}^{-1}\hat{\Phi}(\Pi)}^{1}\alpha(x\hat{\Phi}(\Pi)^{-1})^q dx\right).
\end{aligned}$$

□

Lemma 62 (Coarse Scaling Distortion vs. Distortion of ℓ_q-Norm)**.** *Given an n-point metric space (X, d_X) and a metric space (Y, d_Y). If there exists an embedding $f : X \to Y$ with coarse scaling distortion α then for any distribution Π over $\binom{X}{2}$:*[23]

$$\mathrm{distnorm}_q^{(\Pi)}(f) \le \left(2\int_{\frac{1}{2}\binom{n}{2}^{-1}\hat{\Phi}(\Pi)}^{1} \alpha(x\hat{\Phi}(\Pi)^{-1})^q dx\right)^{1/q} + \alpha(\hat{\Phi}(\Pi)^{-1}).$$

[22] Assuming the integral is defined. We note that lemma is stated using the integral for presentation reasons.
[23] Assuming the integral is defined.

Proof. We may restrict to the case $\Phi(\Pi) \leq \binom{n}{2}$. Otherwise $\hat{\Phi}(\Pi) > \binom{n}{2}$ and therefore $\mathrm{distnorm}_q^{(\Pi)}(f) \leq \mathrm{dist}(f) \leq \alpha(\hat{\Phi}(\Pi)^{-1})$. Recall that

$$\mathrm{distnorm}_q^{(\Pi)}(f) = \frac{\mathbb{E}_\Pi[d_Y(f(u), f(v))^q]^{1/q}}{\mathbb{E}_\Pi[d_X(u,v)^q]^{1/q}}.$$

For $\epsilon \in (0,1)$ recall that $\hat{G}(\epsilon) = \{\{x,y\} \in \binom{X}{2} | d(x,y) \geq \max\{r_{\epsilon/2}(x), r_{\epsilon/2}(y)\}\}$. Since $(f, \hat{G})$ is a $(1-\epsilon)$-partial embedding for any $\epsilon \in (0,1)$ we have that for each $\{u,v\} \in \hat{G}(\epsilon)$, $\mathrm{dist}_f(u,v) \leq \alpha(\epsilon)$. Let $\hat{G}_i = \hat{G}(2^{-i}\hat{\Phi}(\Pi)^{-1}) \setminus \hat{G}(2^{-(i-1)}\hat{\Phi}(\Pi)^{-1})$. We first need to prove the following property:

$$\sum_{\{u,v\} \in \hat{G}_i} d_X(u,v)^q \leq 2^{-i}\hat{\Phi}(\Pi)^{-1} \sum_{u \neq v \in X} d_X(u,v)^q.$$

To prove this fix some $u \in X$. Let $S = \{v | \{u,v\} \notin \hat{G}(2^{-(i-1)}\hat{\Phi}(\Pi)^{-1})\}$. Then $S = B(u, r_{2^{-i}\hat{\Phi}(\Pi)^{-1}}(u))$. Thus, $|S| = 2^{-i}\hat{\Phi}(\Pi)^{-1}n$ and for each $v \in S$, $v' \in \bar{S}$ we have $d(u,v) \leq d(u,v')$. It follows that:

$$\begin{aligned}
\sum_{v; u \neq v \in X} d_X(u,v)^q &= \sum_{v \in S} d_X(u,v)^q + \sum_{v \in \bar{S}} d_X(u,v)^q \\
&\geq |S| \cdot \frac{\sum_{v \in S} d_X(u,v)^q}{|S|} + |\bar{S}| \cdot \frac{\sum_{v \in S} d_X(u,v)^q}{|S|} = \frac{n}{|S|} \sum_{v \in S} d_X(u,v)^q.
\end{aligned}$$

Since α is a monotonic non-increasing function, it follows that

$$\begin{aligned}
\mathbb{E}_\Pi[d_Y(f(u),f(v))^q] &= \sum_{u\neq v\in X} \pi(u,v) d_Y(f(u),f(v))^q \\
&= \sum_{u\neq v\in X} \pi(u,v) d_X(u,v)^q \mathrm{dist}_f(u,v)^q \\
&\leq \sum_{\{u,v\}\in \hat{G}(\hat{\Phi}(\Pi)^{-1})} \pi(u,v) d_X(u,v)^q \alpha(\hat{\Phi}(\Pi)^{-1})^q + \\
&\quad \sum_{i=1}^{\lfloor \log\left(\binom{n}{2}\hat{\Phi}(\Pi)^{-1}\right)\rfloor} \sum_{\{u,v\}\in\hat{G}_i} \pi(u,v) d_X(u,v)^q \alpha(2^{-i}\hat{\Phi}(\Pi)^{-1})^q \\
&\leq \sum_{u\neq v\in X} \pi(u,v) d_X(u,v)^q \cdot \alpha(\hat{\Phi}(\Pi)^{-1})^q + \\
&\quad \sum_{i=1}^{\lfloor \log\left(\binom{n}{2}\hat{\Phi}(\Pi)^{-1}\right)\rfloor} \sum_{\{u,v\}\in\hat{G}_i} d_X(u,v)^q \cdot \hat{\Phi}(\Pi) \cdot \min_{w\neq z\in X} \pi(w,z) \cdot \alpha(2^{-i}\hat{\Phi}(\Pi)^{-1})^q \\
&\leq \sum_{u\neq v\in X} \pi(u,v) d_X(u,v)^q \cdot \alpha(\hat{\Phi}(\Pi)^{-1})^q + \\
&\quad \sum_{i=1}^{\lfloor \log\left(\binom{n}{2}\hat{\Phi}(\Pi)^{-1}\right)\rfloor} \sum_{u\neq v\in X} 2^{-i} d_X(u,v)^q \cdot \min_{w\neq z\in X} \pi(w,z) \cdot \alpha(2^{-i}\hat{\Phi}(\Pi)^{-1})^q \\
&\leq \sum_{u\neq v\in X} \pi(u,v) d_X(u,v)^q \cdot \alpha(\hat{\Phi}(\Pi)^{-1})^q + \\
&\quad \sum_{i=1}^{\lfloor \log\left(\binom{n}{2}\hat{\Phi}(\Pi)^{-1}\right)\rfloor} \sum_{u\neq v\in X} \pi(u,v) d_X(u,v)^q \cdot 2^{-i} \cdot \alpha(2^{-i}\hat{\Phi}(\Pi)^{-1})^q \\
&\leq \mathbb{E}_\Pi[d_X(u,v)^q] \cdot \left[\alpha(\hat{\Phi}(\Pi)^{-1})^q + \left(2\int_{\frac{1}{2}\binom{n}{2}^{-1}\hat{\Phi}(\Pi)}^{1} \alpha(x\hat{\Phi}(\Pi)^{-1})^q dx\right)\right].
\end{aligned}$$

□

8.1 Distortion of ℓ_q-Norm for Fixed q

Lemma 63. *Let $1 \leq q \leq \infty$. For any finite metric space (X,d), there exists an embedding f from X into a* star *metric such that for any non-degenerate distribution Π:* $\mathrm{distnorm}_q^{(\Pi)}(f) \leq 2^{1/q}(2^q-1)^{1/q}\Phi(\Pi)^{1/q}$. *In particular:* $\mathrm{distnorm}_q(f) \leq 2^{1/q}(2^q-1)^{1/q} \leq \sqrt{6}$.

Proof. Let $w \in X$ be the point that minimizes $(\sum_{x\in X} d(w,x)^q)^{1/q}$. Let $Y = X \cup \{r\}$. Define a star metric (Y,d') where r is the center and for every $x \in X$: $d'(r,x) = d(w,x)$. Thus $d'(x,y) =$

$d(w,x)+d(w,y)$. Then

$$\begin{aligned}
\mathbb{E}_\Pi[d'(u,v)^q] &= \sum_{u\neq v\in X}\pi(u,v)d'(u,v)^q \leq \sum_{u\neq v\in X}\pi(u,v)\left(d(u,w)+d(w,v)\right)^q \\
&\leq (2^q-1)\sum_{u\neq v\in X}\pi(u,v)\left(d(u,w)^q+d(w,v)^q\right) \\
&\leq (2^q-1)\sum_{u\neq v\in X}\left(\Phi(\Pi)\min_{s\neq t\in X}\pi(s,t)\right)\cdot\left(d(u,w)^q+d(w,v)^q\right) \\
&= (2^q-1)\cdot\Phi(\Pi)\min_{s\neq t\in X}\pi(s,t)\cdot\frac{n-1}{2}\left(\sum_{u\in X}d(u,w)^q+\sum_{v\in X}d(w,v)^q\right) \\
&\leq (2^q-1)\cdot\Phi(\Pi)\cdot(n-1)\cdot\frac{1}{n}\sum_{z\in X}\sum_{u\in X}\min_{s\neq t\in X}\pi(s,t)\cdot d(u,z)^q \\
&\leq 2(2^q-1)\cdot\Phi(\Pi)\cdot\sum_{u\neq v\in X}\pi(u,v)\cdot d(u,v)^q \\
&= 2(2^q-1)\cdot\Phi(\Pi)\cdot\mathbb{E}_\Pi[d(u,v)^q].
\end{aligned}$$

□

9 Probabilistic Embedding with Scaling Distortion into Trees

In this section we prove Theorem 19.[24]

Theorem 19. *For any n-point metric space (X,d) there exists a probabilistic embedding into a distribution over ultrametrics with coarse scaling distortion $O(\log\frac{2}{\epsilon})$.*

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

Definition 23. An *ultrametric* (HST[25]) is a metric space whose elements are the leaves of a rooted tree T. Each vertex $u\in T$ is associated with a label $\Delta(u)\geq 0$ 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 $\Delta(u)\leq\Delta(v)$. 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.

Proof of Theorem 19. Let $\Delta=\Delta(X)$. For every $i\in\mathbb{N}$ let P_i be a $\Delta 2^{-i}$ bounded probabilistic partition given by Corollary 6, and let η_i be as in the corollary. We build an ultrametric U by defining a labeled tree, in the following manner. For every $i>1$ we iteratively alter P_i into P_i' by replacing each $C\in P_i$ with the clusters $\{C\cap D\mid D\in P_{i-1}\}$. Each cluster $C\in P_i'$ defines a node in the tree, its parent is the cluster in P_{i-1} that contains it, and the label of every cluster in P_i' is $\Delta 2^{-i}$. The root has label Δ and is connected to all the clusters in P_1. Finally, leaves are formed by clusters that contain only one node.

[24] A similar theorem was independently given in [CDG+09].

[25] A k-HST [Bar96] is defined similarly while requiring that $\Delta(u)\leq\Delta(v)/k$. Any ultrametric k-embeds [Bar98] and $O(k/\log k)$-probabilistically embeds [BCR01] in a k-HST.

For any $u,v \in G(\epsilon)$ let t be the integer such that $\Delta 2^{-(t+1)} \le d(u,v) < \Delta 2^{-t}$. Let $\rho_i(u) = \rho(u, 2\Delta 2^{-i}, 2, 1/32)$. Choose for each $1 \le i \le t-6$, $\delta_i = \exp\{-\frac{2^6 d(u,v)\ln\rho_i(u)}{\Delta 2^{-i}}\}$ and note that $\delta_i \le 1$. Recall that in Corollary 6 $\eta_i(u) = \min\left\{\frac{\ln(1/\delta_i)}{2^6\ln\rho_i(u)}, 2^{-6}\right\} = \frac{d(u,v)}{\Delta 2^{-i}}$ (because $\frac{d(u,v)}{\Delta 2^{-i}} \le 2^{-6}$, and if $\ln(\rho_i(u)) = 0$ we define $\eta_i(u)$ in a continuous manner as $\frac{d(u,v)}{\Delta 2^{-i}}$).

If it is the case that $\delta_i \ge 1/2$ we may use the padding property shown in Corollary 6 and argue that for any $1 \le i \le t-6$

$$\Pr[B(u,d(u,v)) \nsubseteq P_i(u)] = \Pr[B(u,\eta_i(u)\Delta 2^{-i}) \nsubseteq P_i(u)] \le 1-\delta \le \frac{2^6 d(u,v)\ln\rho_i(u)}{\Delta 2^{-i}} ,$$

however if $\delta < 1/2$ it will imply that $\Delta 2^{-i} < 2^6 d(u,v)(\ln\rho_i(u))/\ln 2 \le 2^7 d(u,v)\ln\rho_i(u)$ and we will use that $\Pr[B(u,d(u,v)) \nsubseteq P_i(u)] \le 1$. Finally write

$$\begin{aligned}
E[d_U(u,v)] &\le \sum_{i=1}^{t} \Pr[B(u,d(u,v)) \nsubseteq P_i(u)]\Delta 2^{-i} \\
&\le \sum_{i=t-5}^{t} \Delta 2^{-i} + \sum_{i=1}^{t-6} 2^7 d(u,v)\ln\rho_i(u) \\
&\le 2^7 d(u,v) + 2^{10}\ln\left(\frac{n}{|B(u,\Delta 2^{-t})|}\right)\cdot d(u,v) \\
&= O\left(\ln\frac{2}{\epsilon}\right)\cdot d(u,v) ,
\end{aligned}$$

where the third inequality follows by a telescopic sum argument. □

10 Partial Embedding

In this section we prove theorems on partial embedding. In particular we show that practically any embedding of a finite metric space (X,d) into l_p can be converted to a $(1-\epsilon)$ partial embedding, where the dependence of the distortion on the cardinality of X is replaced with $2/\epsilon$.[26]

10.1 Partial Embedding into l_p

Definition 24. We say that a family of metric spaces $\mathcal{X}$ is *subset-closed*, if for any $X \in \mathcal{X}$ every sub-metric $Y \subseteq X$ is also in $\mathcal{X}$.

Theorem 24 (Partial Embedding Upper Bound)**.** *Let $\mathcal{X}$ be a subset-closed family of finite metric spaces. If for any $m \ge 1$ and any m-point metric space from $\mathcal{X}$ there exists an embedding into l_p with distortion $\alpha(m)$ and dimension $\beta(m)$. Then there exists is a universal constant $C > 0$, such that for any $X \in \mathcal{X}$ and for any $\epsilon \in (0,1)$ there exists a $(1-\epsilon)$ partial embedding into l_p with distortion $\alpha(\frac{C\log(2/\epsilon)}{\epsilon})$ and dimension $\beta(\frac{C\log(2/\epsilon)}{\epsilon}) + O(\log(2/\epsilon))$.*

[26] Results similar to those appearing in the section have been independently shown in [CDG+09].

Proof. The idea of the proof is to choose a constant set of beacons, embed them, then embed all the other points according to the nearest beacon, and add some auxiliary coordinates. Formally, given $\epsilon > 0$ let $\hat{\epsilon} = \epsilon/20$, and $t = \lceil 100 \log\left(\frac{1}{\hat{\epsilon}}\right) \rceil$. Let B be a uniformly distributed random set of $\frac{t}{\hat{\epsilon}}$ points in X (the beacons). Let g be an embedding from B into l_p with distortion $\alpha(\frac{t}{\hat{\epsilon}})$ and dimension $\beta(\frac{t}{\hat{\epsilon}})$, which exists since $B \in \mathcal{X}$. Let $\{\sigma_j(u) \mid u \in X, 1 \leq j \leq t\}$ be i.i.d symmetric $\{0,1\}$-valued Bernoulli random variables. Define the following functions:

$$\forall u \in X, 1 \leq j \leq t \quad h_j(u) = \sigma_j(u) r_{\hat{\epsilon}}(u) t^{-1/p}$$

$$\forall u \in X \quad f(u) = g(b) \quad \text{where } b \in B \text{ such that } d_X(u,b) = d_X(u,B)$$

The embedding will be $\varphi = f \oplus h$. Let $G' = \binom{X}{2} \setminus (D_1 \cup D_2)$ where $D_1 = \{(u,v) \mid d_X(u,v) \leq \max\{r_{\hat{\epsilon}}(u), r_{\hat{\epsilon}}(v)\}\}$ and $D_2 = \{(u,v) \mid d_X(u,B) \geq r_{\hat{\epsilon}}(u), d_X(v,B) \geq r_{\hat{\epsilon}}(v)\}$. Observe that $|D_1| \leq \hat{\epsilon} n^2$. For any $u \in X$ $\Pr[d_X(u,B) \geq r_{\hat{\epsilon}}(u)] \leq (1 - t/(n\hat{\epsilon}))^{\hat{\epsilon} n} \leq e^{-t} \leq \hat{\epsilon}$ so by Markov inequality with probability at least 1/2, $|D_2| \leq 2\hat{\epsilon} n^2$. We begin with an upper bound on φ for all $(x,y) \in G'$:

$$\begin{aligned}
\left\|\varphi(u) - \varphi(v)\right\|_p^p &= \|f(u) - f(v)\|_p^p + \sum_{j=1}^{t} |h_j(u) - h_j(v)|^p \\
&\leq (3 d_X(u,v))^p + \sum_{j=1}^{t} \left| t^{-1/p} \max\{r_{\hat{\epsilon}}(u), r_{\hat{\epsilon}}(v)\} - 0 \right|^p \\
&\leq (3^p + 1)\,(d_X(u,v))^p .
\end{aligned}$$

We now partition G' into two sets $G_1 = \{(u,v) \in G' \mid \max\{r_{\hat{\epsilon}}(u), r_{\hat{\epsilon}}(v)\} \geq d_X(u,v)/4\}$ and $G_2 = G' \setminus G_1$. For any $(u,v) \in G_1, 1 \leq j \leq t$, assume w.l.o.g that $r_{\hat{\epsilon}}(u) \geq r_{\hat{\epsilon}}(v)$, and let $\mathcal{E}_j(u,v)$ be the event

$$\mathcal{E}_j(u,v) = \{h_j(u) = \frac{r_{\hat{\epsilon}}(u)}{t^{1/p}} \wedge h_j(v) = 0\}.$$

Then $\Pr[\mathcal{E}_j(u,v)] = \frac{1}{4}$. Let $A(u,v) = \sum_{j=1}^{t} \mathbf{1}_{\mathcal{E}_j(u,v)}$, then $\mathbb{E}[A(u,v)] = t/4$, using Chernoff's bound we can bound the probability that $A(u,v)$ is smaller than half it's expectation:

$$\Pr[A(u,v) \leq t/8] \leq e^{-t/50} \leq \hat{\epsilon}.$$

Let $D_3 = \{(u,v) \in G_1 \mid A(u,v) \leq t/8\}$ so by Markov inequality with probability at least 1/2, $|D_3| \leq 2\hat{\epsilon} n^2$. Therefore, for any $(u,v) \in G_1 \setminus D_3$ we lower bound the contribution.

$$\left\|\varphi(u) - \varphi(v)\right\|_p^p \geq \sum_{j=1}^{t} |h_j(u) - h_j(v)|^p \geq (t/8) \left(r_{\hat{\epsilon}}(u) t^{-1/p}\right)^p \geq 1/8 (d_X(u,v)/4)^p.$$

For any $(u,v) \in G_2$ let b_u, b_v be the beacons such that $f(u) = g(b_u), f(v) = g(b_v)$. Due to the definition of D_2 and G_2 and from the triangle inequality follows

$$d_X(b_u,b_v) \geq d_X(u,v) - d_X(u,b_u) - d_X(v,b_v) \geq d_X(u,v) - \frac{d_X(u,v)}{2} = \frac{d_X(u,v)}{2}.$$

Therefore, we lower bound the contribution of $(u,v) \in G_2$.

$$\begin{aligned}
\left\|\varphi(u) - \varphi(v)\right\|_p^p \geq \left\|f(u) - f(v)\right\|_p^p &= \left\|g(b_u) - g(b_v)\right\|_p^p \\
&\geq \frac{1}{\alpha(\frac{t}{\hat{\epsilon}})} d_X(b_u,b_v) \geq \frac{d_X(u,v)}{2\alpha(\frac{t}{\hat{\epsilon}})}.
\end{aligned}$$

Finally we note that $G = \binom{X}{2} \setminus (D_1 \cup D_2 \cup D_3)$ so with probability at least $1/4$ we have $|G| \geq \binom{n}{2} - 5\hat{\epsilon}n^2 \geq \binom{n}{2} - \epsilon n/4 \geq (1-\epsilon)\binom{n}{2}$ as required. $\square$

Corollary 64 (Partial Embedding Upper Bounds)**.** *For any $\epsilon \in (0,1)$:*

1. *Any finite metric space has a $(1-\epsilon)$ partial embedding into l_p with distortion $O(\log \frac{1}{\epsilon})$ and dimension $O(\log \frac{1}{\epsilon})$.*

2. *Any finite metric space has a $(1-\epsilon)$ partial embedding into l_p with distortion $O(\lceil (\log \frac{2}{\epsilon})/p \rceil)$ and dimension $e^{O(p)} \log \frac{1}{\epsilon}$.*

3. *Any negative type metric (in particular l_1 metrics) has a $(1-\epsilon)$ partial embedding into ℓ_2 with distortion $O\left(\sqrt{\log \frac{1}{\epsilon}} \log\log \frac{1}{\epsilon}\right)$ and dimension $O(\log \frac{1}{\epsilon})$.*

4. *Any tree metric has a $(1-\epsilon)$ partial embedding into ℓ_2 with distortion $O\left(\sqrt{\log\log \frac{1}{\epsilon}}\right)$ and dimension $O(\log \frac{1}{\epsilon})$.*

This follows from known upper bounds. (1) and (2) from [Bou85, Mat90] with dimension bound due to Theorem 10, (3) from [ALN05], and (4) from [Bou86, Mat99].

10.2 Coarse Partial Embedding into l_p

We now consider the coarse version of partial embedding into l_p. The trade off in getting a coarse $(1-\epsilon)$ partial embedding is in higher dimension and stronger requirements.

Definition 25 (Strongly non-expansive)**.** Let f is an embedding from X into l_p^k, where $f = (\eta_1 f_1, \ldots, \eta_k f_k)$ and $\sum_{i=1}^k \eta_i^p = 1$, we say that f is *strongly non-expansive* if it is non-expansive and

$$\forall u, v \in X, i = 1 \ldots k, \quad |f_i(u) - f_i(v)| \leq d(u,v).$$

Notice that the requirement of strongly non-expansion is not so restricting, since almost every known embedding can be converted to a strongly non-expansive one. In particular any generalized Fréchet embedding is strongly non-expansive.

Theorem 25. *Consider a fixed space l_p, $p \geq 1$. Let $\mathcal{X}$ be a subset-closed family of finite metric spaces such that for any $n \geq 1$ and any n-point metric space $X \in \mathcal{X}$ there exists a strongly non-expansive embedding $\phi_X : X \to l_p$ with distortion $\alpha(n)$ and dimension $\beta(n)$. Then there exists a universal constant $C > 0$ such that for any metric space $X \in \mathcal{X}$ and any $\epsilon > 0$ we have a* coarse *$(1-\epsilon)$ partial embedding into l_p, with distortion $O(\alpha(\frac{C}{\epsilon}))$ and dimension $\beta(\frac{C}{\epsilon}) \cdot O(\log n)$.*

Proof. This embedding is quite similar to the previous one, only this time we choose $O(\log n)$ sets of beacons in order to succeed in some events with high probability - depending on n instead of ϵ. This makes the proof more complex, and we need to embed each point according to the "best" beacon *in each coordinate*. Given $\epsilon > 0$ let $\hat{\epsilon} = \epsilon/4$, let $\tau = \lceil 100 \log n \rceil$ and denote $T = \{t \in \mathbb{N} \mid 1 \leq t \leq \tau\}$. Let $m = \lceil \frac{1}{\hat{\epsilon}} \rceil$. For each $t \in T$, let B_t be an independent uniformly distributed random set of m points in X. For each $t \in T$ let $\vec{\phi^{(t)}} = (\eta_1^{(t)} \phi_1^{(t)}, \ldots, \eta_{\beta(m)}^{(t)} \phi_{\beta(m)}^{(t)})$ be a strongly non-expansive embedding from B_t into l_p with distortion $\alpha(m)$ and dimension $\beta(m)$. Let $I = \{i \in \mathbb{N} \mid 1 \leq i \leq \beta(m)\}$. When clear from the context we omit the $\vec{\phi}^{(t)}$ superscript and simply write $\vec{\phi}$. Let $\{\sigma_t(u) \mid u \in X, t \in T\}$ be i.i.d symmetric $\{0,1\}$-valued Bernoulli random variables. Define the following functions:

$$\forall u \in X, t \in T \quad h^{(t)}(u) = \sigma_t(u) r_{\hat{\epsilon}}(u) \tau^{-1/p}$$

$$\forall u \in X, i \in I, t \in T \qquad f_i^{(t)}(u) = \eta_i^{(t)} \min_{b \in B_t} \{d(u,b) + \phi_i^{(t)}(b)\} \tau^{-1/p}$$

Let $f^{(t)} = (f_1^{(t)}, \ldots, f_{\beta(m)}^{(t)})$, $f = (f^{(1)}, \ldots, f^{(\tau)})$, and $h = (h^{(1)}, \ldots, h^{(\tau)})$, the final embedding will be $\varphi = f \oplus h$. Let $D = \{(u,v) \mid d(u,v) \le \max\{r_{\hat{\epsilon}}(u), r_{\hat{\epsilon}}(v)\}\}$ and $G = \binom{X}{2} \setminus D$, as in Theorem 24 before: $|D| \le \hat{\epsilon} n^2$. We begin by an upper bound for all $(u,v) \in G$: For any $t \in T, i \in I$ let $b_i^t \in B_t$ be the beacon that minimizes $f_i^{(t)}(v)$:

$$\begin{aligned}
\Big\|\varphi(u) - \varphi(v)\Big\|_p^p &= \Big\|f(u) - f(v)\Big\|_p^p + \Big\|h(u) - h(v)\Big\|_p^p \\
&\le \sum_{t \in T} \sum_{i \in I} \Big|f_i^{(t)}(u) - f_i^{(t)}(v)\Big|^p + \sum_{t \in T} \Big(\tau^{-1/p} \max\{r_{\hat{\epsilon}}(u), r_{\hat{\epsilon}}(v)\}\Big)^p \\
&\le \sum_{t \in T} \tau^{-1} \sum_{i \in I} \Big|\eta_i^{(t)} \min_{b \in B_t}\{d(u,b) + \phi_i^{(t)}(b)\} - \eta_i^{(t)} \min_{b \in B_t}\{d(v,b) + \phi_i^{(t)}(b)\}\Big|^p + d(u,v)^p \\
&\le \sum_{t \in T} \tau^{-1} \sum_{i \in I} \eta_i^{(t)p} \Big|\big(d(u,b_i^t) + \phi_i^{(t)}(b_i^t) - d(v,b_i^t) - \phi_i^{(t)}(b_i^t)\big)\Big|^p + d(u,v)^p \\
&\le \sum_{t \in T} \tau^{-1} \sum_{i \in I} \eta_i^{(t)p} d(u,v)^p + d(u,v)^p \\
&\le 2d(u,v)^p.
\end{aligned}$$

(recall that for any $t \in T$, $\sum_{i \in I} \eta_i^{(t)p} = 1$) We now partition G into two sets $G_1 = \{(u,v) \in G \mid \max\{r_{\hat{\epsilon}}(u), r_{\hat{\epsilon}}(v)\} \ge \frac{d(u,v)}{16\alpha(m)}$ and $G_2 = G \setminus G_1$. For any $(u,v) \in G_1, t \in T$, assume w.l.o.g that $r_{\hat{\epsilon}}(u) \ge r_{\hat{\epsilon}}(v)$, and let $\mathcal{E}_t(u,v)$ be the event

$$\mathcal{E}_t(u,v) = \{h^{(t)}(u) = r_{\hat{\epsilon}}(u) \wedge h^{(t)}(v) = 0\}.$$

Then $\Pr[\mathcal{E}_t(u,v)] = \frac{1}{4}$. Let $A(u,v) = \sum_{t \in T} \mathbf{1}_{\mathcal{E}_t(u,v)}$, then $\mathbb{E}[A(u,v)] = \tau/4$, using Chernoff bound we can bound the probability that $A(u,v)$ is smaller than half it's expectation:

$$\Pr[A(u,v) \le \tau/8] \le e^{-\tau/50} \le 1/n^2.$$

Therefore with probability greater than $1/2$, for any $(u,v) \in G_1$, $A(u,v) \ge \tau/8$. Assume that this happens, then we can lower bound the contribution for any $(u,v) \in G_1$:

$$\Big\|\varphi(u) - \varphi(v)\Big\|_p^p \ge \sum_{t \in T} \Big|h^{(t)}(u) - h^{(t)}(v)\Big|^p \ge (\tau/8)\,(r_{\hat{\epsilon}}(u))^p \ge \frac{\tau}{8}\left(\frac{d(u,v)}{16\alpha(m)}\right)^p.$$

For any $(u,v) \in G_2, t \in T$ let $b_u, b_v \in B_t$ the nearest beacons to u, v respectively. Let

$$\mathcal{F}_t(u,v) = \{b_u \in B\big(u, r_{\hat{\epsilon}}(u)\big) \wedge b_v \in B\big(v, r_{\hat{\epsilon}}(v)\big)\}.$$

Then $\Pr[\mathcal{F}_t(u,v)] \geq 1-2/e > 1/4$, since for any $u \in X$, $\Pr[d(u,B_t) > r_{\hat{\epsilon}}(u)] = (1-\hat{\epsilon})^{1/\hat{\epsilon}} \leq e^{-1}$. Let $Z(u,v) = \sum_{t\in T} \mathbf{1}_{\mathcal{F}_t(u,v)}$, then $\mathbb{E}[Z(u,v)] \geq \tau/4$, using Chernoff bound we can bound the probability that $Z(u,v)$ is smaller than half it's expectation:

$$\Pr[Z(u,v) \leq \tau/8] \leq e^{-\tau/50} \leq 1/n^2$$

Therefore with probability greater than 1/2 for any $(u,v) \in G_2$, $Z(u,v) \geq \tau/8$, assume from now on that this is the case. Fix a $t \in T$ such that $\mathcal{F}_t(u,v)$ happened. We have

$$\max\left\{d(u,b_u), d(v,b_v)\right\} \leq \frac{d(u,v)}{16\alpha(m)}.$$

Claim 65.

$$\tau^{1/p}\eta_i^{-1}|f_i(u)-f_i(v)| \geq \Big||\phi_i(b_u)-\phi_i(b_v)| - (d(u,b_u)+d(v,b_v))\Big|.$$

Proof. W.l.o.g assume that $f_i(u) \geq f_i(v)$, then let $b_i \in B_t$ be the beacon minimizing $f_i(u)$. Since for every $i \in I$, $\quad \phi_i(b_u) - \phi_i(b_i) \leq d(b_u,b_i)$ we get

$$\tau^{1/p}\eta_i^{-1}f_i(u) = d(u,b_i)+\phi_i(b_i) \geq d(u,b_i)+\phi_i(b_u)-d(b_u,b_i) \geq \phi_i(b_u)-d(u,b_u)$$

and

$$\tau^{1/p}\eta_i^{-1}f_i(v) \leq d(v,b_v)+\phi_i(b_v).$$

$\square$

Let $J = \{i \in I \mid \quad |\phi_i(b_u)-\phi_i(b_v)| \geq \frac{d(u,v)}{4\alpha(m)}\}$. We claim that $\sum_{i\in J}\eta_i^p\,|\phi_i(b_u)-\phi_i(b_v)|^p \geq \left[\frac{d(u,v)}{4\alpha(m)}\right]^p$. Assume by contradiction that it is not the case, then

$$\begin{aligned}
\left\|\vec{\phi}(b_u)-\vec{\phi}(b_v)\right\|_p^p &= \sum_{i\in J}\eta_i^p\Big|\phi_i(b_u)-\phi_i(b_v)\Big|^p + \sum_{i\notin J}\eta_i^p\Big|\phi_i(b_u)-\phi_i(b_v)\Big|^p \\
&< \left[\frac{d(u,v)}{4\alpha(m)}\right]^p + \sum_{i\notin J}\eta_i^p\left[\frac{d(u,v)}{4\alpha(m)}\right]^p \\
&\leq 2\left[\frac{d(u,v)}{4\alpha(m)}\right]^p < \left[\frac{d(b_u,b_v)}{\alpha(m)}\right]^p.
\end{aligned}$$

The last inequality follows since $d(b_u,b_v) \geq d(u,v) - 2\frac{d(u,v)}{16\alpha(m)} \geq \frac{7}{8}d(u,v)$.

Thus contradicting the fact that $\vec{\phi}$ has distortion $\alpha(m)$ on B_t. Now

$$\begin{aligned}
\left\| f^{(t)}(u) - f^{(t)}(v) \right\|_p^p &= \sum_{i \in I} \left| f_i^{(t)}(u) - f_i^{(t)}(v) \right|^p \\
&\geq \tau^{-1} \sum_{i \in J} \eta_i^p \left| \phi_i(b_u) - d(u, b_u) - d(v, b_v) - \phi_i(b_v) \right|^p \\
&\geq \tau^{-1} \sum_{i \in J} \eta_i^p \Big| |\phi_i(b_u) - \phi_i(b_v)| - |d(u, b_u) + d(v, b_v)| \Big|^p \\
&\geq \tau^{-1} \sum_{i \in J} \eta_i^p \Big| |\phi_i(b_u) - \phi_i(b_v)| - 2 \max\{d(u, b_u), d(v, b_v)\} \Big|^p \\
&\geq \tau^{-1} \sum_{i \in J} \eta_i^p \Big| |\phi_i(b_u) - \phi_i(b_v)| - \frac{2}{4} \frac{d(u, v)}{4\alpha(m)} \Big|^p \\
&\geq \tau^{-1} \sum_{i \in J} \eta_i^p \Big| |\phi_i(b_u) - \phi_i(b_v)| - \frac{1}{2} |\phi_i(b_u) - \phi_i(b_v)| \Big|^p \\
&\geq \tau^{-1} \left(\frac{d(u, v)}{8\alpha(m)} \right)^p .
\end{aligned}$$

Since we assumed that $\mathcal{F}_t(u, v)$ happened for at least $\tau/8$ indexes from T we have the lower bound

$$\begin{aligned}
\left\| \varphi(u) - \varphi(v) \right\|_p^p &\geq \sum_{t \in T} \left\| f^{(t)}(u) - f^{(t)}(v) \right\|_p^p \\
&\geq 1/8 \left(\frac{d(u, v)}{8\alpha(m)} \right)^p .
\end{aligned}$$

□

10.3 Low Degree k-HST and Embeddings of Ultrametrics

In this section we study partial embedding of ultrametrics into low degree HSTs and into l_p.

Claim 66. *Let $0 < \epsilon < 1$. Given a set $|X| = n$ and a partition of X into pair-wise disjoint sets $(X_1, \ldots, X_k)$ such that $|X_i| \leq \epsilon n$ for all $1 \leq i \leq k$ then*

$$\sum_{i=1}^{k} \binom{|X_i|}{2} \leq \epsilon \binom{n}{2}.$$

Proof.

$$\sum_{i=1}^{k} \binom{|X_i|}{2} = \sum_{i=1}^{k} \frac{|X_i|(|X_i| - 1)}{2} \leq \frac{\epsilon n - 1}{2} \sum_{i=1}^{k} |X_i| = \frac{\epsilon n - 1}{2} n = \epsilon \binom{n}{2}.$$

□

A k-HST is special type of ultrametric defined in [Bar96], which is an ultrametric T as defined in Definition 23, and has the additional requirement that if $u \in T$ is a descendant of v then $\Delta(u) \leq \Delta(v)/k$.

Lemma 67. *Any ultrametric has a coarse $(1-\epsilon)$-partial embedding into a 6-HST, such that the internal nodes' maximum degree is $O(1/\epsilon)$, with distortion $O(1)$.*

Proof. First we apply a lemma from [Bar96] and create a 6-HST by distorting any distance by no more than 6. Let r be the root, denote the *weight* of a node as the number of leaves in the tree below it. Let $b_1, \dots, b_m$ be all the children of r such that $weight(b_j) < \frac{\epsilon n}{2}$. Do the following process recursively: create a cluster C_i, while $weight(C_i) < \frac{\epsilon n}{2}$ insert any b_j into the cluster. when the cluster is big enough, start filling another until all b_j are clustered. We create sets $C_1, \dots, C_k$, that will replace $b_1, \dots, b_m$ as children of r. note that the weight of each C_i and each remaining child is at least $\frac{\epsilon n}{2}$ (except for maybe one), therefore we have at most $\frac{2}{\epsilon} + 1$ degree of internal node in the HST. Observe that distances between any clusters C_i, C_j are preserved, only distances inside clusters are discarded. By construction, the weight of each C_i is at most ϵn, therefore by Claim 66 there are less than $2\epsilon\binom{n}{2}$ such distances, and we have a 6-HST with the desired distortion. □

The next step is to apply the following lemma [BLMN05c]

Lemma 68. *For any $k > 5$, any k-HST can be $\left(\frac{k+1}{k-5}\right)$-embedded in l_p^h where $h = \lceil C(1 + k/p)^2 \log D \rceil$, where D is maximal out degree of a vertex in the tree defining the k-HST, and $C > 0$ is a universal constant.*

Corollary 69. *Any ultrametric has a $(1-\epsilon)$-partial embedding into l_p with $O(1)$ distortion and $O(\log(1/\epsilon))$ dimension.*

Proof. We first embed the ultrametric in a 6-HST of degree $O(1/\epsilon)$. Choosing $\hat{\epsilon} = \epsilon/4$ for this embedding then further embedding into l_p we discard at most $\epsilon\binom{n}{2}$ distances. □

11 Lower Bounds

11.1 Lower Bound on Dimension

The following theorem will show that the bound given in Theorem 5 is tight up to constant factors.[27]

Theorem 13. *For any $1 \le p < \infty$ and any $\theta > 0$, if the metric of an n-node constant degree expander embeds into l_p with distortion $O(\log^{1+\theta} n)$ then the dimension of the embedding is $\Omega(\log n / \lceil \log(\min\{p, \log n\}) + \theta \log\log n \rceil)$.*

Proof. Let $G = (V, E)$ be a 3-regular expander graph on n vertices and let (X, d) denote the shortest path metric on G. W.l.o.g let $\theta > 1/\log\log n$ and assume that $f : X \to l_p$ is a non-expansive embedding with distortion $C \log^{1+\theta} n$ for a constant C. Note that $\frac{1}{|E|}\sum_{(u,v)\in E} \|f(u) - f(v)\|_p^p \le 1$. Matoušek [Mat97] extended a theorem of [LLR94] and showed that there exists a number $c = O(\min\{p, \log n\})^p$ where the constant in the big O notation depends only on the expansion of G, such that $\frac{1}{\binom{n}{2}}\sum_{u \ne v} \|f(u) - f(v)\|_p^p \le c$.

Define a graph H on X where two vertices are connected iff $\|f(u) - f(v)\|_p^p \le 2c$. There must be a vertex u with degree at least $n/2$, as otherwise the average of all pairs will be larger than c. Denote the set of u and its neighbors in H by M.

[27] We thank an anonymous referee for providing us with the idea for this theorem.

We claim that there exists a subset $M' \subseteq M$ of cardinality at least $\sqrt{n}/2$ such that for any $x, y \in M'$ we have $d(x, y) \geq (1/2)\log_3 n$. To see this, greedily choose some point $x \in M$, add x to M', and remove all points $z \in M$ such that $d(x, z) < (1/2)\log_3 n$ (note that there are at most $\sqrt{n}$ such points). Continue while $M \neq \emptyset$. Since there are at least $n/2$ points in M we must have chosen at least $\sqrt{n}/2$ points before M was exhausted.

Note that for any $x, y \in M'$, it must be that $(\log^{-\theta} n)/(4C) < \|f(x) - f(y)\|_p$. This holds since $d(x, y) > (\log n)/4$, so it cannot be contracted by the embedding to less than $(\log^{-\theta} n)/(4C)$.

Now a volume argument suggests that having the points of M' in l_p space requires dimension at least $\Omega(\frac{\log n}{\theta \log\log n})$, by the following reasoning. Assume we embed into D dimensions, then for all $x \in M'$, by definition of M we have that $f(x) \in B_{l_p}(f(u), (2c)^{1/p})$, let $\alpha = (2c)^{1/p} = O(\min\{p, \log n\})$. The ball $B_{l_p}(f(u), \alpha)$ can be covered by $2^{O(D \cdot \log(8C\alpha/\log^{-\theta} n))}$ balls of radius $(\log^{-\theta} n)/(8C)$, each of the small balls contains no more than a single image of a point in M'. As $|M'| \geq \sqrt{n}/2$ it follows that $2^{O(D(\log\alpha + \theta\log\log n))} \geq \sqrt{n}/2$, or $D \geq \Omega(\frac{\log n}{\log\alpha + \theta\log\log n})$.

$\square$

For $1 \leq p \leq O(\log^\theta n)$ the dimension required is at least $\Omega\left(\frac{\log n}{\theta\log\log n}\right)$, which implies that the trade-off between distortion and dimension given in Theorem 5 is tight up to constant factors.

11.2 Lower Bound for Weighted Average distortion

In this section we show that the upper bound on weighted average distortion from Theorem 10 is tight up to a constant factor.

Theorem 14. *For any $p \geq 1$ and any large enough $n \in \mathbb{N}$ there exists a metric space (X, d) on n points, and non-degenerate probability distributions Π, Π' on $\binom{X}{2}$ with $\Phi(\Pi) = n$ and $\Phi(\Pi') = n^2$, such that any embedding f of X into l_p will have* $\mathrm{dist}_p^{(\Pi)}(f) \geq \Omega(\log(\Phi(\Pi))/p)$, *and* $\mathrm{distnorm}_p^{(\Pi')}(f) \geq \Omega(\log(\Phi(\Pi'))/p)$.

Proof. Let $G = (V, E)$ be a 3-regular expander graph on n vertices, *i.e.* the second eigenvalue λ of the Laplace matrix of G is bounded below by a constant independent of n, let (X, d) be the usual shortest path metric on G. Let $F = \binom{V}{2} \setminus E$. We define Π as Z/n on all pairs in E and Z/n^2 on all pairs in F, where $Z = \frac{n}{2(n-1)} \geq \frac{1}{2}$ is some normalizing factor. It follows that $\log(\Phi(\Pi)) = \log n$. It is an easy fact that at least $1/2$ of the distances in F are at least $\lfloor \log_3(n/2) \rfloor$, hence $\sum_{(u,v)\in F} d(u, v) \geq |F|(\log n)/4 \geq n^2(\log n)/16$ (for n large enough), and of course $\sum_{(u,v)\in E} d(u, v) = 3n/2$. By [Mat97] (which generalized the proof of [LLR94] for l_p), we know that if β is such that $\sum_{(u,v)\in E} \|f(u) - f(v)\|_p^p = \beta$, then $\frac{1}{n}\sum_{(u,v)\in F} \|f(u) - f(v)\|_p^p \leq O(\lambda\beta p^p)$. Note that since f is a non-contractive embedding we have that $\sum_{(u,v)\in F} \|f(u) - f(v)\|_p^p \geq \Omega(n^2 \log^p n)$ thus $\beta \geq \Omega((n \log^p n)/p^p)$.

$$
\begin{aligned}
\mathrm{dist}_q^{(\Pi)}(f)^p &= \sum_{u,v\in X} \Pi(u, v) \frac{\|f(u) - f(v)\|_p^p}{d(u, v)^p} \\
&= \sum_{(u,v)\in E} \frac{Z\|f(u) - f(v)\|_p^p}{n} + \sum_{(u,v)\in F} \frac{Z\|f(u) - f(v)\|_p^p}{n^2 \cdot d(u, v)^p} \\
&\geq \frac{Z\beta}{n} \geq \Omega((\log n)/p)^p.
\end{aligned}
$$

For the distortion of ℓ_p-norm we use the following distribution Π' which is Z' on edges and Z'/n^2 on $(u,v) \in F$, for some normalizing factor Z'. In this case $\log(\Phi(\Pi')) = 2\log n$. Then

$$\begin{aligned}
\text{distnorm}_p^{(\Pi')}(f)^p &= \frac{\sum_{u,v\in X} \Pi'(u,v)\|f(u)-f(v)\|_p^p}{\sum_{u,v\in X} \Pi'(u,v)d(u,v)^p} \\
&= \frac{\sum_{(u,v)\in E} \Pi'(u,v)\|f(u)-f(v)\|_p^p + \sum_{(u,v)\in F} \Pi'(u,v)\|f(u)-f(v)\|_p^p}{\sum_{(u,v)\in E} \Pi'(u,v) + \sum_{(u,v)\in F} \Pi'(u,v)d(u,v)^p} \\
&= \frac{\sum_{(u,v)\in E} \|f(u)-f(v)\|_p^p + (1/n^2)\sum_{(u,v)\in F} \|f(u)-f(v)\|_p^p}{\sum_{(u,v)\in E} 1 + (1/n^2)\sum_{(u,v)\in F} d(u,v)^p} \\
&\geq \frac{\sum_{(u,v)\in E} \|f(u)-f(v)\|_p^p}{2\sum_{(u,v)\in E} 1} \\
&\geq \frac{\beta}{6n} \geq \Omega((\log n)/p)^p.
\end{aligned}$$

In the third equality the normalizing factor Z' cancels out, and the first inequality: first note that there exist a constant $c = c(\lambda)$ such that for all $(u,v) \in E$, $d(u,v) \leq c\log n$, and if n is large enough so that $p \leq \log n/(\log c + \log\log n)$ then $(1/n^2)\sum_{(u,v)\in F} d(u,v)^p \leq (c\log n)^p \leq n \leq \sum_{(u,v)\in E} 1$. $\square$

11.3 Partial Embedding Lower Bounds

Recall the definition of metric composition Definition 9 and composition closure Definition 10. The theorem we prove shows a tight relation between lower bounds on the distortion and lower bounds for partial distortion.[28]

Theorem 15. *Let Y be a target metric space, let $\mathcal{X}$ be a family of metric spaces nearly closed under composition. If for any $k > 1$, there is $Z \in \mathcal{X}$ of size k such that any embedding of Z into Y has distortion at least $\alpha(k)$, then for all $n > 1$ and $\frac{1}{n} \leq \epsilon \leq 1$ there is a metric space $X \in \mathcal{X}$ on n points such that the distortion of any $(1-\epsilon)$ partial embedding of X into Y is at least $\alpha\left(\lceil \frac{1}{4\sqrt{\epsilon}} \rceil\right)/2$.*

Proof. Given ϵ, let Z be a metric space on $k = \lceil \frac{1}{4\sqrt{\epsilon}} \rceil$ points satisfying the assumptions of Theorem 15, choose $m = \lceil 4\sqrt{\epsilon}n \rceil$ for n large enough, so that m is strictly bigger than $2k$, let $\mathcal{C} = \{C_x\}_{x\in Z}$ where each $C_x \in \mathcal{X}$ with size m, and let $X = \mathcal{C}_\beta[Z]$ be its β-composition space for β satisfying that X can be embedded into some $\hat{X} \in \mathcal{X}$ with distortion 2.
Recall that a family of sets $\mathcal{F}$ is called *almost disjoint* if for any $A, B \in \mathcal{F}$ $\quad |A \bigcap B| \leq 1$.
Let $\mathcal{H} = \{(x_1, \ldots, x_k) : \forall i, x_i \in C_i\}$, we shall use the following basic lemma, similar arguments can be found in [BLMN05b].

Lemma 70. *For any integer k let $S_1, \ldots, S_k$ be disjoint sets of size m, where $m/2 > k$. Then there is a family $\mathcal{F}$ of representatives, i.e. a family of almost disjoint sets of size k containing a single element from each S_i, such that $|\mathcal{F}| \geq (m/2)^2$.*

Proof. Let p be a prime satisfying $m/2 < p \leq m$ Assume any p elements in each S_i are numbered $0, 1, 2 \ldots p-1$ (we ignore the others). denote x_{ij} the j-th element in the set S_i.
for each $a, b \in \mathbb{Z}_p$ let

$$A_{a,b} = \{x_{ij} : 1 \leq i \leq k, j = b + ai \pmod{p}\}.$$

[28] A similar theorem was independently given in [CDG+09].

$A_{a,b}$ is indeed a set of representatives - there is a unique $0 \le j \le p-1$ for each i satisfying the condition. Then take $\mathcal{F} = \{A_{a,b} : a, b \in \mathbb{Z}_p\}$, $|\mathcal{F}| = p^2$.
Assume by contradiction that for $A_{a,b} \neq A_{a',b'}$ we have $|A_{a,b} \bigcap A_{a',b'}| > 1$, then there must be $x_{ji}, x_{j'i'} \in A_{a,b} \bigcap A_{a',b'}$, then $j = b+ai \pmod p = b'+a'i \pmod p$ and $j' = b+ai' \pmod p = b'+a'i' \pmod p$. Now if $a = a'$ we have $b = b'$ (since p is prime), contradiction.
otherwise w.l.o.g assume $a' > a$

$$\begin{aligned}
b + ai &= b' + a'i \pmod p \\
b &= b' + (a'-a)i \pmod p \\
\big(b' + (a'-a)i\big) + ai' &= b' + a'i' \pmod p \\
(a'-a)i &= (a'-a)i' \pmod p
\end{aligned}$$

and since $a \neq a'$ we have $i = i'$ - contradiction. □

Consider a $(1-\epsilon)$ partial embedding of X in Y. By the lemma there is an almost disjoint family $\mathcal{F} \subseteq \mathcal{H}$ of size at least $(m/2)^2 > 2\epsilon n^2$, each pair $(u, v) \in X$ belongs to at most one set in $\mathcal{F}$. Since $\left|\binom{X}{2} \setminus G\right| < \epsilon n^2$, let $Z' \in \mathcal{F}$ be a set such that for all $u, v \in Z'$, $\quad (u, v) \in G$. up to scaling, Z' is isomorphic to Z, therefore the $(1-\epsilon)$ partial embedding of X into Y must incur distortion at least $\alpha(|Z|)$, and since X can be embedded into some $\hat{X} \in \mathcal{X}$ with distortion 2, $(1-\epsilon)$ partially embedding $\hat{X}$ into Y requires distortion at least $\alpha(|Z|)/2 = \alpha(\lceil \frac{1}{4\sqrt{\epsilon}} \rceil)/2$. □

Notice that if we are dealing with probabilistic embedding into a set of metric spaces $\mathcal{S}$, the claim hold for embedding into every $Y \in S$, and the theorem follows from our definition of *probabilistic* $(1-\epsilon)$ *partial* embedding.

The next Lemma gives an improved lower bound for coarse partial embeddings.

Lemma 71. *Let Y be a target metric space, let $\mathcal{X}$ be a family of metric spaces nearly closed under composition. If for any $k > 1$, there is $Z \in \mathcal{X}$ of size k such that any embedding of Z into Y has distortion at least $\alpha(k)$, then for all $n > 1$ and $\frac{1}{n} \le \epsilon \le 1$ there is a metric space $X \in \mathcal{X}$ on n points such that the distortion of any coarse $(1-\epsilon)$ partial embedding of X into Y is at least $\alpha\left(\lceil \frac{1}{2\epsilon} \rceil\right)/2$.*

The proof is immediate using the same method of metric composition. Let Z be a metric space on $k = \lceil \frac{1}{2\epsilon} \rceil$ points, and $m = \lceil 2\epsilon n \rceil$ be the composition sets' size. Then from the coarse property only distances inside each C_x can be discarded, so many isomorphic Z' have for all $u, v \in Z'$, $(u, v) \in G$.

Corollary 72. *For any $n > 1$ and $1/n < \epsilon < 1$:*

1. *There exists a metric space (X, d) on n points that requires $\Omega\left(\frac{\log(\frac{1}{\epsilon})}{p}\right)$ distortion for $(1-\epsilon)$ partial embedding into l_p.*
2. *There exists a metric space (X, d) on n points for which any $(1-\epsilon)$ partial embedding with distortion α into l_p requires dimension $\Omega(\log_\alpha \frac{1}{\epsilon})$.*
3. *There exists a metric space (X, d) on n points that requires $\Omega(\frac{1}{\sqrt{\epsilon}})$ distortion for $(1-\epsilon)$ partial embedding into trees.*

4. *There exists a metric space (X,d) on n points that requires $\Omega(\frac{1}{\epsilon})$ distortion for coarse $(1-\epsilon)$ partial embedding into trees.*

5. *There exists a metric space (X,d) on n points that requires $\Omega\left(\log\left(\frac{1}{\epsilon}\right)\right)$ distortion for any probabilistic $(1-\epsilon)$ partial embedding to trees.*

6. *There exists an n point subset of L_1 that requires $\Omega(\sqrt{\log(2/\epsilon)})$ distortion for $(1-\epsilon)$ partial embedding into L_2.*

7. *There exists a tree metric on n points that requires $\Omega(\sqrt{\log\log(2/\epsilon)})$ distortion for $(1-\epsilon)$ partial embedding into L_2.*

8. *There exists a metric space (X,d) on n points that requires $\Omega(\min\{q,\log n\}/p)$ q-norm of the distortion in an embedding into l_p.*

9. *There exists a metric space (X,d) on n points that requires $\Omega(\min\{q,\log n\})$ q-norm of expected distortion in any probabilistic embedding into trees.*

This follows from known lower bounds: (1) from [Mat97], (2) from equilateral dimension considerations, (3) and (4) from [RR98a], (5) from [Bar96], (6) from [Enf69] and with (7) also from the fact shown in [BLMN05c] that every normed space and trees are almost closed under composition, (7) also from [Bou86], (8) and (9) from Lemma 2.

Lemma 2. *Let Y be a target metric space, let $\mathcal{X}$ be a family of metric spaces. If for any $\epsilon \in (0,1)$, there is a lower bound of $\alpha(\epsilon)$ on the distortion of $(1-\epsilon)$ partial embedding of metric spaces in $\mathcal{X}$ into Y, then for any $1 \le q \le \infty$, there is a lower bound of $\frac{1}{2}\alpha(2^{-q})$ on the ℓ_q-distortion of embedding metric spaces in $\mathcal{X}$ into Y.*

Proof. For any $1 \le q \le \infty$ set $\epsilon = 2^{-q}$ and let $X \in \mathcal{X}$ be a metric space such that any $(1-\epsilon)$ partial embedding into Y has distortion at least $\alpha(\epsilon)$. Now, let f be an embedding of X into Y. It follows that there are at least $\epsilon\binom{n}{2}$ pairs $(u,v) \in \binom{X}{2}$ such that $\mathrm{dist}_f(u,v) \ge \alpha(\epsilon)$. Therefore:

$$\left(\mathbb{E}\left[\mathrm{dist}_f(u,v)^q\right]\right)^{1/q} \ge \left(\epsilon\alpha(\epsilon)^q\right)^{1/q} \ge \left(2^{-q}\alpha\left(2^{-q}\right)^q\right)^{1/q} = \frac{1}{2}\alpha\left(2^{-q}\right).$$

□

12 Applications

Consider an optimization problem defined with respect to weights $c(u,v)$ in a graph or in a metric space, where the solution involves minimizing the sum over distances weighted according to c: $\sum_{u,v} c(u,v)d(u,v)$. It is common for many optimization problem that such a term appears either in the objective function or alternatively it may come up in the linear programming relaxation of the problem.

These weights can be normalized to define the distribution Π where $\pi(u,v) = \frac{c(u,v)}{\sum_{x,y} c(x,y)}$ so that the goal translates into minimizing the *expected distance* according to the distribution Π. We can now use our results to construct embeddings with small *distortion of average* provided in Theorem 10, Theorem 20 and Theorem 21. Thus we get embeddings f into l_p and into ultrametrics with $\mathrm{distavg}^{(\Pi)}(f) = O(\log \hat{\Phi}(\Pi))$. In some of these applications it is crucial that the result holds for all such distributions Π (Theorems 10 and 20).

Define $\Phi(c) = \Phi(\Pi)$ and $\hat{\Phi}(c) = \hat{\Phi}(\Pi)$. Note that if for all $u \neq v$, $c(u,v) > 0$ then $\Phi(c) = \frac{\max_{u,v} c(u,v)}{\min_{u,v} c(u,v)}$. Using this paradigm we obtain $O(\log \hat{\Phi}(c)) = O(\min\{\log(\Phi(c)), \log n\})$ approximation algorithms.

This lemma below summarizes the specific propositions which will be useful in most of the applications in the sequel:

Lemma 73. *Let X be a metric space. For a weight function on the pairs $c : \binom{X}{2} \to \mathbb{R}_+$. Then:*

1. *There exists an embedding $f : X \to l_p$ such that for any weight function c:*

$$\sum_{\{u,v\}\in\binom{X}{2}} c(u,v)\|f(u) - f(v)\|_p \leq O(\log \hat{\Phi}(c)) \sum_{\{u,v\}\in\binom{X}{2}} c(u,v) d_X(u,v).$$

2. *There is a set of ultrametrics $\mathcal{S}$ and a probabilistic embedding $\hat{\mathcal{F}}$ of X into $\mathcal{S}$ such that for any weight function c:*

$$\mathbb{E}_{f\sim\hat{\mathcal{F}}}\left[\sum_{\{u,v\}\in\binom{X}{2}} c(u,v) d_Y(f(u), f(v))\right] \leq O(\log \hat{\Phi}(c)) \sum_{\{u,v\}\in\binom{X}{2}} c(u,v) d_X(u,v).$$

3. *For any given weight function c, there exists an ultrametric (Y, d_Y) and an embedding $f : X \to Y$ such that*

$$\sum_{\{u,v\}\in\binom{X}{2}} c(u,v) d_Y(f(u), f(v)) \leq O(\log \hat{\Phi}(c)) \sum_{\{u,v\}\in\binom{X}{2}} c(u,v) d_X(u,v).$$

Note that our result are particularly strong when the weights are uniform or close to uniform.

12.1 Sparsest cut

We show an approximation for the sparsest cut problem for complete weighted graphs, i.e., for the following problem:

Given a complete graph $G(V,E)$ on n vertices with capacities $c(u,v) : E \to \mathbb{R}_+$ and demands $D(u,v) : E \to \mathbb{R}_+$. Define the weight of a cut $(S, \overline{S})$ as

$$\frac{\sum_{u\in S, v\in\overline{S}} c(u,v)}{\sum_{u\in S, v\in\overline{S}} D(u,v)}.$$

We seek a subset $S \subseteq V$ minimizing the weight of the cut.

The uniform demand case of the problem was first given an approximation algorithm of $O(\log n)$ by Leighton and Rao [LR99]. For the general case $O(\log n)$ approximation algorithms were given by Aumann and Rabani [AR98] and London, Linial and Rabinovich [LLR94] via embeddings into l_1 of Bourgain. Recently Arora, Rao and Vazirani improved the uniform case bound to $O(\sqrt{\log n})$

and subsequently Arora, Lee and Naor gave an $O(\sqrt{\log n}\log\log n)$ approximation for the general demand case based on embedding of negative-type metrics into l_1.

We show an $O(\log\hat{\Phi}(c))$ approximation. We apply the method of [LLR94]: build the following linear program:

$$\begin{aligned} \min_{\tau} \sum_{u,v} c(u,v)\tau(u,v) \\ \text{subject to: } \sum_{u,v} D(u,v)\tau(u,v) \geq 1 \\ \text{for all } x,y,z: \ \tau(x,y) \leq \tau(x,z)+\tau(y,z) \\ \tau \geq 0 \end{aligned}$$

If the solution would yield a cut metric it would be the optimal solution. We solve the relaxed program for all metrics, obtaining a metric (V,τ), then embed (V,τ) into ℓ_1, using assertion (1.) of Lemma 73. Since the embedding f is non-contractive $\tau(u,v) \leq \|f(u)-f(v)\|_1$, hence

$$\frac{\sum_{u,v} c(u,v)\|f(u)-f(v)\|_1}{\sum_{u,v} D(u,v)\|f(u)-f(v)\|_1} \leq O(\log\hat{\Phi}(c))\frac{\sum_{u,v} c(u,v)\tau(u,v)}{\sum_{u,v} D(u,v)\tau(u,v)}.$$

Following [LLR94], we can obtain a cut that provides a $O(\log\hat{\Phi}(c))$ approximation.

12.2 Multicut

The multicut problem is: given a complete graph $G(V,E)$ with weights $c(u,v): E \to \mathbb{R}_+$, and k set of pairs $(s_i,t_i) \subseteq V \times V \quad i=1,\ldots,k$ find a minimal weight subset $E' \subseteq E$, such that removing every edge in E' disconnects every pair (s_i,t_i).

The best approximation algorithm for this problem is due to Garg, Vazirani and Yannakakis [GVY93] has performance $O(\log k)$.

We show a $O(\log\hat{\Phi}(c))$ approximation. We slightly change the methods of [GVY93], create a linear program:

$$\begin{aligned} \min_{\tau} \sum_{(u,v)\in\binom{V}{2}} c(u,v)\tau(u,v) \\ \text{subject to: } \forall i,j \sum_{(u,v)\in p_i^j} \tau(u,v) \geq 1 \\ \text{for all } x,y,z: \ \tau(x,y) \leq \tau(x,z)+\tau(y,z) \\ \tau \geq 0 \end{aligned}$$

where p_i^j is the j-th path from s_i to t_i. Now solve the relaxed version obtaining metric space (V,τ). Using (3.) of Lemma 73 we get an embedding $f: V \to Y$ into an HST (Y,d_Y) satisfying

$$\sum_{(u,v)\in\binom{V}{2}} c(u,v)d_Y(u,v) \leq O(\log\hat{\Phi}(c)) \sum_{(u,v)\in\binom{V}{2}} c(u,v)\tau(u,v) \ .$$

We use this metric to partition the graph instead of the region growing method introduced by [GVY93].

We build a multicut E': for every pair (s_i, t_i) find their $\text{lca}(s_i, t_i) = r_i$, and create two clusters containing all the vertices under each child: insert into E' all the edges between the points in each subtree and the rest of the graph. Since we have the constraint that $\sum_{(u,v)\in p_i^j} \tau(u,v) \geq 1$, we get from the fact that f is non-contractive that $\Delta(r_i) = d_Y(s_i, t_i) \geq 1$. It follows that if an edge $(u,v) \in E'$ then $d(u,v) \geq 1$. It follows that

$$\sum_{(u,v)\in E'} c(u,v) \leq \sum_{(u,v)\in \binom{V}{2}} c(u,v) d_Y(u,v) \leq O(\log \hat{\Phi}(c)) OPT.$$

12.3 Minimum Linear Arrangement

The same idea can be used in the minimum linear arrangement problem, where we have an undirected graph $G(V,E)$ with capacities $c(e)$ for every $e \in E$, we wish to find a one to one arrangement of vertices $h : V \to \{1, \ldots, |V|\}$, minimizing the total edge length: $\sum_{(u,v)\in E} c(u,v)|h(u) - h(v)|$.

This problem was first given an $O(\log n \log\log n)$ approximation by Even, Naor, Rao and Schieber [ENRS00], which was subsequently improved by Rao and Richa [RR98b] to $O(\log n)$.

As shown in [ENRS00], this can be done using the following LP:

$$\begin{aligned}
\min \quad & \sum_{u\neq v\in V} c(u,v) d(u,v) \\
s.t. \quad & \forall U \subseteq V, \quad \forall v \in U : \sum_{u\in U} d(u,v) \geq \frac{1}{4}(|U|^2 - 1) \\
& \forall (u,v) : d(u,v) \geq 0
\end{aligned}$$

which is proven there to be a lower bound to the optimal solution. Even et. al [ENRS00] use this LP formulation to define a *spreading metric* which they use to recursively solve the problem in a divide-and-conquer approach. Their method can be in fact viewed as an embedding into an ultrametric (HST) (the argument is similar to the one given for the special case of the *multicut* problem) and so by using assertion (3.) of Lemma 73 we obtain an $O(\log \hat{\Phi}(c))$ approximation.

The problem of embedding in d-dimensional meshes is basically an expansion of h to d dimensions, and can be solved in the same manner.

12.4 Multiple sequence alignment

Multiple sequence alignments are important tools in highlighting similar patterns in a set of genetic or molecular sequences.

Given n strings over a small character set, the goal is to insert gaps in each string as to minimize the total number of different characters between all pairs of strings, when the cost of gap is considered 0.

In their paper, [WLB+99] showed an approximation algorithm for the generalized version, where each pair of string has an importance parameter $c(u,v)$, they phrased the problem as finding a minimum *communication cost spanning tree*, i.e. finding a tree that minimizes $\sum_{u,v} c(u,v)d(u,v)$, where d is the edit distance. They apply probabilistic embedding into trees to bound the cost of such a tree. This gives an approximation ratio of $O(\log n)$.

Using assertion (3.)[29] of Lemma 73 we get an $O(\log \hat{\Phi}(c))$ approximation.

[29] We could use assertion (2.) here but since the parameter $c(u,v)$ is fixed assertion (3.) suffices.

12.5 Uncapacitated quadratic assignment

The uncapacitated quadratic assignment problem is one of the main studied problems in operations research (see the survey [PRW94]) and is once of the main applications of metric labeling [KT02]. Given three $n \times n$ input matrices C, D, F, such that C is symmetric with 0 in the diagonal, D is a metric and all matrices are non-negative. The objective is to minimize

$$\min_{\sigma \in \mathcal{S}_n} \sum_{i,j} C(i,j) D\left(\sigma(i), \sigma(j)\right) + \sum_{i} F\left(i, \sigma(i)\right)$$

where $\mathcal{S}_n$ is the set of all permutations over n elements.
One of the major applications of uncapacitated quadratic assignment is in location theory: where $C(i,j)$ is the material flow from facility i to j, $D\left(\sigma(i), \sigma(j)\right)$ is their distance after locating them and $F\left(i, \sigma(i)\right)$ is the cost for positioning facility i at location $\sigma(i)$.

Unlike the previous applications here C is not a fixed weight function on the metric D, but the actual weights depend on σ which is determined by the algorithm. Hence we require the probabilistic embedding given by assertion (2.) of Lemma 73 which is oblivious to the weight function C.
Kleinberg and Tardos [KT02] gave an approximation algorithm based on probabilistic embedding into ultrametrics. They give an $O(1)$ approximation algorithm for an ultrametric (they in fact use a 3-HST). This implies an $O(\log k)$ approximation for general metrics, where k is the number of labels.

As uncapacitated quadratic assignment is a special case of metric labeling it can be solved in the same manner, yielding a $O(\log \hat{\Phi}(C))$ approximation ratio by applying assertion (2.) of Lemma 73 together with the $O(1)$ approximation for ultrametrics of [KT02].

12.6 Min-sum k-clustering

Recall the min-sum k-clustering problem, where one has to partition a graph H to k clusters $C_1, \ldots, C_k$ as to minimize

$$\sum_{i=1}^{k} \sum_{u,v \in C_i} d_H(u,v).$$

[BCR01] showed a dynamic programming algorithm that gives a constant approximation factor for graphs that can be represented as HST. Then they used probabilistic embedding into a family of HST to give approximation with a factor of $O\left(\frac{1}{\epsilon}(\log n)^{1+\epsilon}\right)$ for general graphs H, with running time $n^{O(1/\epsilon)}$. Let $\Phi = \Phi(d)$.

Lemma 74. *For a graph H equipped with the shortest path metric, there is a $\log^{O(\log \Phi)} n$ time algorithm that gives $O\left(\log(k\Phi)\right)$ approximation for min-sum k-clustering problem.*

Proof. Denote by OPT the optimum solution for the problem with clusters C_i^{OPT}, and OPT_T the optimum solution for an HST T with clusters $C_i^{OPT_T}$. Also denote ALG for the result of [BCR01] algorithm with clusters $C_i^{ALG_T}$.

By Theorem 24 there exists a probabilistic $(1-\epsilon)$ partial embedding of H into a family of HST $\mathcal{T}$. Recall that G is the set of pairs distorted by at most $O(\log \frac{1}{\epsilon})$. Note that edges $e \in G$ are expanded by $O(\log \frac{1}{\epsilon})$ and for $e \notin G$ the maximum expansion is Φ (no distance is contracted), therefore choosing $\epsilon = \frac{1}{k^2\Phi}$ yields:

$$\begin{aligned}
\mathbb{E}[ALG] &= \sum_{T\in\mathcal{T}} \Pr[T] \sum_{i=1}^{k} \sum_{u,v\in C_i^{ALG_T}} d_H(u,v) \\
&\le \sum_{T\in\mathcal{T}} \Pr[T] \sum_{i=1}^{k} \sum_{u,v\in C_i^{ALG_T}} d_T(u,v) \\
&\le O(1) \sum_{T\in\mathcal{T}} \Pr[T] \sum_{i=1}^{k} \sum_{u,v\in C_i^{OPT_T}} d_T(u,v) \\
&\le O(1) \sum_{T\in\mathcal{T}} \Pr[T] \sum_{i=1}^{k} \sum_{u,v\in C_i^{OPT}} d_T(u,v) \\
&\le O(1) \left(\sum_{i=1}^{k} \sum_{u,v\in C_i^{OPT}\cap G} \sum_{T\in\mathcal{T}} \Pr[T] d_T(u,v) + \sum_{i=1}^{k} \sum_{u,v\in C_i^{OPT}\setminus G} \sum_{T\in\mathcal{T}} \Pr[T] d_T(u,v) \right) \\
&\le O(1) \left(\sum_{i=1}^{k} \sum_{u,v\in C_i^{OPT}\cap G} O\left(\log\left(1/\epsilon\right)\right) d_H(u,v) + \sum_{i=1}^{k} \sum_{u,v\in C_i^{OPT}\setminus G} \Phi \right) \\
&\le O((\log\left(1/\epsilon\right)) OPT + k\epsilon n^2 \Phi \\
&= O\left(\log(k\Phi)\right) OPT + n^2/k = O\left(\log(k\Phi)\right) OPT \ ,
\end{aligned}$$

the last equation follows from the fact that $\frac{n^2}{2k} \le OPT$ (assuming we scaled the distances such that $\min_{u\neq v\in H} d_H(u,v) \ge 1$), in what follows we show this fact. Let the clusters of the optimal solution be of sizes $a_1,\dots,a_k$, naturally $\sum_{i=1}^k a_i = n$, and there are at least $\sum_{i=1}^k a_i^2/2$ pairs of distance 1 inside clusters. Let $b = (1,1,\dots,1) \in \mathbb{R}^k$. From Cauchy-Schwartz we get

$$\left(\sum_{i=1}^{k} a_i\right)^2 = (\langle a,b\rangle)^2 \le \|a\|^2\|b\|^2 = \sum_i (a_i^2)k,$$

and therefore $\sum_i (a_i^2) \ge \frac{n^2}{k}$, meaning $OPT \ge \frac{n^2}{2k}$.

The running time of the algorithm is shown in [BCR01] to be $\log^L n$, where L is the maximal number of levels in the HST family $\mathcal{T}$. and this is at most $O(\log^{O(\log\Phi)} n + n^2)$ (which is $n^{O(1)}$ for $\Phi \le 2^{\frac{\log n}{\log\log n}}$), (see [BCR01] for details).

□

13 Distance Oracles

A distance oracle for a metric space (X,d), $|X| = n$ is a data structure that given any pair returns an estimate of their distance. In this section we study scaling distance oracles and partial distance oracles.

Let us begin by recalling the following consequence of Theorem 17:

Theorem 23. *Let (X, d) be a finite metric space. Let $k = O(\ln n)$ be a parameter. The metric space can be preprocessed in polynomial time, producing a data structure of size $O(n \cdot \lambda^{\frac{\log k}{k}} \log \lambda \log^2 k)$, such that distance queries can be answered in $O(\lambda^{\frac{\log k}{k}} \log \lambda \log^2 k)$ time, with worst case distortion $O(k)$.*

13.1 Distance oracles with scaling distortion

Given a distance oracle with $O(n^{1/k})$ bits, the worst case stretch can indeed be $2k-1$ for some pairs in some graphs. However we prove the existence of distance oracles with a *scaling stretch* property. For these distance oracles, the average stretch over all pairs is only $O(1)$.

We repeat the same preprocessing and distance query algorithm of Thorup and Zwick [TZ05] with sampling probability $3n^{-1/k} \ln n$ for the first set and $n^{-1/k}$ thereafter.

Given (X, d) and parameter k:
  $A_0 := X$; $A_k = \emptyset$;
  for $i = 1$ to $k-1$
    let A_i contain each element of A_{i-1},
    independently with probability $\begin{cases} 3n^{-1/k} \ln n & i = 1 \\ n^{-1/k} & i > 1 \end{cases}$;
  for every $x \in X$
    for $i = 0$ to $k-1$
      let $p_i(x)$ be the nearest node in A_i,
      so $d(x, A_i) = d(x, p_i(x))$;
      let $B_i(x) := \{y \in A_i \setminus A_{i+1} \mid d(x, y) < d(x, A_{i+1})\}$;

Figure 1: Preprocessing algorithm.

Given $x, y \in X$:
  $z := x$; $i := 0$;
  while $z \notin B_i(y)$
    $i := i + 1$;
    $(x, y) := (y, x)$;
    $z := p_i(x)$;
  return $d(x, z) + d(z, y)$;

Figure 2: Distance query algorithm.

Theorem 26. *Let (X, d) be a finite metric space. Let $k = O(\ln n)$ be a parameter. The metric space can be preprocessed in polynomial time, producing a data structure of $O(n^{1+1/k} \log n)$ size, such that distance queries can be answered in $O(k)$ time. The distance oracle has coarse scaling distortion bounded by $\left(2 \left\lceil \frac{\log(2/\epsilon)k}{\log n} \right\rceil + 1\right)$.*

Proof. For an integer $0 \le i < k$ let $\mathcal{E}_i(x)$ be the event that

$$B(x, r_{n^{(i-k)/k}}(x)) \cap A_i = \emptyset \ .$$

Note that as $|B(x, r_{n^{(i-k)/k}}(x))| \geq n^{i/k}$ it follows that $\Pr[\mathcal{E}_i(x)] \leq (1 - 3n^{-i/k}\ln n)^{n^{i/k}} \leq 1/n^3$, hence by the union bound there is high probability that none of the bad events $\mathcal{E}_i(x)$ happen for any $x \in X$ and $0 \leq i < k$, so from now on assume it is so.

Fix $\epsilon \in (0,1)$, and $x, y \in \hat{G}(\epsilon)$. Let j be the integer such that $n^{j/k} \leq \epsilon n/2 < n^{(j+1)/k}$. We prove by induction that at the end of the ℓth iteration of the while loop of the distance query algorithm:

1. $d(x,z) \leq d(x,y)\max\{1, \ell - j\}$
2. $d(z,y) \leq d(x,y)\max\{2, \ell - j + 1\}$.

First note that (1.) holds for any $\ell < j$ since we assume that $\mathcal{E}_\ell(x)$ did not happen, so $p_\ell(x) \in B(x, r_{n^{(\ell-k)/k}}(x))$, which suggests that $d(x, p_\ell(x)) \leq r_{n^{(\ell-k)/k}}(x) \leq r_{n^{j/k-1}}(x) \leq r_{\epsilon/2}(x) \leq d(x,y)$ and (2.) follows from (1.) and the triangle inequality. For $\ell \geq j$, from the induction hypothesis, at the beginning of the ℓth iteration, $d(z', y) \leq d(x,y)\max\{1, \ell - j\}$, where $z' = p_\ell(x)$, $z' \in A_\ell$. Since $z' \notin B_\ell(y)$ then after the swap (the line $(x,y) := (y,x)$) we have

$$d(x,z) = d(x, p_{\ell+1}(x)) \leq d(x,y)\max\{1, \ell - j\}$$

and $d(z,y) \leq d(x,y)\max\{2, \ell - j + 1\}$ follows from the triangle inequality. This competes the inductive argument. Since $p_{k-1}(x) \in A_{k-1} = B_{k-1}(y)$ then $\ell \leq k-1$ and therefore the stretch of the response is bounded by $2(k-j) - 1 \leq 2\left\lceil \frac{\log(2/\epsilon)k}{\log n} \right\rceil + 1$. $\square$

We note that a similar argument showing scaling stretch can be given for variation of Thorup and Zwick's compact routing scheme [TZ01].

13.2 Partial distance oracles

We construct a distance oracle with linear memory that guarantees stretch to $1-\epsilon$ fraction of the pairs. Recall the definition of $\hat{G}(\epsilon)$ given in Definition 6.

Theorem 27. *Let (X,d) be a finite metric space. Let $0 < \epsilon < 1$ be a parameter. Let $k \leq O(\log \frac{2}{\epsilon})$. The metric space can be preprocessed in polynomial time, producing data structure with either one of the following properties:*

1. *Either with $O\left(n\log(2/\epsilon) + k\left(\frac{\log(2/\epsilon)}{\epsilon}\right)^{1+1/k}\right)$ size, $O(k)$ query time and stretch $6k-1$ for some set $G \subseteq \binom{X}{2}$, $|G| \geq (1-\epsilon)\binom{n}{2}$.*
2. *Or, with $O\left(n\log n\log(2/\epsilon) + k\log n(1/\epsilon)^{1+1/k}\right)$ size, $O(k\log n)$ query time and stretch $6k-1$ for the set $\hat{G}(\epsilon)$.*

Proof. We begin with a proof of (1.). Let $b = \lceil (8/\epsilon)\ln(16/\epsilon)\rceil$. Let B be a set of b beacons chosen uniformly at random. Construct a distance oracle of [TZ05] on the subspace (B,d) with parameter $k \leq \log b$ yielding stretch $2k-1$ and using $O(kb^{1+1/k})$ storage. For every $x \in X$ we store $p(x)$, which is the closest node to x in B. The resulting data structure's size is $O(n\log b) + O(kb^{1+1/k}) = O(n\log b + kb^{1+1/k})$. Queries are processed as follows: given two nodes $x, y \in X$ let r be the response of the distance oracle on the beacons $p(x), p(y)$ then return $d(x,p(x)) + r + d(p(y),y)$.

Observe that from triangle inequality the response is at least $d(x,y)$. Let $\mathcal{E}_x$ for any $x \in X$ be the event

$$\mathcal{E}_x = \{d(x,B) > r_{\epsilon/8}(x)\}\ .$$

Then $\Pr[\mathcal{E}_x] \le (1-b/n)^{\epsilon n/8} \le \epsilon/16$ and so by Markov inequality, $\Pr[|\{\mathcal{E}_x \mid x \in X\}| \le \epsilon n/8] \ge 1/2$. In such a case let

$$G = \{(x,y) \in \binom{X}{2} \mid \neg\mathcal{E}_x \wedge \neg\mathcal{E}_y \wedge d(x,y) \ge \max\{r_{\epsilon/8}(x), r_{\epsilon/8}(y)\}\} \ .$$

We bound the size of G. For every point $x \in X$ at most $\epsilon n/8$ pairs (x,z) are removed due to $\mathcal{E}_z$ occurring and at most $\epsilon n/8$ pairs (x,z) are removed because $z \in B(x, r_{\epsilon/8}(x))$, so $|G| \ge (1-\epsilon/4)n^2 \ge (1-\epsilon)\binom{X}{2}$. For $(x,y) \in G$, we have $d(p(x),p(y)) \le d(p(x),x) + d(x,y) + d(p(y),y) \le d(x,y) + r_{\epsilon/8}(x) + r_{\epsilon/8}(y) \le 3d(x,y)$ so from the distance oracle $r \le (6k-3)d(x,y)$ and in addition $\max\{d(x,p(x)), d(y,p(y))\} \le d(x,y)$ so the stretch is bounded by $6k-1$.

The proof of (2.) is a slight modification of the above procedure. Let $m = \lceil 3\ln n \rceil$. Let $B_1, \dots, B_m$ be sets each containing $b = \lceil 16/\epsilon \rceil$ beacons, chosen independently and uniformly at random. Let DO_i be the distance oracle on (B_i, d). For every $x \in X$ we store $p_1(x), \dots, p_m(x)$ where $p_i(x)$ is the closest node in B_i. The resulting data structure's size is $O(n \log b \ln n) + O(kb^{1+1/k} \ln n) = O(n \log b \ln n + kb^{1+1/k} \ln n)$. Queries are processed as follows: given two nodes $x, y \in X$ let r_i be the response of the distance oracle DO_i on the beacons $p_i(x), p_i(y)$ then return $\min_{1 \le i \le m} d(x, p_i(x)) + r_i + d(p_i(y), y)$.

For every $(x,y) \in \binom{X}{2}, 1 \le i \le m$ define the event $\mathcal{E}^i_{x,y} = \{d(x,B_i) > r_{\epsilon/8}(x) \vee d(y,B_i) > r_{\epsilon/8}(y)\}$. Then $\Pr[\mathcal{E}^i_{x,y}] \le 2(1-b/n)^{\epsilon n/8} \le 1/e$, by independency $\Pr[\forall i, \mathcal{E}^i_{x,y}] \le 1/e^m \le 1/n^3$, and so by the union bound, $\Pr[\forall x,y \in X, \exists i \mid \neg\mathcal{E}^i_{x,y}] \ge 1/n$.

By a similar argument as in (1.) above, the stretch of $d(x,p_i(x)) + r_i + d(p_i(y),y)$ is at most $6k-1$.

□

References

[ABC+05] Ittai Abraham, Yair Bartal, T-H. Hubert Chan, Kedar Dhamdhere Dhamdhere, Anupam Gupta, Jon 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 Neimany. Advances in metric embedding theory. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, STOC '06, pages 271–286, New York, NY, USA, 2006. ACM.

[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.

[ABN08] 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.

[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.

[ABNS10] Ittai Abraham, Yair Bartal, Ofer Neiman, and Leonard J. Schulman. Volume in general metric spaces. In *Proceedings of the 18th annual European conference on Algorithms: Part II*, ESA'10, pages 87–99, Berlin, Heidelberg, 2010. Springer-Verlag.

[ALN05] Sanjeev Arora, James R. Lee, and Assaf Naor. Euclidean distortion and the sparsest cut. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, STOC '05, pages 553–562, New York, NY, USA, 2005. ACM.

[AR98] Yonatan Aumann and Yuval Rabani. An o(log k) approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301, 1998.

[ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, STOC '04, pages 222–231, New York, NY, USA, 2004. ACM.

[AS03] Vassilis Athitsos and Stan Sclaroff. Database indexing methods for 3d hand pose estimation. In *Gesture Workshop*, pages 288–299, 2003.

[Ass83] P. Assouad. Plongements lipschitziens dans $\mathbb{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.

[Bar96] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of the 37th Annual Symposium on Foundations of Computer Science*, pages 184–, Washington, DC, USA, 1996. IEEE Computer Society.

[Bar98] Yair Bartal. On approximating arbitrary metrices by tree metrics. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, STOC '98, pages 161–168, New York, NY, USA, 1998. ACM.

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

[BBM06] Yair Bartal, Béla Bollobás, and Manor Mendel. Ramsey-type theorems for metric spaces with applications to online problems. *J. Comput. Syst. Sci.*, 72:890–921, August 2006.

[BCR01] Yair Bartal, Moses Charikar, and Danny Raz. Approximating min-sum k-clustering in metric spaces. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, STOC '01, pages 11–20, New York, NY, USA, 2001. ACM.

[BFM86] J. Bourgain, T. Figiel, and V. Milman. On Hilbertian subsets of finite metric spaces. *Israel J. Math.*, 55(2):147–152, 1986.

[BHV08] Erik G. Boman, Bruce Hendrickson, and Stephen Vavasis. Solving elliptic finite element systems in near-linear time with support preconditioners. *SIAM J. Numer. Anal.*, 46:3264–3284, October 2008.

[BLMN05a] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric ramsey-type dichotomies. *J. London Math. Soc.*, 71(2):289–303, 2005.

[BLMN05b] Y. Bartal, N. Linial, M. Mendel, and A. Naor. Some low distortion metric ramsey problems. *Discrete Comput. Geom.*, 33(2):25–41, 2005.

[BLMN05c] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric ramsey-type phenomena. *Annals Math*, 162(2):643–709, 2005.

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

[Bou86] J. Bourgain. The metrical interpretation of superreflexivity in Banach spaces. *Israel J. Math.*, 56(2):222–230, 1986.

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

[CCG+98] Moses Charikar, Chandra Chekuri, Ashish Goel, Sudipto Guha, and Serge Plotkin. Approximating a finite metric by a small number of tree metrics. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, FOCS '98, pages 379–, Washington, DC, USA, 1998. IEEE Computer Society.

[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.

[CDG06] T.-H. Hubert Chan, Michael Dinitz, and Anupam Gupta. Spanners with slack. In *Proceedings of the 14th conference on Annual European Symposium - Volume 14*, pages 196–207, London, UK, 2006. Springer-Verlag.

[CDG+09] T.-H. Hubert Chan, Kedar Dhamdhere, Anupam Gupta, Jon Kleinberg, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. *SIAM Journal on Computing*, 38(6):2303–2329, 2009.

[CDK+04] Russ Cox, Frank Dabek, Frans Kaashoek, Jinyang Li, and Robert Morris. Practical, distributed network coordinates. *SIGCOMM Comput. Commun. Rev.*, 34:113–118, January 2004.

[CGT08] T-H. Hubert Chan, Anupam Gupta, and Kunal Talwar. Ultra-low-dimensional embeddings for doubling metrics. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '08, pages 333–342, Philadelphia, PA, USA, 2008. Society for Industrial and Applied Mathematics.

[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.

[Din07] Michael Dinitz. Compact routing with slack. In *Proceedings of the twenty-sixth annual ACM symposium on Principles of distributed computing*, PODC '07, pages 81–88, New York, NY, USA, 2007. ACM.

[DPK82] Narsingh Deo, G. Prabhu, and M. S. Krishnamoorthy. Algorithms for generating fundamental cycles in a graph. *ACM Trans. Math. Softw.*, 8:26–42, March 1982.

[ELR07] Michael Elkin, Christian Liebchen, and Romeo Rizzi. New length bounds for cycle bases. *Information Processesing Letters*, 104(5):186–193, 2007.

[Enf69] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8:103–105 (1969), 1969.

[ENRS00] Guy Even, Joseph Seffi Naor, Satish Rao, and Baruch Schieber. Divide-and-conquer approximation algorithms via spreading metrics. *J. ACM*, 47:585–616, July 2000.

[FHL05] Uriel Feige, MohammadTaghi Hajiaghayi, and James R. Lee. Improved approximation algorithms for minimum-weight vertex separators. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, STOC '05, pages 563–572, New York, NY, USA, 2005. ACM.

[FHRT03] Jittat Fakcharoenphol, Chris Harrelson, Satish Rao, and Kunal Talwar. An improved approximation algorithm for the 0-extension problem. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '03, pages 257–265, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics.

[FJJ+01] Paul Francis, Sugih Jamin, Cheng Jin, Yixin Jin, Danny Raz, Yuval Shavitt, and Lixia Zhang. Idmaps: a global internet host distance estimation service. *IEEE/ACM Trans. Netw.*, 9:525–540, October 2001.

[FLM77] T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1-2):53–94, 1977.

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

[FT03] Jittat Fakcharoenphol and Kunal Talwar. An improved decomposition theorem for graphs excluding a fixed minor. In *In Proceedings of 6th Workshop on Approximation, Randomization, and Combinatorial Optimization*, pages 36–46. Springer, 2003.

[GK11] L. Gottlieb and R. Krauthgamer. A nonlinear approach to dimension reduction. In *SODA '11: Proceedings of the 24th annual ACM-SIAM Symposium on Discrete Algorithms*, 2011.

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

[Gro83] Mikhael Gromov. Filling Riemannian manifolds. *J. Differential Geom.*, 18(1):1–147, 1983.

[GVY93] Naveen Garg, Vijay V. Vazirani, and Mihalis Yannakakis. Approximate max-flow min-(multi)cut theorems and their applications. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, STOC '93, pages 698–707, New York, NY, USA, 1993. ACM.

[HBK+03] Eran Halperin, Jeremy Buhler, Richard M. Karp, Robert Krauthgamer, and B. Westover. Detecting protein sequence conservation via metric embeddings. In *In 11th Conference on Intelligent Systems for Molecular Biology*, pages 122–129, 2003.

[HFC00] Gabriela Hristescu and Martin Farach-Colton. Cofe: A scalable method for feature extraction from complex objects. In *Proceedings of the Second International Conference on Data Warehousing and Knowledge Discovery*, DaWaK 2000, pages 358–371, London, UK, 2000. Springer-Verlag.

[HS03] Gísli R. Hjaltason and Hanan Samet. Properties of embedding methods for similarity searching in metric spaces. volume 25, pages 530–549, Washington, DC, USA, May 2003. IEEE Computer Society.

[IM04] Piotr Indyk and Jiri Matousek. Low-distortion embeddings of finite metric spaces. In *in Handbook of Discrete and Computational Geometry*, pages 177–196. CRC Press, 2004.

[Ind01] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *Proceedings of the 42nd IEEE symposium on Foundations of Computer Science*, FOCS '01, pages 10–, Washington, DC, USA, 2001. IEEE Computer Society.

[JL84] W. B. Johnson and J. 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.

[JS82] W.B. Johnson and G. Schechtman. Embedding l_p^m into l_1^n. *Acta Mathematica*, 149(1):71–85, 1982.

[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 *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 434–443, Washington, DC, USA, 2004. IEEE Computer Society.

[KMP] Ioannis Koutis, Gary L. Miller, and Richard Peng. Approaching optimality for solving sdd linear systems. In *51th Annual IEEE Symposium on Foundations of Computer Science, October 23-26, 2010*, pages 235–244, Las Vegas, Nevada, USA.

[KMST10] Alexandra Kolla, Yury Makarychev, Amin Saberi, and Shang-Hua Teng. Subgraph sparsification and nearly optimal ultrasparsifiers. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 57–66, New York, NY, USA, 2010. ACM.

[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.

[Kru64] J.B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[KRXY07] Goran Konjevod, Andrea Werneck Richa, Donglin Xia, and Hai Yu. Compact routing with slack in low doubling dimension. In *Proceedings of the twenty-sixth annual ACM symposium on Principles of distributed computing*, PODC '07, pages 71–80, New York, NY, USA, 2007. ACM.

[KSW09] Jon Kleinberg, Aleksandrs Slivkins, and Tom Wexler. Triangulation and embedding using small sets of beacons. *J. ACM*, 56(6):1–37, 2009.

[KT02] Jon Kleinberg and Éva Tardos. Approximation algorithms for classification problems with pairwise relationships: metric labeling and markov random fields. *J. ACM*, 49:616–639, September 2002.

[KW78] Joseph B. Kruskal and Myron Wish. *Multidimensional Scaling*. M. Sage Publications, CA, 1978.

[LHC05] Hyuk Lim, Jennifer C. Hou, and Chong-Ho Choi. Constructing internet coordinate system based on delay measurement. volume 13, pages 513–525, Piscataway, NJ, USA, June 2005. IEEE Press.

[Lin02] Nathan Linial. Finite metric spaces: combinatorics, geometry and algorithms. In *Proceedings of the eighteenth annual symposium on Computational geometry*, SCG '02, pages 63–63, New York, NY, USA, 2002. ACM.

[LLR94] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. pages 577–591, 1994.

[LMN05] James R. Lee, Manor Mendel, and Assaf Naor. Metric structures in l1: dimension, snowflakes, and average distortion. volume 26, pages 1180–1190, London, UK, UK, November 2005. Academic Press Ltd.

[LN05] J. R. Lee and A. Naor. Extending lipschitz functions via random metric partitions. *Inventiones Mathematicae*, 160(1):59–95, 2005.

[LR99] Tom Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM*, 46:787–832, November 1999.

[LS91] Nathan Linial and Michael Saks. Decomposing graphs into regions of small diameter. In *Proceedings of the second annual ACM-SIAM symposium on Discrete algorithms*, SODA '91, pages 320–330, Philadelphia, PA, USA, 1991. Society for Industrial and Applied Mathematics.

[LS08] Emmanuelle Lebhar and Nicolas Schabanel. Graph augmentation via metric embedding. In *OPODIS*, pages 217–225, 2008.

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

[Mat96] J. Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel J. Math.*, 93:333–344, 1996.

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

[Mat99] J. Matoušek. On embedding trees into uniformly convex Banach spaces. *Israel J. Math.*, 114:221–237, 1999.

[Mat02] Jiri Matousek. *Lectures on Discrete Geometry.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002.

[Mat05] Jiri Matoušek. Open problems on low-distortion embeddings of finite metric spaces, 2005. Avaliable in http://kam.mff.cuni.cz/ matousek/metrop.ps.gz.

[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.

[NZ02] T. S. Eugene Ng and Hui Zhang. Predicting internet network distance with coordinates-based approaches. In *21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, pages 178–187, 2002.

[PRW94] P. Pardalos, F. Rendl, and H. Wolkowicz. The quadratic assignment problem: a survey and recent developments. In P. Pardalos and H. Wolkowicz, editors, *Quadratic assignment and related problems (New Brunswick, NJ, 1993)*, pages 1–42. Amer. Math. Soc., Providence, RI, 1994.

[Rab03] Yuri Rabinovich. On average distortion of embedding metrics into the line and into l1. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, STOC '03, pages 456–462, New York, NY, USA, 2003. ACM.

[Rao99] Satish Rao. Small distortion and volume preserving embeddings for planar and euclidean metrics. In *Proceedings of the fifteenth annual symposium on Computational geometry*, SCG '99, pages 300–306, New York, NY, USA, 1999. ACM.

[RR98a] Yuri Rabinovich and Ran Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *Discrete & Computational Geometry*, 19(1):79–94, 1998.

[RR98b] Satish Rao and Andréa W. Richa. New approximation techniques for some ordering problems. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, SODA '98, pages 211–218, Philadelphia, PA, USA, 1998. Society for Industrial and Applied Mathematics.

[Sem96] S. Semmes. On the nonexistence of bilipschitz parameterizations and geometric problems about a_∞ weights. *Revista Matemática Iberoamericana*, 12:337–410, 1996.

[ST04] Yuval Shavitt and Tomer Tankel. Big-bang simulation for embedding network distances in euclidean space. *IEEE/ACM Trans. Netw.*, 12:993–1006, December 2004.

[TC04] Liying Tang and Mark Crovella. Geometric exploration of the landmark selection problem. In *Passive and Active Network Measurement, 5th International Workshop, PAM 2004, Antibes*, pages 63–72, 2004.

[TZ01] Mikkel Thorup and Uri Zwick. Compact routing schemes. In *Proceedings of the thirteenth annual ACM symposium on Parallel algorithms and architectures*, SPAA '01, pages 1–10, New York, NY, USA, 2001. ACM.

[TZ05] Mikkel Thorup and Uri Zwick. Approximate distance oracles. *J. ACM*, 52(1):1–24, 2005.

[WLB+99] Bang Ye Wu, Giuseppe Lancia, Vineet Bafna, Kun-Mao Chao, R. Ravi, and Chuan Yi Tang. A polynomial-time approximation scheme for minimum routing cost spanning trees. volume 29, pages 761–778, Philadelphia, PA, USA, December 1999. Society for Industrial and Applied Mathematics.