Local Global Tradeoffs in Metric Embeddings*

Moses Charikar† Konstantin Makarychev‡ Yury Makarychev§

Preprint

Abstract

Suppose that every k points in a n point metric space X are D-distortion embeddable into ℓ_1. We give upper and lower bounds on the distortion required to embed the entire space X into ℓ_1. This is a natural mathematical question and is also motivated by the study of relaxations obtained by lift-and-project methods for graph partitioning problems. In this setting, we show that X can be embedded into ℓ_1 with distortion $O(D \times \log(n/k))$. Moreover, we give a lower bound showing that this result is tight if D is bounded away from 1. For $D = 1+\delta$ we give a lower bound of $\Omega(\log(n/k)/\log(1/\delta))$; and for $D = 1$, we give a lower bound of $\Omega(\log n/(\log k + \log\log n))$. Our bounds significantly improve on the results of Arora, Lovász, Newman, Rabani, Rabinovich and Vempala, who initiated a study of these questions.

1 Introduction

In this paper we study the following question raised by Arora, Lovász, Newman, Rabani, Rabinovich and Vempala [5]:

> Suppose that every k points in a metric space X are D-distortion embeddable into ℓ_1. What is the least distortion with which we can embed the entire space X into ℓ_1?

In other words, what do *local* properties (embeddability of subsets) of the space tell us about *global* properties (embeddability of the entire space)? This is a natural question about

*This work was done while K. Makarychev and Y. Makarychev were graduate students at Princeton University. The authors were supported by NSF ITR grant CCR-0205594, NSF CAREER award CCR-0237113, MSPA-MCS award 0528414, and an Alfred P. Sloan Fellowship. K. Makarychev was supported by a Gordon Wu fellowship and an IBM Ph.D. Fellowship. Y. Makarychev was supported by a Gordon Wu fellowship.

†Department of Computer Science, Princeton University, Princeton, NJ 08540. Email: moses@cs.princeton.edu.

‡IBM T.J. Watson Research Center, Yorktown Heights, NY 10598. Email: konstantin@us.ibm.com.

§Microsoft Research, One Memorial Drive, Cambridge, MA 02142. Email: yurym@microsoft.com.

metric spaces and our research is motivated by numerous applications of low distortion metric embeddings in computer science and mathematics[1].

The study of embeddings of metric spaces into normed spaces has played an important role in the development of approximation algorithms. In particular, finite metrics arise naturally in mathematical programming relaxations of graph partitioning problems and the distortion required for embedding into ℓ_1 is directly related to the approximation factor achievable using this approach. The challenge here is to find computationally tractable metrics that admit low distortion embeddings into ℓ_1. The classical theorem due to Bourgain [8] states that any n point metric is embeddable into ℓ_1 with distortion $O(\log n)$. Linial, London and Rabinovich [16] and independently Aumann and Rabani [7] showed that this result is tight and exploited it to design an $O(\log n)$ approximation for sparsest cut.

Considering tighter relaxations is one potential route to improved results. Indeed the study of SDP relaxations and the so called ℓ_2^2 metrics that arise from them have given rise to better approximations for sparsest cut [6, 4]. One avenue for further improvement is the application of lift-and-project methods (such as those given by Lovász–Schrijver [18] and Sherali–Adams [25]) which are systematic ways to design a sequence of increasingly tighter relaxations. The metrics that arise from k rounds of lift-and-project satisfy the property that every subset of size k is isometrically embeddable into ℓ_1. This naturally leads to the question of how such local embeddability affects global embeddability of the metric.

Independent of the motivation from combinatorial optimization, the question of the relationship between local and global properties of metric spaces is fairly natural. It has been studied before in the context of other properties of metric spaces. Menger's theorem [23] states that the embeddability of a metric into ℓ_2^n is characterized by embeddability of all subsets of size $n+3$ into ℓ_2^n. Similarly, it is known that metric is a tree metric if and only if every subset of size four is a tree metric. Tradeoffs between local and global distortions for tree metrics were recently studied by Abraham, Balakrishnan, Kuhn, Malkhi, Ramasubramanian, and Talwar [1].

Analogous local-global questions have been studied in other realms, including analysis, combinatorics, geometry, topology, and mathematical logic. One such example is Helly's theorem [15] on intersections of bounded convex sets. The influential theory of graph minors studies global graph properties that arise from the local property of subgraphs excluding a given set of minors. Other examples are numerous compactness theorems that state that if some property holds for every finite subset of elements, then the property holds for the entire set (e.g. if every finite subtheory is consistent, then the entire theory is consistent). In functional analysis, many properties of a Banach space are deduced from properties of its finite dimensional subspaces. In particular, one important parameter is the largest Banach–Mazur distance between a k dimensional subspace of a Banach space and the k dimensional Euclidean space (see e.g. [27, Section 6]). In computer science, such local-global considerations arise in property testing and in the study of PCPs.

[1]A good introduction to the area of low distortion metric embeddings and their applications in computer science is Matoušek's book "Lectures on Discrete Geometry" [21, Section 15].

1.1 Our results

We show that if every k point subset of n point metric space X can be embedded into ℓ_1 with distortion D, then X can be embedded into ℓ_1 with distortion[2] $O(D \times \log(n/k))$. Moreover, we give a lower bound showing that this result is tight if D is bounded away from 1. For $D = 1 + \delta$ we give a lower bound of $\Omega(\log(n/k)/\log(1/\delta))$; and for $D = 1$, we give a lower bound of $\Omega(\log n/(\log k + \log\log n))$. We summarize our results and compare them with the results obtained by Arora, Lovász, Newman, Rabani, Rabinovich and Vempala [5] in the table below.

Upper Bounds

	This paper	**Arora et al [5]**
D	$O(D\log(n/k))$	$O\left(D\,(n/k)^2\right)$

Lower Bounds

1	$\Omega\left(\frac{\log n}{\log k + \log\log n}\right)$	$(\log n)^{\Omega(1/k)}$
$1+\delta$	$\Omega\left(\frac{\log(n/k)}{\log(1/\delta)}\right)$	
$D \geq 3/2$	$\Omega(D\log(n/k))$	$\Omega\left(D\frac{\log^2(n/k)}{\log n}\right)$ for $D \sim C\frac{\log^2 n}{\log^2(n/k)}$

Figure 1: The local distortion (the distortion with which every k points are embeddable into ℓ_1) is given in the first column. In the last row we assume that $D \leq \log n/\log(n/k)$.

Our results significantly improve the results obtained in [5]. We completely solve the problem for every D bounded away from 1. We also answer the main open question posed by Arora, Lovász, Newman, Rabani, Rabinovich, and Vempala [5]: we construct a metric space that requires large distortion to embed into ℓ_1, such that every subset of size $n^{o(1)}$ embeds *isometrically* into ℓ_1. There is still a gap between our lower and upper bounds for metric spaces locally embeddable into ℓ_1 isometrically. Closing this gap is an interesting open question.

We also show that even if a small fraction α (say 1%) of all subsets of size k embeds into ℓ_p with distortion at most D, then we still can embed the entire space into ℓ_p with distortion at most

$$D \cdot O(\log(n/k) + \log\log(1/\alpha) + \log p).$$

In subsequent work [12], we show that our results imply strong lower bounds for Sherali–Adams relaxations of the Sparsest Cut, MAX CUT, Vertex Cover, Maximum Acyclic Subgraph and other problems. Namely, we prove that the integrality gap for the Sparsest Cut

[2]We also generalize this result to all ℓ_p spaces.

problem is at least

$$\Omega\left(\max\left(\sqrt{\frac{\log n}{\log r + \log\log n}}, \frac{\log n}{r + \log\log n}\right)\right)$$

after r rounds of the Sherali–Adams lift-and-project. The integrality gap for MAX CUT, Vertex Cover and Maximum Acyclic Subgraph remains $2 - \varepsilon$ even after n^γ rounds (for every positive ε and some γ that depends on ε). We also show how to obtain Sherali–Adams gap examples for any problem that has a Unique Games based hardness result (with some additional conditions on the reduction from Unique Games).

1.2 Overview

Before we describe the technical details of our results, we give a brief (and imprecise) overview of our techniques.

Our upper bound is based on combining two embeddings, one that handles large distances and the other that handles small distances. In order to construct the first embedding, we construct a hitting set S of size k such that the local neighborhood of every point contains a point in S. The local embeddability property guarantees that S is embeddable into ℓ_1 with small distortion. We extend this embedding to an embedding of the entire set. In order to do this, we construct a random clustering of the points in the space such that nearby points are likely to fall in the same cluster; each cluster is then mapped to a point in the hitting set S. We show that this embedding does not stretch any pair of points by a large amount and is a good embedding for pairs of points at large distances. The second embedding is obtained by taking the first $O(\log(n/k))$ densities in Bourgain's embedding. Again, this does not stretch any pair of points by a large amount. On the other hand, it is a good embedding for small distances.

Our lower bounds are based on constant degree expander graphs used in previous work [3]. Instead of the commonly used shortest path metric however, we define a different metric particularly convenient for the purpose of constructing embeddings into ℓ_1. Our choice of this metric was inspired by the papers of de la Vega and Kenyon-Mathieu [13] and Schoenebeck, Trevisan, and Tulsiani [26], who used a similar metric (distribution of cuts) in their integrality gap examples. We exploit the fact that subgraphs of constant degree expander graphs are sparse and show that such sparse graphs equipped with the new metric embed well into ℓ_1. All short distances are preserved exactly (which is convenient in our iterative construction) and long distances are distorted by a small factor. By appropriately choosing parameters, we are able to construct embeddings with distortion at most $1 + \delta$. However the metric on the entire expander still requires high distortion for embedding into ℓ_1. In order to obtain examples with high global distortion where the metric on subsets is isometrically embeddable into ℓ_1, we show that obtaining distortion sufficiently close to 1 is enough. In this case, we guarantee that a slightly different metric (obtained by adding a small constant to all distances) is in fact isometrically embeddable into ℓ_1. This changes all distances by a factor of at most 2 and still has high global distortion.

2 Embedding Theorem

In this section, we prove the following theorem.

Theorem 2.1. *Let k be a positive integer and $p \geq 1$. Suppose that every subset of size k of a finite metric space (X, d) is embeddable into ℓ_p with distortion D. Then the metric space (X, d) is embeddable into ℓ_p with distortion $O(D \cdot \log(|X|/k))$.*

For simplicity of presentation, we assume throughout the proof that all distances in our metric space are distinct. This is a standard assumption and we may make it without loss of generality. We denote the ball of radius R about x by $\mathrm{B}(x, R) = \{y : d(x, y) \leq R\}$. Finally, we define the local radius for every point as follows.

Definition 2.2. *For every point x of the metric space X, define radius $R_{x,m}$ to be the minimum radius R for which the ball $\mathrm{B}(x, R)$ contains m points:*

$$R_{x,m} = \min(R : |\mathrm{B}(x, R)| = m).$$

Note that for a point x, the local neighborhood (mentioned in the overview) is the ball $\mathrm{B}(x, R_{x,m})$.

2.1 Hitting Set

In this section we describe a greedy algorithm for finding a set S of size at most k that intersects with every ball $\mathrm{B}(x, 2R_{x,m})$. The following lemma first appeared in the paper of Chan, Dinitz and Gupta [11].

Lemma 2.3. *For every finite metric space (X, d) and every positive integer m there exists a subset $S \subset X$ of size at most $\lfloor |X|/m \rfloor$ such that for every point x in X the ball about x of radius $2R_{x,m}$ contains at least one point from S. In other words, for every x in X*

$$\mathrm{B}(x, 2R_{x,m}) \cap S \neq \varnothing.$$

Moreover, for every x and y in S the balls $\mathrm{B}(x, R_{x,m})$ and $\mathrm{B}(y, R_{y,m})$ do not intersect.

We need the following simple observation.

Lemma 2.4. *For every finite metric space (X, d), every positive integer m and every two points x and y in X, the following inequality holds:*

$$|R_{x,m} - R_{y,m}| \leq d(x, y).$$

Proof. Notice, that the ball $\mathrm{B}(x, R_{y,m} + d(x, y))$ contains the ball $\mathrm{B}(y, R_{y,m})$. Therefore, $|\mathrm{B}(x, R_{y,m} + d(x, y))| \geq |\mathrm{B}(y, R_{y,m})| = m$ and $R_{x,m} \leq R_{y,m} + d(x, y)$. Similarly, $R_{y,m} \leq R_{x,m} + d(x, y)$. □

Proof of Lemma 2.3. We give an explicit (deterministic) algorithm for finding the set S. The algorithm maintains three sets: a set of "active" points A, a set of "unsatisfied" balls $\mathcal{B}$, and a set of selected points S. Initially the set A contains all points of the metric space X, the set $\mathcal{B}$ contains all balls $\mathrm{B}(x, R_{x,m})$:

$$\mathcal{B} = \{\mathrm{B}(x, R_{x,m}) : x \in X\};$$

and the set S is empty. At each iteration we pick the ball $\mathrm{B}(x, R_{x,m})$ of smallest radius from $\mathcal{B}$ and add the center of this ball, the point x, to S. Then we remove all points of the ball $\mathrm{B}(x, R_{x,m})$ from the set of active points A:

$$A = A \setminus \mathrm{B}(x, R_{x,m});$$

we also remove all balls that intersect with $\mathrm{B}(x, R_{x,m})$ from $\mathcal{B}$:

$$\mathcal{B} = \{\mathrm{B}(y, R_{y,m}) \in \mathcal{B} : \mathrm{B}(y, R_{y,m}) \cap \mathrm{B}(x, R_{x,m}) = \varnothing\}.$$

When the set $\mathcal{B}$ becomes empty, the algorithm stops and returns the set S.

Let us analyze the algorithm. Observe that after every iteration all balls in $\mathcal{B}$ contain points only from the set A. Hence at every step we remove exactly m points from A (recall that every ball $\mathrm{B}(x, R_{x,m})$ contains exactly m points from X by the definition of $R_{x,m}$). Therefore, after $\lfloor |X|/m \rfloor$ iterations A will contain less than m elements and thus the set $\mathcal{B}$ will be empty. Hence the set S contains at most $\lfloor |X|/m \rfloor$ points.

We now need to check that S intersects with every ball $\mathrm{B}(y, 2R_{y,m})$. Consider an arbitrary point y and the step at which $\mathrm{B}(y, R_{y,m})$ was removed from $\mathcal{B}$. Let x be the point that was added to the set S at this step. Since $\mathrm{B}(y, R_{y,m})$ was removed from $\mathcal{B}$, the ball $\mathrm{B}(y, R_{y,m})$ intersects with the ball $\mathrm{B}(x, R_{x,m})$. Hence

$$d(x, y) \leq R_{x,m} + R_{y,m}.$$

Notice that $R_{x,m} \leq R_{y,m}$, since at every step we choose the ball of smallest radius. Thus x lies in the ball of radius $2R_{y,m}$ about y. This concludes the proof. □

Corollary 2.5. *For every finite metric space (X, d) and every positive integer m there exists a subset $S \subset X$ of size at most $\lfloor |X|/m \rfloor$ and a mapping $g : X \to S$ such that for every point x in X,*

$$d(x, g(x)) \leq 2R_{x,m}.$$

2.2 Partitioning

In this section, we describe a randomized mapping of the metric space X into itself that "glues" together points at small distances with high probability. Our algorithm is based on the clustering technique of Calinescu, Karloff and Rabani [10] and Fakcharoenphol, Rao, and Talwar [14].

Lemma 2.6. *For every finite metric space (X, d) and every positive integer m there exists a random mapping $f\colon X \to X$ such that for every x and y in X,*

1. $d(x, f(x)) \leq R_{x,m}$ *(always);*
2. $\Pr\left(f(x) \neq f(y)\right) \leq O(\log m) \times \frac{d(x,y)}{R_{x,m}+R_{y,m}}$.

Proof. We present a probabilistic algorithm that finds the mapping f.

---

1. Pick a random number α uniformly distributed in $(0, 1)$.
2. Pick a random (uniform) total order $<_\pi$ on the elements of the space X.
3. Now define the mapping $f\colon X \to X$ as follows: For every point x, let $f(x)$ be the minimal point z in the ball $\mathrm{B}(x, \alpha \cdot R_{x,m})$ with respect to the order $<_\pi$.
4. Return the mapping f.

Clearly, the mapping returned by the algorithm always satisfies the first property:

$$d(x, f(x)) \leq \alpha \cdot R_{x,m} \leq R_{x,m}.$$

To verify the second property, consider two points x and y in X. We will show that

$$\Pr\Big(f(x) <_\pi f(y)\Big) \leq (2\log m + O(1)) \times \frac{d(x,y)}{R_{y,m}}.$$

Enumerate all points in the ball $\mathrm{B}(x, R_{x,m})$ in the order of increasing distance from the point x: $z_1, \dots, z_m$. Write

$$\Pr\left(f(x) <_\pi f(y)\right) = \sum_{i=1}^{m} \Pr\left(f(x) = z_i \text{ and } f(y) >_\pi z_i\right). \tag{1}$$

Observe that if $f(y) >_\pi z_i$, then z_i does not belong to the ball $\mathrm{B}(y, \alpha \cdot R_{y,m})$. Therefore,

$$\begin{aligned}(f(x) = z_i \text{ and } f(y) >_\pi z_i) &\leq \Pr\left(f(x) = z_i \mid z_i \in \mathrm{B}(x, \alpha R_{x,m}) \text{ and } z_i \notin \mathrm{B}(y, \alpha R_{y,m})\right) \\ &\quad \times \Pr\left(z_i \in \mathrm{B}(x, \alpha R_{x,m}) \text{ and } z_i \notin \mathrm{B}(y, \alpha R_{y,m})\right).\end{aligned}$$

Let us estimate the probabilities in the right hand side. We have

$$\begin{aligned}\Pr(z_i \in \mathrm{B}(x, \alpha R_{x,m}) \text{ and } z_i \notin \mathrm{B}(y, \alpha R_{y,m})) &= \Pr\left(d(x, z_i) \leq \alpha R_{x,m} \text{ and } d(y, z_i) > \alpha R_{y,m}\right) \\ &= \Pr\left(\frac{d(z_i, x)}{R_{x,m}} \leq \alpha < \frac{d(z_i, y)}{R_{y,m}}\right) \\ &\leq \max\left(\frac{d(z_i, y)}{R_{y,m}} - \frac{d(z_i, x)}{R_{x,m}}, 0\right).\end{aligned}$$

Applying the triangle inequality $d(z_i, y) \leq d(z_i, x) + d(x, y)$ and the inequality $R_{x,m} - R_{y,m} \leq d(x, y)$ we get

$$\begin{aligned} \frac{d(z_i, y)}{R_{y,m}} - \frac{d(z_i, x)}{R_{x,m}} &\leq \frac{d(z_i, x) + d(x, y)}{R_{y,m}} - \frac{d(z_i, x)}{R_{x,m}} \\ &= \frac{d(z_i, x)}{R_{x,m}} \cdot \frac{R_{x,m} - R_{y,m}}{R_{y,m}} + \frac{d(x, y)}{R_{y,m}} \leq \frac{2d(x, y)}{R_{y,m}}. \end{aligned}$$

We now show that

$$\Pr\left(f(x) = z_i \mid z_i \in \mathrm{B}(x, \alpha R_{x,m}) \text{ and } z_i \notin \mathrm{B}(y, \alpha R_{y,m})\right) \leq \frac{1}{i}.$$

Indeed for any fixed $\alpha = \alpha_0$, if z_i lies in the ball $\mathrm{B}(x, \alpha_0 R_{x,m})$, then the points $z_1, \ldots, z_{i-1}$ also lie in this ball. Therefore, conditionally on $\alpha = \alpha_0$ the probability that z_i is the minimal point with respect to the order $<_\pi$ is at most $1/i$.

Now we can bound (1) as follows:

$$\Pr\left(f(x) <_\pi f(y)\right) = \sum_{i=1}^{m} \frac{1}{i} \times \frac{2d(x, y)}{R_{y,m}} \leq (2 \log m + O(1)) \times \frac{d(x, y)}{R_{y,m}}.$$

Hence

$$\Pr\left(f(x) \neq f(y)\right) \leq (2 \log m + O(1)) \times d(x, y) \times \left(\frac{1}{R_{x,m}} + \frac{1}{R_{y,m}}\right).$$

We are almost done. Assume without loss of generality that $R_{x,m} \leq R_{y,m}$. If $R_{x,m} \geq d(x, y)$, then $R_{y,m} \leq 2R_{x,m}$ and hence

$$\begin{aligned} \Pr\left(f(x) \neq f(y)\right) &\leq (2 \log m + O(1)) \times d(x, y) \times \left(\frac{1}{R_{x,m}} + \frac{1}{R_{y,m}}\right) \\ &\leq (9 \log m + O(1)) \times \frac{d(x, y)}{R_{x,m} + R_{y,m}}. \end{aligned}$$

If $R_{x,m} \leq d(x, y)$, then

$$\Pr\left(f(x) \neq f(y)\right) \leq 1 \leq 3\frac{d(x, y)}{R_{x,m} + R_{y,m}}.$$

□

2.3 Embedding for Large Scales

We combine the results of the previous two sections and obtain an embedding of X into ℓ_p that separates points at large distances.

Lemma 2.7. *For every finite metric space (X, d) and every positive integer m there exists a subset $S \subset X$ of size at most $\lfloor |X|/m \rfloor$ and a probabilistic mapping $h : X \to S$ such that for every two points x and y in X the following conditions hold:*

1. $d(x, h(x)) \leq 5R_{x,m}$ *(always);*

2. $\Pr(h(x) \neq h(y)) \leq O(\log m) \times \frac{d(x,y)}{R_{x,m}+R_{y,m}}$;

3. $\mathbb{E}[d(h(x), h(y))] \leq O(\log m) \times d(x, y)$;

4. $\mathbb{E}[d(h(x), h(y))] \geq d(x, y) - 5(R_{x,m} + R_{y,m})$.

Proof. Choose the set S as in Corollary 2.5 and let f and g be mappings from Lemma 2.6 and Corollary 2.5. Define $h(x) = g(f(x))$. Let us verify that conditions 1-4 are satisfied.

1. We have

$$d(x, h(x)) \leq d(x, f(x)) + d(f(x), g(f(x))) \leq R_{x,m} + 2R_{f(x),m} \leq R_{x,m} + 4R_{x,m} = 5R_{x,m}.$$

Here we used the following simple observation:

$$R_{f(x),m} \leq R_{x,m} + d(x, f(x)) \leq 2R_{x,m}.$$

2. The probability of the event $h(x) \neq h(y)$ is not greater than the probability of the event $f(x) \neq f(y)$. Therefore, the second condition follows from Lemma 2.6.

3. Verify the third condition:

$$\begin{aligned}
\mathbb{E}[d(h(x), h(y))] &= \mathbb{E}[d(h(x), h(y)) \mid h(x) \neq h(y)] \Pr(h(x) \neq h(y)) \\
&\leq \mathbb{E}[d(x, y) + d(x, h(x)) + d(y, h(y)) \mid h(x) \neq h(y)] \Pr(h(x) \neq h(y)) \\
&\leq d(x, y) + 5(R_{x,m} + R_{y,m}) \times O(\log m) \times \frac{d(x, y)}{R_{x,m} + R_{y,m}} \\
&= O(\log m) \times d(x, y).
\end{aligned}$$

4. Finally,

$$\mathbb{E}[d(h(x), h(y))] \geq d(x, y) - \mathbb{E}[d(x, h(x)) + d(y, h(y))] \geq d(x, y) - 5(R_{x,m} + R_{y,m}).$$

□

We are ready to finish the construction of the embedding that handles large distances.

Lemma 2.8. *Let (X, d) be a finite metric space and let k be a positive integer. Suppose that every subset $S \subset X$ of size k is embeddable into a normed space $(V, \|\cdot\|)$ with distortion D. Then there exists an embedding $\varphi : X \hookrightarrow V$ such that for all x and y,*

1. $\|\varphi(x) - \varphi(y)\| \leq D \cdot O(\log(|X|/k)) \times d(x, y)$;

2. $\|\varphi(x) - \varphi(y)\| \geq d(x, y) - (7D + 2) \times (R_{x,m} + R_{y,m})$.

Proof. Set $m = \lceil |X|/k \rceil$. Pick a set S and a random mapping h as in Lemma 2.7. Since the size of S is at most k there exists a distortion D embedding $\nu : S \hookrightarrow V$. By rescaling it we may assume that for every x and y in S:

$$d(x,y) \leq \|\nu(x) - \nu(y)\| \leq D \times d(x,y).$$

Define φ as follows:

$$\varphi(x) = \mathbb{E}\left[\nu(h(x))\right].$$

Verify that it satisfies condition 1:

$$\begin{aligned} \|\varphi(x) - \varphi(y)\| &= \|\mathbb{E}\left[\nu(h(x)) - \nu(h(y))\right]\| \leq \mathbb{E}\|\nu(h(x)) - \nu(h(y))\| \\ &\leq \mathbb{E}\left[D \times d(h(x), h(y))\right] \leq D \times O(\log(|X|/k)) \times d(x,y). \end{aligned}$$

Consider an arbitrary x' in the intersection $S \cap \mathrm{B}(x, 2R_{x,m})$ and y' in $S \cap \mathrm{B}(y, 2R_{y,m})$. Then

$$\begin{aligned} \|\varphi(x) - \nu(x')\| &\leq \mathbb{E}\,\|\nu(h(x)) - \nu(x')\| \leq \mathbb{E}\left[D \times d(h(x), x')\right] \\ &\leq \mathbb{E}\left[D \times (d(h(x), x) + d(x, x'))\right] \leq 7D \times R_{x,m}, \end{aligned}$$

and similarly $\|\varphi(y) - \nu(y')\| \leq 7D \times R_{y,m}$. Therefore,

$$\begin{aligned} \|\varphi(x) - \varphi(y)\| &\geq \|\nu(x') - \nu(y')\| - \|\varphi(x) - \nu(x')\| - \|\varphi(y) - \nu(y')\| \\ &\geq d(x', y') - 7D \times (R_{x,m} + R_{y,m}) \\ &\geq d(x,y) - (7D + 2) \times (R_{x,m} + R_{y,m}). \end{aligned}$$

□

2.4 Bourgain's Embedding for Small Scales

We show that Bourgain's embedding applied at the first $\log(n/k)$ scales preserves short distances and does not expand long distances. The proof is a slight modification of Bourgain's original argument [8].

Lemma 2.9. *For every finite metric space (X,d), every real $p \geq 1$ and every positive integer m there exists an embedding $\psi : X \hookrightarrow \ell_p$ such that for every x and y in X*

1. $\|\psi(x) - \psi(y)\|_p \leq d(x,y)$;
2. $\|\psi(x) - \psi(y)\|_p \geq \Omega(1/\log m) \times \min(d(x,y), R_{x,m} + R_{y,m})$.

Proof. Let $\ell = \lceil \log m \rceil$. Pick a random integer number r from 1 to ℓ. Then choose a random subset $W_r \subset X$, where each point of X belongs to W_r with probability 2^{-r} (the choices are independent for distinct points). Now, for every x in X, define $\psi(x)$ to be the distance from x to the set W_r:

$$\psi(x) = d(x, W_r) \equiv \min_{w \in W_r} d(x,w).$$

Note that $\psi(x)$ is a random variable. The ℓ_p norm is defined in the standard way:

$$\|\psi(x) - \psi(y)\|_p = \left(\mathbb{E}\,|\psi(x) - \psi(y)|^p\right)^{1/p}.$$

Fix two arbitrary points x and y in X and verify that ψ satisfies conditions 1 and 2. By the triangle inequality, $|\psi(x) - \psi(y)|$ is always less than or equal to $d(x, y)$. Hence the first condition holds.

By Lyapunov's inequality,

$$\|\psi(x) - \psi(y)\|_p \geq \|\psi(x) - \psi(y)\|_1.$$

Therefore, we need to prove the second condition only for $p = 1$. Let $\Delta = \min(d(x, y), R_{x,m} + R_{y,m})/2$. We will show that for every $0 < t < \Delta$,

$$\Pr\left(\psi(x) \leq t \leq \psi(y) \text{ or } \psi(y) \leq t \leq \psi(x)\right) \geq \Omega(1/\log m); \tag{2}$$

and hence,

$$\|\psi(x) - \psi(y)\|_1 = \int_0^\infty \Pr\left(\psi(x) \leq t \leq \psi(y) \text{ or } \psi(y) \leq t \leq \psi(x)\right) dt \geq \Omega(1/\log m) \times \Delta.$$

Fix an arbitrary t in the segment $[0, \Delta]$. Let m' be the size of the smallest of the balls $\mathrm{B}(x, t)$ and $\mathrm{B}(y, t)$. Without loss of generality assume that $|\,\mathrm{B}(x, t)| = m'$ and $|\,\mathrm{B}(y, t)| \geq m'$. Note that $m' \leq m$, since $t \leq \max(R_{x,m}, R_{y,m})$. Hence there exists a "scale" $i \in \{1, \dots, \ell\}$ such that $2^{i-1} \leq m' \leq 2^i$. Assume that $r = i$, then with a constant probability the set W_i contains no points from $\mathrm{B}(x, t)$ and with a constant probability W_i contains at least one point from $\mathrm{B}(y, t)$. Moreover, since the balls $\mathrm{B}(x, t)$ and $\mathrm{B}(y, t)$ are disjoint, with a constant probability (conditional on $r = i$) both events happen:

$$\Pr\left(\mathrm{B}(x, t) \cap W_r = \varnothing \text{ and } \mathrm{B}(y, t) \cap W_r \neq \varnothing \mid r = i\right) \geq \Omega(1),$$

which implies

$$\Pr\left(\psi(x) \geq t \text{ and } \psi(y) \leq t\right) \geq \Omega(1/\log m).$$

□

Remark 2.10. *A similar statement was independently proved by Abraham, Bartal, and Neiman [2, Theorem 2]. Our lemma is a generalization of the result of Mendel and Naor [22, Lemma 3.4], stating that every metric space that* has an m-center *(which means, in our terms, that all balls* $\mathrm{B}(x, R_{x,m})$ *have a nonempty (mutual) intersection) is embeddable into* ℓ_1 *with distortion* $\log m$*. Notice that if a metric space has an* m *center, then* $d(x, y) \leq R_{x,m} + R_{y,m}$ *for every* x *and* y*. Therefore, in this case, our embedding works for all distances. Mendel and Naor also used the first* $\log m$ *scales of Bourgain's embedding in their proof.*

2.5 Proof of Theorem 2.1

In this section, we finish the proof of Theorem 2.1.

Proof of Theorem 2.1. Let φ and ψ be the embeddings from Lemmas 2.8 and 2.9. After appropriate rescaling we get that

- for all x and y,
$$\|\varphi(x) - \varphi(y)\|_p \leq O(D \log(|X|/k) \times d(x,y));$$
$$\|\psi(x) - \psi(y)\|_p \leq O(D \log(|X|/k) \times d(x,y));$$
- for all x and y with $d(x,y) \geq 10D \times (R_{x,m} + R_{y,m})$,
$$\|\varphi(x) - \varphi(y)\|_p \geq d(x,y);$$
- for all x and y with $d(x,y) \leq (R_{x,m} + R_{y,m})$,
$$\|\psi(x) - \psi(y)\|_p \geq 10D \times d(x,y) \geq d(x,y);$$
- finally, for all x and y with $d(x,y) \geq (R_{x,m} + R_{y,m})$,
$$\|\psi(x) - \psi(y)\|_p \geq 10D \times (R_{x,m} + R_{y,m}).$$

Notice that ψ expands all distances by a factor at least $10D$. The desired embedding is the direct sum of the embeddings φ and ψ. It is easy to see that it is expanding, but does not increase distances more than $D \cdot O(\log(|X|/k))$ times. □

2.6 Embedding using random samples

Suppose that only an α fraction of all subsets of size k embeds into ℓ_p with distortion at most D. We show that the entire space embeds into ℓ_p with distortion at most

$$D \cdot O(\log(|X|/k) + \log\log(1/\alpha) + \log p).$$

Lemma 2.11. *Let (X,d) be a metric space and $\gamma \in (0,1)$. Consider a probabilistic distribution $\mathcal{D}_\gamma$ of subsets $Y \subset X$ with measure $\Pr(\{Y\}) = \gamma^{|Y|}(1-\gamma)^{|X \setminus Y|}$, that is, each x in X belongs to Y with probability γ. Denote the event "Y* embeds into ℓ_p with distortion D*" by $\mathcal{E}$. Then the entire space X embeds into ℓ_p with distortion at most*

$$O(D \times (\log(1/\gamma) + \log\log(1/\Pr(\mathcal{E})) + \log p)). \tag{3}$$

Proof. We use the same embedding as in Theorem 2.1 with

$$k = \left\lfloor \frac{\gamma |X|}{\log(1/\Pr(\mathcal{E})) + 2p} \right\rfloor.$$

The distortion of this embedding is at most $O(D\log(|X|/k)) = (3)$. The only property of X we used in the proof of Theorem 2.1 is that the set S from Lemma 2.3 embeds into ℓ_p with distortion D (see Lemma 2.8). We shall prove a slightly weaker statement which is still sufficient for the proof. Namely, we show that there exists an embedding ν of S into ℓ_p such that for all x and y in S,

A: $\|\nu(x) - \nu(y)\|_p \le O(D) \times d(x, y)$;

B: $\|\nu(x) - \nu(y)\|_p \ge \Omega(1) \times (d(x, y) - (R_{x,m} + R_{y,m}))$.

As before, let $m = \lceil |X|/k \rceil$. Let T_Y be the set of those points x in S for which there exists a point y in Y at distance at most $R_{x,m}$. Since each ball $\mathrm{B}(x, R_{x,m})$ contains m points, the probability that x belongs to T_Y is equal to $1 - \delta$, where

$$\delta = (1-\gamma)^m \le \frac{\Pr(\mathcal{E})}{e^{2p}}.$$

Moreover, since for every distinct x and y in S, the balls $\mathrm{B}(x, R_{x,m})$ and $\mathrm{B}(y, R_{y,m})$ are disjoint, the events $x \in T_Y$ and $y \in T_Y$ are independent.

Let $\mathcal{E}'$ be the event "*Y embeds into ℓ_p with distortion D and T_Y is not empty*". Then $\Pr(\mathcal{E}') \ge \Pr(\mathcal{E}) - \Pr(T_Y = \varnothing) \ge \Pr(\mathcal{E}) - \delta > 4\Pr(\mathcal{E})/5$. Let $\tilde{Y}$ be a random set distributed according to $\mathcal{D}_\gamma$ conditional on the event $\mathcal{E}'$; denote the random set $T_{\tilde{Y}}$ by $\tilde{T}$.

Fix now a subset Y of X that embeds into ℓ_p with distortion D. Consider the embedding $\nu_Y : T_Y \hookrightarrow \ell_p$ that maps each x in T_Y to the closest point y in Y and then embeds Y into ℓ_p (using a non-contracting embedding). Notice that ν_Y satisfies properties A and B. Below we describe an algorithm that *extends* ν_Y to the set S: for every nonempty set T and every point x in S it returns a point $q_T(x)$ in T. We show that the mapping of x to the random variable $\nu_{\tilde{Y}}(q_{\tilde{T}}(x))$ is an embedding of S into ℓ_p satisfying properties A and B.

Input: a nonempty set T;
Output: mapping $q_T : S \to T$;

1. Enumerate all points in S with numbers 1 to k:
   - Pick an arbitrary starting point x_1 in S;
   - For all $i < k$, let x_{i+1} be the closest point to x_i in the set $S \setminus \{x_1, \dots, x_i\}$.

   Remark: the ordering of points $x_1, \dots, x_k$ does not depend on the set T.

2. Consider the following *walk*: x_i goes to x_{i+1} unless there is a point y_i in T at distance less than $d(x_i, x_{i+1})$, in which case x_i goes to y_i. More precisely, at every step, x_i goes to

$$N(x_i) = \begin{cases} y_i, & \text{if } d(x_i, y_i) \le d(x_i, x_{i+1}); \\ x_{i+1}, & \text{otherwise;} \end{cases}$$

where y_i is the closest point to x_i in T. Note that $N(x_k) = y_k$. Continue to move x_i till it hits the set T:

$$x_i \mapsto x_{i+1} \mapsto \cdots \mapsto x_{i+t} \mapsto y_{i+t} \in T.$$

Denote the number of steps needed to reach T by t_i:

$$t_i = \min \left\{ t : N^t(x_i) \in T \right\};$$

and let $q_T(x_i)$ be the last point of the walk:

$$q_T(x_i) = N^{t_i}(x_i).$$

3. Return the mapping q_T.

Notice that the algorithm does not move elements of the set T. If $i < j$ then

$$d(N(x_i), x_j) \leq d(x_i, N(x_i)) + d(x_i, x_j) \leq 2d(x_i, x_j);$$

and if $x_i \in T$ then similarly $d(x_i, N(x_j)) \leq 2d(x_i, x_j)$. Hence

$$d(q_T(x_i), q_T(x_j)) \leq 2^{t_i+t_j} d(x, y).$$

Estimate the probability of the event $t_i \geq t$ (for a random Y). If $t_i \geq t$ then $x_i, \ldots, x_{i+t-1}$ do not belong to T_Y. Hence, this probability is at most δ^t and the probability that $t_i + t_j \geq t$ is at most $2\delta^{\lceil t/2 \rceil}$. We have for positive t,

$$\Pr\left(t_i + t_j \geq t \mid \mathcal{E}'\right) = \frac{2\delta^{\lceil t/2 \rceil}}{\Pr\left(\mathcal{E}'\right)} \leq \frac{5}{2} e^{-pt};$$

and

$$\Pr\left(q_{\tilde{T}}(x) = x \text{ and } q_{\tilde{T}}(y) = y\right) = \Pr\left(t_i + t_j = 0 \mid \mathcal{E}'\right) \geq 1 - \frac{5}{2e} = \Omega(1).$$

Therefore,

$$\begin{aligned}
\|\nu_{\tilde{Y}}(q_{\tilde{T}}(x)) - \nu_{\tilde{Y}}(q_{\tilde{T}}(y))\|_p &= \mathbb{E}\left[\|\nu_Y(q_{T_Y}(x)) - \nu_Y(q_{T_Y}(y))\|_p^p \mid \mathcal{E}'\right]^{1/p} \\
&\leq O(D\, d(x, y)) \times \left(1 + \sum_{t=1}^{\infty} 2^{tp} \Pr\left(t_i + t_j \geq t \mid \mathcal{E}'\right)\right)^{1/p} \\
&\leq O(D\, d(x, y)).
\end{aligned}$$

and

$$\begin{aligned}
\|\nu_{\tilde{Y}}(q_{\tilde{T}}(x)) - \nu_{\tilde{Y}}(q_{\tilde{T}}(y))\|_p &= \mathbb{E}\left[\|\nu_Y(q_{T_Y}(x)) - \nu_Y(q_{T_Y}(y))\|_p^p \mid \mathcal{E}'\right]^{1/p} \\
&\geq \left(d(x, y) - R_{x,m} - R_{y,m}\right) \Pr\left(q_{\tilde{T}}(x) = x \text{ and } q_{\tilde{T}}(y) = y\right) \\
&\geq \Omega(1) \times (d(x, y) - R_{x,m} - R_{y,m}).
\end{aligned}$$

□

The following theorem is an easy corollary of the lemma we proved.

Theorem 2.12. *Let (X, d) be a metric space of size n. Suppose that an α fraction of all subsets of size k embeds into ℓ_p with distortion at most D. Then entire space (X, d) embeds into ℓ_p with distortion at most*

$$D \cdot O(\log(n/k) + \log\log(1/\alpha) + \log p).$$

Proof. Let $\gamma = k/(2n)$. Then a random set Y distributed as $\mathcal{D}_\gamma$ contains at most k elements with probability at least half. Therefore, Y embeds into ℓ_p with distortion D with probability at least $\alpha/2$. □

Remark 2.13. *In the bound above, one can replace* $\log p$ *with* $\log D$*. The proof of the new bound is similar to the previous one.*

3 Lower Bound Constructions

3.1 Embedding Sparse Graphs into ℓ_1

Now we present our lower bounds. We will construct metric spaces that embed locally into ℓ_1 with small distortion, but whose global distortion is large. In this section, we prove that every sparse graph with high girth equipped with an appropriate metric embeds into ℓ_1 with very small distortion. Later we will use this to show that the metric spaces we construct locally embed into ℓ_1 with small distortion.

Definition 3.1. *A graph G is α-sparse, if every subgraph on k vertices contains at most αk edges.*

To embed a sparse graph, we will decompose it to several pieces, embed them separately, and then combine the embeddings. We will need the following definition, which was implicitly introduced in [3].

Definition 3.2. *We say that a graph G is l-path decomposable if every 2-connected subgraph H of G (other than an edge) contains a path of length l such that every vertex of the path has degree 2 in H.*

Every l-path decomposable graph is either

1. a vertex or an edge; or
2. the union of (more than one) connected components, each of which is also l-path decomposable; or
3. a one point union of l-path decomposable (proper) subgraphs (we say that a union of several graphs is a *one point union* if there exists a vertex belonging to all graphs; and the intersection of every two of these graphs contains only this vertex); or

4. the union of an l-path decomposable subgraph and a path of length l (that do not have common vertices except for the endpoints of the path).

Arora, Bollobás, Lovász, and Tourlakis [3] proved that every $1+\eta$ sparse graph with girth $\Omega(1/\eta)$ is $\Omega(1/\eta)$-path decomposable.

Theorem 3.3. *Suppose $G=(V,E)$ is an l-path decomposable graph. Let $d(\cdot,\cdot)$ be the shortest path distance on G, and $L=\lfloor l/9\rfloor$; $\mu\in[1/L,1]$. Then there exists a probabilistic distribution of multicuts of G (or in other words random partition of G in pieces) such that the following properties hold. For every two vertices u and v,*

1. *If $d(u,v)\leq L$, then the probability that u and v are separated by the multicut (i.e. lie in different parts) equals*

$$\rho(u,v)=\rho_\mu(u,v)=1-(1-\mu)^{d(u,v)};$$

moreover, if u and v lie in the same part, then the unique shortest path between u and v also lies in that part.

2. *If $d(u,v)>L$, then the probability that u and v are separated by the multicut is at least $1-(1-\mu)^L$.*

3. *Every piece of the multicut partition is a tree.*

Proof idea. Each multicut is a subset S of edges (i.e. the edges removed to obtain the partition): the multicut S separates two vertices if every path between them intersects S. We will ensure that every edge belongs to S with probability μ. Additionally, if the distance between two edges e_1,e_2 is less than L, the events that $e_1\in S$, and $e_2\in S$ will be independent. We will also ensure that if there is more than one simple path between u and v, all but one path will be cut with probability 1. Our proof will be by induction: using the path-decomposability of G, we will reduce the problem to smaller subproblems. In order to argue about paths of length l which we will encounter during the decomposition we need the following lemma.

Lemma 3.4. *Suppose a graph H is a path of length at least $3L$. Then there exists a distribution of multicuts S of H that satisfies the property of the theorem. Moreover, the endpoints of the path are separated with probability 1.*

Proof. Let us subdivide the path into three paths P_1, P_2, and P_3, each of length at least L. We now add every edge to S with probability μ. However, our decisions are not independent, and we add edges so that: all decisions for P_1 and P_2 are independent; all decisions for P_2 and P_3 are independent; we add to S at least one edge either from P_1 or from P_3. This coupling is possible since the probability that we add at least one edge from P_1 plus the probability that we add at least one edge from P_3 is at least $1-(1-\mu)^L+1-(1-\mu)^L\geq 2-2/e>1$.

First assume that both vertices u and v lie in the same path P_i or they lie in the neighboring paths P_i and P_{i+1}. Then all our choices for the edges between u and v are independent.

Therefore, the probability that S separates u and v is $1-(1-\mu)^{d(u,v)} = \rho(u,v)$. Hence S satisfies condition 1 in this case. Now assume that u lies in P_1 and v lies in P_3 (or vice versa). Then $d(u,v) > L$. If S contains an edge from P_2 then u and v are separated, so

$$\Pr(u \text{ and } v \text{ are separated}) \geq \Pr\left(S \cap E(P_2) \neq \varnothing\right) = 1-(1-\mu)^{|E(P_2)|} \geq 1-(1-\mu)^L.$$

Finally, since S contains either an edge from P_1 or from P_3, the endpoints of H are always separated. □

Proof of Theorem 3.3. We prove by induction that there exists the required distribution on multicuts $S \subset E$. We first verify the base case. If G consists of two vertices connected by an edge then let S contain this edge with probability μ. Otherwise, G is decomposable into the union of smaller l-path connected subgraphs. Consider three cases.
1. The graph G is the union of connected components C_i. Since each C_i has less vertices than G, by the induction hypothesis, there exists a probability distribution on multicuts S_i in each C_i. Let $S = \cup_i S_i$, where all S_i are drawn independently. Then if u and v lie in the same connected component C_i then $\Pr(S \text{ separates } u \text{ and } v) = \Pr(S_i \text{ separates } u \text{ and } v)$; if u and v lie in distinct connected components $\Pr(S \text{ separates } u \text{ and } v) = 1 \geq 1-(1-\mu)^L$ whereas $d(u,v) = \infty > L$.
2. The graph G has a cut vertex c. Represent G as the union of subgraphs C_i that have only one common vertex c. Construct a distribution of multicuts S_i for each C_i. Let $S = \cup_i S_i$, where all S_i are drawn independently. Assume first that u and v lie in the same subgraph C_i. Then every simple path between them lies entirely in C_i. Therefore, S separates u and v if and only if S_i separates them. By the induction hypothesis, the probability that the multicut S separates u and v satisfies condition 1. Now assume that u and v lie in distinct subgraphs: $u \in C_i$, $v \in C_j$. Then each path between u and v must visit c. We have

$$\begin{aligned}
\Pr(S \text{ separates } u \text{ and } v) &= \Pr\left(S \text{ separates } u \text{ and } c \text{ or } S \text{ separates } c \text{ and } v\right)\\
&= 1-(1-\Pr\left(S \text{ separates } u \text{ and } c\right))(1-\Pr\left(S \text{ separates } c \text{ and } v\right))\\
&= 1-(1-\Pr\left(S_i \text{ separates } u \text{ and } c\right))(1-\Pr\left(S_j \text{ separates } c \text{ and } v\right)).
\end{aligned}$$

If $d(u,c) \leq L$ and $d(v,c) \leq L$, then

$$\begin{aligned}
\Pr(S \text{ separates } u \text{ and } v) &= 1-(1-\rho(u,c))(1-\rho(c,v))\\
&= 1-(1-\mu)^{d(u,c)}(1-\mu)^{d(c,v)} = 1-(1-\mu)^{d(u,c)+d(c,v)} = \rho(u,v).
\end{aligned}$$

Here we used that $d(u,v) = d(u,c) + d(c,v)$. Now consider the case when either $d(u,c) \geq L$ or $d(v,c) \geq L$. Assume without loss of generality $d(u,c) \geq L$. Note that $d(u,v) = d(u,c) + d(c,v) > L$. Hence we have to show that u are v are separated with probability at least $1-(1-\mu)^L$. Indeed,

$$\begin{aligned}
\Pr(S \text{ separates } u \text{ and } v) &= 1-(1-\Pr(S_i \text{ separates } u \text{ and } c))(1-\Pr(S_j \text{ separates } c \text{ and } v))\\
&\geq 1-(1-\Pr(S_i \text{ separates } u \text{ and } c)) \cdot 1\\
&= \Pr(S_i \text{ separates } u \text{ and } v) \geq 1-(1-\mu)^L.
\end{aligned}$$

Finally, since each piece of every multicut partition S_i is a tree, every piece of S is also a tree.

3. The graph G is the union of a subgraph H and a path P of length l. Denote the endpoints of the path by x and y. Let S_H be a distribution on multicuts in H that satisfies condition 1. Subdivide P into three pieces A_1, A_2 and A_3, each of length at least $3L$. Let S_i be the multicut whose existence is guaranteed by Lemma 3.4 for the path A_i (we choose multicuts S_H, S_1, S_2, S_3 independently). Let $S = S_H \cup S_1 \cup S_2 \cup S_3$. Consider two vertices u and v. Note that either both of them lie in $H \cup A_1 \cup A_2$, or in $H \cup A_2 \cup A_3$, or in $H \cup A_1 \cup A_3$ (of course, these possibilities are not mutually exclusive). First, assume that u and v lie in $H \cup A_1 \cup A_2$. Since A_3 is always cut by S, the multicut S separates u and v in G if and only if the multicut $S_H \cup S_1 \cup S_2$ separates them in $H \cup A_1 \cup A_2$. Additionally, if $d_G(u, v) \leq L$ then $d_{H\cup A_1\cup A_2}(u, v) = d_G(u, v)$, and if $d_G(u, v) > L$ then $d_{H\cup A_1\cup A_2}(u, v) > L$. Therefore, it suffices to verify that condition 1 holds for $H \cup A_1 \cup A_2$. Indeed, the graph $H \cup A_1 \cup A_2$ is a one point union of graphs H and $A_1 \cup A_2$; in turn, the graph $A_1 \cup A_2$ is a one point union of graphs A_1 and A_2. We already proved that a one point union of graphs satisfying condition 1 satisfies condition 1. Therefore, the multicut $S_H \cup S_1 \cup S_2$ of the graph $H \cup A_1 \cup A_2$ satisfies condition 1. Similarly, condition 1 holds when u and v lie in $H \cup A_2 \cup A_3$, or u and v lie in $H \cup A_1 \cup A_3$.

Finally, we know that every cycle in H is cut by the multicut (by the induction hypothesis) and the path P is cut by the multicut. Therefore, every piece of the constructed multicut partition is a tree. □

Corollary 3.5. *Let $G = (V, E)$, L, μ and ρ be as in Theorem 3.3. Then the metric space (V, ρ) embeds into ℓ_1 with distortion $1+O(e^{-\mu L})$. Moreover, if $d(u, v) \leq L$ then the embedding preserves the distance between u and v; if $d(u, v) > L$ then the distance between images of u and v lies between $1 - (1 - \mu)^L$ and 1.*

Proof. The distribution of multicuts from Theorem 3.3 defines the desired embedding[3]: the distance between images of u and v equals the probability that the multicut separates u and v. If $d(u, v) \leq L$ then the distance between u and v is preserved by the embedding. If $d(u, v) > L$ then the distance between u and v in the cut metric lies between $1 - (1 - \mu)^L$ and 1. Hence the distortion is at most

$$\frac{1}{1 - (1 - \mu)^L} = 1 + O(e^{-\mu L}).$$

□

3.2 Lower Bound for Non-Isometric Case

In this section, we construct a metric space that locally embeds into ℓ_1 almost isometrically, but the minimum distortion with which the entire space embeds into ℓ_1 is high. The metric

[3]Given a distribution of multicuts we construct an embedding to ℓ_1 as follows: pick a random multicut from the distribution; map all vertices in every part to either 1 or -1 with probability a half. Then the expected distance between images of two vertices exactly equals the probability that the vertices are separated by the multicut.

space will be based on a 3-regular expander graph whose k vertex subgraphs are sparse. The underlying expander graph was used by Arora, Lovász, Newman, Rabani, Rabinovich, and Vempala [5] in their construction of such metric space. However, we equip this graph with a different metric. We need the following lemma that was proved by Arora, Bollobás, Lovász, and Tourlakis [3].

Lemma 3.6 ([3], Lemma 2.8, Lemma 2.12; see also [5], Lemma 3.3)**.** *There exists a 3-regular expander graph on n vertices with girth $\Omega(\log n)$ such that every subset of k points has sparsity $1 + O(\frac{1}{\log(n/k)})$. Therefore, this expander is $\Omega(\log(n/k))$-path decomposable.*

We are ready to prove the following theorem.

Theorem 3.7.
I. For every n, $k < n$ and $\delta \in (0, 1/2]$, there exists a metric space (X, ρ) on n points such that

- *every embedding of (X, ρ) into ℓ_1 requires distortion $\Omega\left(\frac{\log(n/k)}{\log 1/\delta}\right)$;*
- *every subset of X of size k embeds into ℓ_1 with distortion $1 + \delta$.*

Moreover, the aspect ratio of X (i.e. the ratio between the diameter of X and the minimal distance between two points) is $O(\log(n/k))$.
II. For every n, $k < n$ and $D \in (1, \frac{\log n}{\log(n/k)})$, there exists a metric space (X, ρ) on n points such that

- *every embedding of (X, ρ) into ℓ_1 requires distortion $\Omega(D \log(n/k))$;*
- *every subset of X of size k embeds into ℓ_1 with distortion $O(D)$.*

Proof.
I. Let G be the expander from Lemma 3.6. Denote the set of its vertices by X. We know that every subgraph of G on $k \cdot 3 \cdot 2^l$ vertices (for every l) is $\Omega(\log(n/(k \cdot 3 \cdot 2^l)))$-path decomposable. Choose $l = \Theta(\log(n/k))$ so that every subgraph on $k \cdot 3 \cdot 2^l$ vertices is l-path decomposable. Let $\mu = c \log(1/\delta)/l$, where c is a sufficiently large constant.

Let us equip X with the metric ρ defined as

$$\rho(u, v) = \rho_\mu(u, v) = 1 - (1 - \mu)^{d(u,v)},$$

where $d(u, v)$ is the shortest path distance between u and v in G. We will now prove that every subset of X of size k embeds into ℓ_1 with distortion at most $1 + \delta$. Let Y be a subset of X of size k. Consider the set of vertices $B_d(Y, l)$ whose distance to Y is at most l: $B_d(Y, l) = \{x : d(x, Y) \le l\}$. Let H be the graph induced by $B_d(Y, l)$ on G. Since degree of each vertex in G is at most 3, the size of $B_d(Y, l)$ is at most $k \cdot 3 \cdot 2^l$. Therefore, H is l-path decomposable. By Corollary 3.5, there exists an embedding $\psi : B_d(Y, l) \to \ell_1$ such that for every $u, v \in B_d(Y, l)$ (for $L = \lfloor l/9 \rfloor$):

1. if $d_H(u, v) \le L$ then $\|\psi(u) - \psi(v)\|_1 = 1 - (1 - \mu)^{d(u,v)}$;
2. if $d_H(u, v) > L$ then $1 \ge \|\psi(u) - \psi(v)\|_1 \ge 1 - (1 - \mu)^L$.

Note that since H is a subgraph of G, the shortest path distance between two vertices in H, $d_H(u,v)$, is at least the shortest path distance between them in G, $d_G(u,v) \equiv d(u,v)$. However, if $u,v \in Y$, and $d_G(u,v) \le l$ then the shortest path between them lies in H. Hence $d_H(u,v) = d_G(u,v)$. Therefore, for $u,v \in Y$, if $d_G(u,v) \le L$ then $\|\psi(u) - \psi(v)\|_1 = \rho(u,v)$; if $d_G(u,v) > L$ then $1 \ge \|\psi(u) - \psi(v)\|_1 \ge 1-(1-\mu)^L$. Hence the distortion of the embedding $\psi : (Y,\rho) \hookrightarrow \ell_1$ is at most $1+O(e^{-\mu L}) = 1+O(e^{-c/9\cdot\log(1/\delta)}) < 1+\delta$ (if we choose c sufficiently large).

As was shown by Linial, London and Rabinovich [16] and Aumann and Rabani [7], the distortion with which a bounded degree expander graph is embeddable into ℓ_1 is at least (up to a constant factor) the ratio of the average distance between all vertices in the graph to the average length of an edge:

$$\frac{2}{n(n-1)} \sum_{u,v\in V(G)} \rho(u,v) \Bigg/ \frac{1}{|E(G)|} \sum_{(u,v)\in E(G)} \rho(u,v)\ .$$

Therefore, the least distortion with which (X,ρ) embeds into ℓ_1 is

$$\frac{\Omega(1)}{(1-(1-\mu))} = \Omega\left(\frac{\log(n/k)}{\log(1/\delta)}\right). \tag{4}$$

Finally, the diameter of X is at most 1; the minimal distance between two points is μ. Hence the aspect ratio is $O(\log(n/k))$.

II. Consider the metric space $(X, \rho \equiv \rho_\mu)$ constructed in part I for $\delta = 1/2$. Recall that $\mu = \Theta(1/\log(n/k))$.

Let us equip X with a new distance function $\rho_{\mu/D}(u,v) = 1-(1-\mu/D)^{d(u,v)}$ (where $d(u,v)$ is the shortest path metric in the underlying expander). Note that since

$$\frac{1}{D}\left(1-(1-\mu)^{d(u,v)}\right) \le 1-(1-\mu/D)^{d(u,v)} \le 1-(1-\mu)^{d(u,v)},$$

and every k point subspace of (X,ρ_μ) embeds into ℓ_1 with a constant distortion, then every k point subspace of $(X,\rho_{\mu/D})$ embeds into ℓ_1 with distortion at most $O(D)$. On the other hand, the entire space $(X,\rho_{\mu/D})$ embeds into ℓ_1 with distortion at least (similarly to (4))

$$\frac{\Omega(1)}{(1-(1-\mu/D))} = \Omega\left(D\log\frac{n}{k}\right).$$

(The average distance in X with respect to $\rho_{\mu/D}$ is a constant, since if $d(u,v) = \Omega(\log n)$ then $\rho_{\mu/D}(u,v) = 1-e^{-\Omega(\mu\log n/D)}$. Recall that $D < \log n/\log(n/k)$. Hence $\rho_{\mu/D}(u,v) = \Omega(1)$.) □

3.3 Lower Bound for Isometric Case

In this section, we present a metric space such that every subset of size k isometrically embeds into ℓ_1, whereas every embedding of the entire space into ℓ_1 requires distortion at least $\Omega(\log n/(\log\log n + \log k))$. Our construction will be a perturbation of the metric space we presented in Section 3.2.

Theorem 3.8. *Consider a metric space (X, ρ) on n points. Let $k < n$; let M be the aspect ratio (i.e. the ratio between the diameter of X and the shortest distance). Suppose now that every subspace of (X, ρ) of size k embeds into ℓ_1 with distortion at most $1 + 1/(2kM)$. Then there exists a 2-Lipschitz equivalent metric $\hat{\rho}$ on X such that every subspace of $(X, \hat{\rho})$ of size k embeds isometrically into ℓ_1.*

We need the following definition and lemma.

Definition 3.9. *Let $S(u, v)$ be the metric defined by $S(u, v) = 1$, if $u \neq v$; and $S(u, v) = 0$, if $u = v$. This metric is often called the discrete metric.*

Lemma 3.10. *Consider a metric space (Y, ρ) on k points. If for every two points u and v from Y:*

$$|\rho(u, v) - S(u, v)| \leq \frac{1}{2k},$$

then (Y, ρ) is isometrically embeddable into ℓ_2.

Proof. We will prove that the matrix

$$G_{uv} = 1 - \rho(u, v)^2/2$$

is positive semidefinite and, therefore, there exists a set of unit vectors z_u in ℓ_2 such that

$$\langle z_u, z_v \rangle = 1 - \rho(u, v)^2/2.$$

This implies that the mapping $u \mapsto z_u$ is an isometric embedding, since

$$\|z_u - z_v\|_2 = \sqrt{\|z_u\|^2 + \|z_v\|^2 - 2\langle z_u, z_v \rangle} = \rho(u, v).$$

Express the matrix G as the sum of three matrices:

$$G = \frac{1}{2} I + \begin{pmatrix} \frac{1}{2} & \cdots & \frac{1}{2} \\ \vdots & \ddots & \vdots \\ \frac{1}{2} & \cdots & \frac{1}{2} \end{pmatrix} + Q.$$

Observe, that the eigenvalues of the matrix $I/2$ are equal to $1/2$; the second matrix is positive semidefinite. Then $|Q_{uv}| \leq 1/(2k) + 1/(8k^2)$ and $Q_{uu} = 0$ for all u and v. Therefore, the eigenvalues of Q are bounded in absolute value by

$$\|Q\|_\infty \leq (k - 1) \times \left[\frac{1}{2k} + \frac{1}{8k^2} \right] = \frac{4k^2 - 3k - 1}{8k^2} \leq \frac{1}{2}.$$

Hence G is positive semidefinite. □

Corollary 3.11. *Consider a metric space (Y, ρ) on k points. If for every two points u and v from Y:*

$$|\rho(u, v) - S(u, v)| \leq \frac{1}{2k},$$

then (Y, ρ) is isometrically embeddable into ℓ_1.

Proof. Every finite subset of ℓ_2 is isometrically embeddable into ℓ_1. □

Remark 3.12. *The condition in this corollary cannot be significantly strengthened: there exists a metric space (Y,ρ) such that $|\rho(u,v) - S(u,v)| = O(1/k)$, however, the space (Y,d) is not isometrically embeddable into ℓ_1.*

Proof of Theorem 3.8. Denote the minimal distance between two distinct vertices of X by δ. Define a new metric $\hat{\rho}$ on X as follows:

$$\hat{\rho}(u,v) = \rho(u,v) + \delta S(u,v).$$

First, the metric $\hat{\rho}$ is 2-Lipschitz equivalent to the metric ρ. Indeed, for $u \neq v$ we have $\rho(u,v) \leq \hat{\rho}(u,v) = \rho(u,v) + \delta \leq 2\rho(u,v)$. Now let Y be a subset of X of size k. By the condition, there is an embedding $\varphi : Y \hookrightarrow \ell_1$ with distortion at most $1 + 1/(2kM)$. Without loss of generality we may assume that

$$\rho(u,v) \leq \|\varphi(u) - \varphi(v)\|_1 \leq (1 + 1/(2kM))\rho(u,v).$$

Since the distance $\rho(u,v)$ is at most $M\delta$ (by the definition of M), we have

$$0 \leq \|\varphi(u) - \varphi(v)\|_1 - \rho(u,v) \leq \delta/(2k).$$

This bound and Corollary 3.11 imply that the set Y equipped with the metric

$$(u,v) \mapsto S(u,v) - (\|\varphi(u) - \varphi(v)\|_1 - \rho(u,v))/\delta$$

embeds into ℓ_1 isometrically. Denote the embedding by ψ. We have

$$\begin{aligned} \hat{\rho}(u,v) &= \|\varphi(u) - \varphi(v)\|_1 + \delta\,(S(u,v) - (\|\varphi(u) - \varphi(v)\|_1 - \rho(u,v))/\delta) \\ &= \|\varphi(u) - \varphi(v)\|_1 + \delta\,\|\psi(u) - \psi(v)\|_1. \end{aligned}$$

Therefore, $\varphi \oplus \delta\psi$ is an isometric embedding of $(Y,\hat{\rho})$ into $\ell_1 \oplus \ell_1 \cong \ell_1$. □

Theorem 3.13. *For every n and $k < n$, there exists a metric space (X,ρ) on n points such that*

- *every embedding of (X,ρ) into ℓ_1 requires distortion $\Omega\left(\frac{\log n}{\log k + \log\log n}\right)$;*
- *every subset of X of size k embeds isometrically into ℓ_1.*

Proof. Let $\delta = \frac{c}{k \log n}$ (where c is sufficiently small). By Theorem 3.7, there exists a metric space (X,ρ) such that X embeds into ℓ_1 with distortion at least $\Omega(\frac{\log(n/k)}{\log(1/\delta)}) = \Omega(\frac{\log n}{\log k + \log\log n})$; every subset of X of size k embeds into ℓ_1 with distortion $1 + \delta$; the aspect ratio of X is $O(\log n)$. Applying Theorem 3.8 to (X,ρ), we get that there exists a 2-Lipschitz equivalent metric $\hat{\rho}$ on X such that every subspace of $(X,\hat{\rho})$ of size k embeds into ℓ_1 isometrically. Since $(X,\hat{\rho})$ is 2-Lipschitz equivalent to (X,ρ), every its embedding into ℓ_1 has distortion at least $\Omega(\frac{\log n}{\log k + \log\log n})$. This concludes the proof. □

3.4 Lower Bounds for Spaces ℓ_p

In this section, we present analogs of the results from Sections 3.2 and 3.3 for spaces ℓ_p.

First we establish lower bounds for metric spaces that locally embed into ℓ_p isometrically or almost isometrically. This result is a simple corollary of our results for embedding into ℓ_1.

Notice that if a metric space (X,ρ) embeds isometrically into ℓ_1, then the metric space $(X,\rho^{1/p})$ embeds isometrically into ℓ_p. Therefore, if (X,ρ) embeds into ℓ_1 with distortion D then $(X,\rho^{1/p})$ embeds into ℓ_p with distortion $D^{1/p}$. We can upper bound the (global) distortion of an embedding $(X,\rho^{1/p})$ into ℓ_p using a theorem by Matoušek [19], which states that every embedding of an expander into ℓ_p has distortion $\Omega(1/p \times \frac{\text{average distance}}{\text{length of edge}})$.

This observation allows us to generalize Theorems 3.7 and 3.13 for ℓ_p spaces.

Theorem 3.14.
I. For every n, $k < n$, $p \geq 1$ and $\delta \in (0, 1/2]$, there exists a metric space (X,ρ) on n points such that

- *every embedding of (X,ρ) into ℓ_1 requires distortion $\Omega\left(\frac{\log(n/k)}{\log 1/\delta}\right)^{1/p}$;*
- *every embedding of (X,ρ) into ℓ_p requires distortion $\frac{1}{p}\cdot\Omega\left(\cdot\frac{\log(n/k)}{\log 1/\delta}\right)^{1/p}$;*
- *every subset of X of size k embeds into ℓ_p with distortion $1+\delta$.*

II. For every n, $k < n$ and $p \geq 1$, there exists a metric space (X,ρ) on n points such that

- *every embedding of (X,ρ) into ℓ_1 requires distortion $\Omega\left(\frac{\log n}{\log k+\log\log n}\right)^{1/p}$;*
- *every embedding of (X,ρ) into ℓ_p requires distortion $\frac{1}{p}\cdot\Omega\left(\frac{\log n}{\log k+\log\log n}\right)^{1/p}$;*
- *every subset of X of size k embeds isometrically into ℓ_p.*

We now construct a metric space that locally embeds into ℓ_p with a small distortion but whose global distortion is almost $\log(n/k)$. Our proof relies on ideas from papers of Bourgain [9] and Matoušek [20] that study low distortion tree embeddings. The authors would like to thank Assaf Naor who suggested how to generalize our preliminary result (which was stated in the conference version of this paper), and who pointed to a relevant paper of Lee, Mendel, and Naor [17].

Theorem 3.15. *For every n, $k < n$, $p \geq 1$ and $D \geq 2$ there exists a metric space (X,ρ) on n points such that*

- *every embedding of (X,ρ) into ℓ_1 requires distortion $\Omega(\log(n/k)^{1-1/D^{\max(p,2)}})$;*
- *every embedding of (X,ρ) into ℓ_p requires distortion $\frac{1}{p}\cdot\Omega(\log(n/k)^{1-1/D^{\max(p,2)}})$;*
- *every subset of X of size k embeds into ℓ_p with distortion $O(D)$.*

In particular, there exists a metric space (X, ρ) such that

- *every embedding of (X, ρ) into ℓ_1 requires distortion $\Omega(\log(n/k))$;*
- *every embedding of (X, ρ) into ℓ_p requires distortion $\frac{1}{p} \cdot \Omega(\log(n/k))$;*
- *every subset of X of size k embeds into ℓ_p with distortion $O((\log\log(n/k))^{\min(1/2,1/p)})$.*

Proof. Since every finite subset of ℓ_2 embeds isometrically into ℓ_p when $p \in [1, 2]$, the metric space for $p = 2$ works for all $p \in [1, 2]$. Thus we may assume below that $p \geq 2$.

We consider the expander graph from Section 3.2 equipped with the metric

$$\rho(u, v) = \min(d(u, v), c \log(n/k))^{1-\varepsilon},$$

where $d(\cdot, \cdot)$ is the shortest path metric, $\varepsilon = 1/D^p$, and c is a sufficiently small constant.

It is immediate that the metric requires distortion at least $\Omega(\log(n/k)^{1-1/D^p})$ for embedding into ℓ_1. In Theorem 3.7, we proved that every k points of the expander graph equipped with a different metric embed into ℓ_1 with distortion at most $1 + \delta$. The proof was based on Corollary 3.5. We can use exactly the same proof to show that every k points in the expander embed with distortion $O(D)$ into ℓ_p. The only thing we need to establish is an analog of Corollary 3.5 for our new metric ρ. We prove it in Theorem 3.16. □

Theorem 3.16. *Let $G = (V, E)$ be a $(20L + 5)$-path decomposable graph with girth at least $4L$, $p \geq 2$, and $\varepsilon \in (0, 1/2)$ be a parameter. Define the distance function*

$$\rho(u, v) = \min(d(u, v), L)^{1-\varepsilon}.$$

Then the metric space (V, ρ) embeds into ℓ_p with distortion $O(D)$, where

$$D = \frac{1}{\varepsilon^{1/p}}.$$

Below we consider *orientations* of graph edges in which every edge (u, v) is directed either from u to v, from v to u or not directed at all (but it cannot be directed both from u to v and from v to u). If the edge is directed from u to v or from v to u, we say that the edge is *regular*; if it is not directed, we say that the edge is *special*. We say that a vertex is a *sink* if it has no regular outgoing edges.

Definition 3.17. *We denote the shortest path between u and v by $\pi(u, v)$. We denote the concatenation of two paths $P^{(1)}$ and $P^{(2)}$ by $P^{(1)} \to P^{(2)}$. The distance between two edges is the distance between the sets of their endpoints. Similarly, the distance between a vertex u and an edge $e = (v_1, v_2)$ is $d(u, e) = d(u, \{v_1, v_2\})$.*

Lemma 3.18. *Let G be an l-path decomposable graph; $R = \lfloor (l - 1)/4 \rfloor$. There exists an orientation of its edges such that*

1. *at most one edge leaves every vertex;*

2. *the distance between every two special edges is at least R.*

Proof. We will prove the theorem by induction. In fact, we will prove that given a vertex s there exists an orientation with the required properties that satisfies two additional conditions: (A) s is a sink vertex, and (B) the distance from s to every special edge is at least R.

If G is a vertex, the statement holds trivially. If G is an edge (s, u), we orient (s, u) from u to s. If the graph is not connected, we orient each connected component separately.

Now we verify the induction step. Since G is l-path decomposable, there are two possibilities:

1. G has a cut vertex c;
2. G is the union of a graph H and a path P of length at least l, which do not have common vertices except for the endpoints of P.

Consider the first case. Then G is union of components C_i whose only common vertex is c. Assume without loss of generality that $s \in C_1$. We recursively orient edges in each C_i so that

- s is a sink in C_1;
- c is a sink in every C_i for $i \geq 2$.

Let us check that the obtained orientation satisfies the required conditions. Since c is a sink in all components $\{C_i\}_{i\geq 2}$, at most one edge leaves c. Every other vertex u belongs to exactly one component C_i, thus also at most one edge leaves u. The distance between every two special edges in one component is at least R. The distance between two special edges e_1 and e_2 in distinct components equals $d(e_1, c) + d(c, e_2)$. One of the summands is at least R by condition (B). The same argument shows that $d(s, e) \geq R$ for every special edge e. Finally, s is clearly a sink vertex in G (either it belongs only to C_1 and it is a sink in C_1; or $s = c$ and it is a sink in every C_i).

Now we consider the other case, G is a union of a subgraph H and a path P. Denote the endpoints of P by x and y. First, assume that s lies in H. Then recursively orient vertices in H. Pick an edge e in the middle of P at distance at least $\lfloor (l-1)/2 \rfloor > R$ from the endpoints of P. Let e be a special edge. Orient all edges between x and e toward x; all edges between y and e toward y. The distance from e to any other special edge or s is at least R. This orientation clearly satisfies all the required properties. Finally, assume that s lies on P. Without loss of generality, we assume that $d(x, s) \leq d(y, s)$. Note that the graph is a union of a subgraph $H' \equiv H \cup \pi(x, s)$ and a path $P' \equiv \pi(s, y)$. Since s lies in H', we can apply the previous argument to the union of H' and P'. Since $\lfloor (\lceil l/2 \rceil - 1)/2 \rfloor \geq R$, the distance between any two special edges is also greater than R. □

Proof of Theorem 3.16. First, we apply Lemma 3.18 (with $l = 20L + 5$). We get an orientation of edges such that the distance between every two special edges is greater then $5L$.

Then we add extra vertices to G: we attach a path of length L to every sink in G, and orient the edges of this path from the sink. We obtain a new graph $\tilde{G} = (\tilde{V}, \tilde{E})$. For every vertex u in G, there exists a unique directed path P_u in $\tilde{G}$ of length L that starts at u and in which all edges have the same orientation as in $\tilde{G}$.

Let $\alpha = p - 1 - \varepsilon p$. Define $f(i) = \min(i+1, L+1-i)^\alpha$. Note that for every integer k between 0 and $L+1$,

$$\sum_{i=0}^{k} f(i) = \Theta(k+1)^{\alpha+1}.$$

We now define a map φ from V to $(\mathbb{R}^{\tilde{V}}, \|\cdot\|_p)$ as follows, the v-coordinate of the image of u equals

$$\varphi_v(u) = \begin{cases} f(d(u,v))^{1/p}, & \text{if } v \in P_u; \\ 0, & \text{otherwise .} \end{cases}$$

Note that

$$\|\varphi(u)\|_p = \left(\sum_{v \in P_u} f(d(u,v))\right)^{1/p} = \left(\sum_{k=0}^{L} f(k)\right)^{1/p} = \Theta(L^{(\alpha+1)/p}) = \Theta(L^{1-\varepsilon}).$$

We will now prove a lemma, which bounds the distortion of the map φ for some pairs of vertices.

Lemma 3.19. *Consider two vertices u and v. If all edges on the path $\pi(u,v)$ are regular then*

$$c\rho(u,v) \le \|\varphi(u) - \varphi(v)\|_p \le CD\rho(u,v).$$

Here c and C are some absolute constants.

Proof. There is a vertex b in $\pi(u,v)$ such that all edges on $\pi(u,b)$ are directed from u to b; all edges on $\pi(v,b)$ are directed from v to b (b can coincide with u or v). Then P_u is a subpath of length L of the path $\pi(u,b) \to P_b$; and P_v is a subpath of length L of the path $\pi(v,b) \to P_b$. Denote $d_1 = \min(d(u,b), L+1)$, $d_2 = \min(d(v,b), L+1)$. Without loss of generality we assume that $d_1 \le d_2$. We now describe how the paths P_u and P_v intersect.

- The first d_1 vertices on P_u (which lie on the path $\pi(u,b)$) do not belong to P_v; the first d_2 vertices on P_v do not belong to P_u.
- The first $L+1-d_2$ vertices on P_b belong to both P_u and P_v.
- The next $d_2 - d_1$ vertices on P_b belong only to P_u.
- P_u and P_v contain no vertices other than listed above.

We have

$$\|\varphi(u)-\varphi(v)\|_p^p = \sum_{k=0}^{d_1-1} f(k)+\sum_{k=0}^{d_2-1} f(k)+\sum_{k=0}^{L-d_2} \left|f(k+d_1)^{1/p} - f(k+d_2)^{1/p}\right|^p + \sum_{k=L+1-d_2}^{L-d_1} f(k+d_1).$$

The first term equals $\Theta(d_1)^{(1-\varepsilon)p}$, the second term equals $\Theta(d_2)^{(1-\varepsilon)p}$. Since $f(L-k) = f(k)$, the last term equals

$$\sum_{k=0}^{d_2-d_1-1} f(k) = O(d_2 - d_1)^{(1-\varepsilon)p}.$$

Finally, let us estimate the third term.

$$\begin{aligned}
\sum_{k=0}^{L-d_2} |f(k+d_1)^{1/p} - f(k+d_2)^{1/p}|^p &\le 2\sum_{k=0}^{\infty} \left|(k+(d_2-d_1))^{\alpha/p} - k^{\alpha/p}\right|^p \\
&\le 2\sum_{k=0}^{d_2-d_1} (k+(d_2-d_1))^{\alpha} + 2\sum_{k=d_2-d_1+q}^{\infty} k^{\alpha}\left(\left(1+\frac{d_2-d_1}{k}\right)^{\alpha/p} - 1\right)^p \\
&\le O\left((2(d_2-d_1))^{\alpha+1} + \left(\frac{\alpha(d_2-d_1)}{p}\right)^p \sum_{k=d_2-d_1+1}^{\infty} k^{\alpha-p}\right) \\
&\le O\left((2(d_2-d_1))^{\alpha+1} + \left(\frac{\alpha(d_2-d_1)}{p}\right)^p \frac{(d_2-d_1)^{\alpha+1-p}}{p-\alpha-1}\right) \\
&= (d_2-d_1)^{\alpha+1} \times O\left(1+\frac{\alpha}{p(p-\alpha-1)^{1/p}}\right)^p.
\end{aligned}$$

Combining all our estimates, we get that (for some positive constants c_1 and C_1),

$$c_1 d_2^{(\alpha+1)/p} \le \|\varphi(u)-\varphi(v)\|_p \le C_2 d_2^{(\alpha+1)/p}\left(1+\frac{1}{(p-\alpha-1)^{1/p}}\right).$$

Taking into the account that $\min(d(u,v), L+1)/2 \le d_2 \le \min(d(u,v), L+1)$ we get the statement of the lemma. □

If there are no special edges in G, we are done. So we assume that there is at least one special edge in the graph.

Given a special edge e and a vertex u, let e_u be the endpoint of e that is further away from u (or an arbitrary endpoint if both endpoints are at the same distance from u). Notice that the shortest path between u and e_u always contains the edge e. In other words, the shortest path first visits the other endpoint of e and then goes to e_u along the edge e. Define a potential function $p_e(u)$ with respect to a special edge e as follows:

$$p_e(u) = 1 - \frac{\rho(u,u_e)}{L^{1-\varepsilon}} = \begin{cases} 1 - \left(\frac{d(u,e_u)}{L}\right)^{1-\varepsilon}, & \text{if } d(u,e_u) < L; \\ 0, & \text{otherwise.} \end{cases}$$

Note that $|p_e(u) - p_e(v)| \le (\rho(u,v)+1)/L^{1-\varepsilon}$. Since the distance between any two special edges is greater then $5L$, the potential of a given point u is positive with respect to at most one special edge.

Now we define the desired embedding of G into $(\mathbb{R}^{\tilde{V}}, \|\cdot\|_p)$.

$$\psi(u) = \begin{cases} \varphi(u) + p_e(u)\varphi(e_u), & \text{if there is a special edge } e \text{ s.t. } p_e(u) > 0; \\ \varphi(u), & \text{otherwise.} \end{cases}$$

We show that the paths P_u and P_{e_u} do not intersect, and thus the supports of $\varphi(u)$ and $\varphi(e_u)$ are disjoint. If P_u and P_{e_u} intersect, then there exists a path between u and P_{e_u} of length at most $2L$ whose edges are regular. Since the girth of the graph is at least $4L$, this path must be the shortest path. But the shortest path between u and e_u goes through the special edge e by the definition of e_u.

We get $\|\psi(u)\|_p = \Theta(L^{1-\varepsilon})$. Now we show that the distortion of ψ is $O(D)$. First we prove that if $d(u,v) > 3L$ then the supports of $\psi(u)$ and $\psi(v)$ do not intersect:

- Paths P_u and P_v do not intersect since $\text{length}(P_u) + \text{length}(P_v) = 2L < d(u,v)$. That is, supports of $\varphi(u)$ and $\varphi(u)$ do not intersect.
- Assume $p_e(v) > 0$ for some special edge e. Then $d(v, e_v) < L$ and $d(u, e_v) \ge d(u,v) - d(v,e_v) > 2L$. Thus paths P_u and P_{e_v} do not intersect. That is, supports of $\varphi(u)$ and $p_e(v)\varphi(e_v)$ do not intersect.
- Similarly, supports of $\varphi(v)$ and $p_e(u)\varphi(e_u)$ do not intersect.
- Assume $p_{e^1}(u) > 0$ and $p_{e^2}(v) > 0$ for special edges e^1 and e^2. Note that $e_1 \ne e_2$ as otherwise we would have $d(u,v) \le d(u, e^1_u) + d(v, e^2_v) + 1 \le 3L$. Then since the distance between every two special edges is at least $5L$, paths $P_{e^1_u}$ and $P_{e^2_v}$ do not intersect. That is, supports of $p_{e^1}(u)\varphi(e^1_u)$ and $p_{e^2}(v)\varphi(e^2_v)$ do not intersect.

Therefore,

$$\|\psi(u) - \psi(v)\|_p = \Theta(\|\psi(u)\|_p + \|\psi(v)\|_p) = \Theta(L^{1-\varepsilon}) = \Theta(\rho(u,v)).$$

Below we assume that $d(u,v) \le 3L$. Since the distance between any two special edges is greater than $5L$, the path $\pi(u,v)$ can contain at most one special edge. Moreover, if u is within distance L from a special edge e, and v is within distance L from a special edge e' then $e = e'$. Therefore, there exists a special edge e such that the distance from u to any other special edge is more than L, and the distance from v to any other special edge is more than L. Thus

$$\psi(u) = \varphi(u) + p_e(u)\varphi(e_u),$$
$$\psi(v) = \varphi(v) + p_e(v)\varphi(e_v).$$

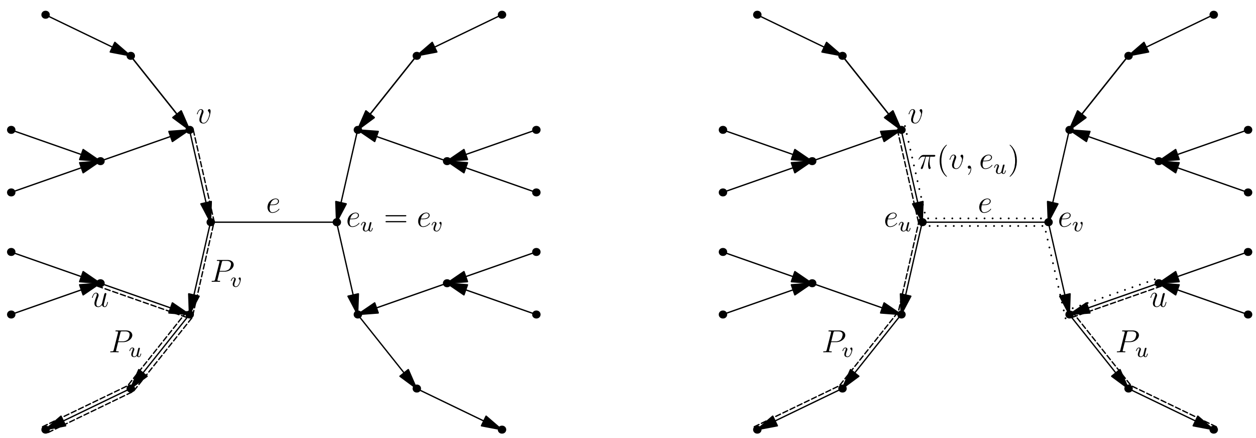

Figure 2: The figure shows the arrangement of vertices u, v, e_u, e_v and paths $\pi(u, e_u)$, $\pi(v, e_v)$ (shown by dotted lines), P_u and P_v (shown by dashed lines) in Case I and Case III.

I. Assume that all edges on $\pi(u, v)$ are regular edges and $e_u = e_v$. Then

$$\begin{aligned}\|\psi(u) - \psi(v)\|_p &= \left(\|\varphi(u) - \varphi(v)\|_p^p + |p_e(u) - p_e(v)|^p \|\psi(e_u)\|_p^p\right)^{1/p} \\ &= \Theta(\|\varphi(u) - \varphi(v)\|_p + \rho(u, v)).\end{aligned}$$

From Lemma 3.19, we get

$$c\rho(u, v) \le \|\psi(u) - \psi(v)\|_p \le CD\rho(u, v)$$

for some positive constants c and C.

II. Assume that all edges on $\pi(u, v)$ are regular edges but $e_u \ne e_v$. Both the shortest path from u to e_u and the shortest path from v to e_v contain the edge e, but these paths go along e in different directions. Thus their union is a simple path from u to v containing the edge e. Since it contains a special edge, it cannot be the shortest path (i.e. $\pi(u, v)$); and therefore $d(u, e_u) > L$ or $d(v, e_v) > L$. Then either $p_e(u) = 0$ or $p_e(v) = 0$. The analysis reduces to the previous case.

III. Assume that there is exactly one special edge on $\pi(u, v)$; and this edge is e. Then $e_u \ne e_v$. The paths P_u and P_v do not intersect; and the supports of vectors $\varphi(u) - p_e(v)\varphi(e_v)$ and $\varphi(v) - p_e(u)\varphi(e_u)$ are disjoint. Thus

$$\|\psi(u) - \psi(v)\|_p = \Theta\left(\|\varphi(u) - p_e(v)\varphi(e_v)\|_p + \|\varphi(v) - p_e(u)\varphi(e_u)\|_p\right).$$

Since all edges on $\pi(u, e_v)$ and $\pi(v, e_u)$ are regular, $\|\varphi(u) - \varphi(e_v)\|_p = \Theta(\rho(v, e_u))$ and $\|\varphi(v) - \varphi(e_u)\|_p = \Theta(\rho(v, e_u))$. We have

$$\begin{aligned}\|\psi(u) - \psi(v)\|_p &\le O\big(\|\varphi(u) - \varphi(e_v)\|_p + (1 - p_e(v))\|\varphi(e_v)\|_p \\ &\quad + \|\varphi(v) - \varphi(e_u)\|_p + (1 - p_e(u))\|\varphi(e_u)\|_p\big) \\ &= O\left(\rho(u, e_v) + \rho(v, e_u) + \frac{\rho(v, e_v) + \rho(u, e_u)}{L^{1-\varepsilon}} \times L^{1-\varepsilon}\right) \\ &\le O(D\rho(u, v)).\end{aligned}$$

On the other hand,

$$\begin{aligned}\|\psi(u)-\psi(v)\|_p &\geq \Omega\left(\|\varphi(u)\|_p - p_e(u)\|\varphi(e_u)\|_p + \|\varphi(v)\|_p - p_e(v)\|\varphi(e_v)\|_p\right)\\ &= \Omega(\|\psi(u)\|_p \times ((1-p_e(u)) + (1-p_e(v))))\\ &= \Omega(\rho(v,e_v)+\rho(u,e_u)) = \Omega(\rho(u,v)).\end{aligned}$$

IV. Finally, assume that there is a special edge $e' \neq e$ on $\pi(u,v)$. Since $d(u,v) < 3L$, we have $d(u,e'_u) < 3L$, $d(v,e'_v) < 3L$. Hence $d(u,e) > L$, $d(v,e) > L$. We can apply the reasoning from item III with $e = e'$ to get the desired bounds on the distortion. □

4 Acknowledgements

The authors thank Assaf Naor for his suggestions how to strengthen the result of Theorem 3.15.

References

[1] I. Abraham, M. Balakrishnan, F. Kuhn, D. Malkhi, V. Ramasubramanian, and K. Talwar. *Reconstructing Approximate Tree Metrics.* Proceedings of the twenty-sixth ACM symposium on Principles of distributed computing, pp. 43–52, 2007.

[2] I. Abraham, Y. Bartal, and O. Neiman. *Local embeddings of metric spaces.* Proceedings of the 39th ACM symposium on Theory of computing, pp. 631–640, 2007.

[3] S. Arora, B. Bollobás, L. Lovász, and I. Tourlakis. *Proving Integrality Gaps without Knowing the Linear Program.* Theory of Computing, vol. 2, pp. 19–51, 2006.

[4] S. Arora, J. Lee, and A. Naor. *Euclidean distortion and the sparsest cut.* Proceedings of the 37th ACM symposium on theory of computing, pp. 553–562, 2005.

[5] S. Arora, L. Lovász, I. Newman, Y. Rabani, Y. Rabinovich, and S. Vempala. *Local versus global properties of metric spaces.* Proceedings of the 17th ACM-SIAM symposium on Discrete algorithms, pp. 41–50, 2006.

[6] S. Arora, S. Rao, and U. Vazirani. *Expander flows, geometric embeddings and graph partitioning.* Proceedings of the 36th ACM symposium on theory of computing, pp. 222–231, 2004.

[7] Y. Aumann and Y. Rabani. *An $O(\log k)$ Approximate Min-cut Max-flow Theorem and Approximation Algorithm for Multi-Commodity Flows.* SIAM J. Comput., vol. 27, pp. 291-301, 1998.

[8] J. Bourgain. *On Lipschitz embedding of finite metric spaces in Hilbert space.* Israel J. Math., vol. 52, pp. 46-52, 1985.

[9] J. Bourgain. *The metrical interpretation of superreflexivity in Banach spaces.* Israel J. Math, vol. 56, pp. 222–230, 1986.

[10] G. Calinescu, H. Karloff, and Y. Rabani. *Approximation Algorithms for the 0–Extension Problem.* Proceedings of the 12th ACM-SIAM symposium on Discrete algorithms, pp. 8–16, 2001.

[11] H. Chan, M. Dinitz, and A. Gupta. *Spanners with Slack.* Proceedings of the 14th European symposium on algorithms, pp. 196–207, 2006.

[12] M. Charikar, K. Makarychev, and Y. Makarychev. *Integrality Gaps for Sherali–Adams Relaxations.* To appear in Proceedings of the 41st ACM symposium on theory of computing, 2009.

[13] W. F. de la Vega and C. Kenyon-Mathieu. *Linear Programming Relaxations of Maxcut.* Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms, pp. 53–61, 2007.

[14] J. Fakcharoenphol, S. Rao, and K. Talwar. *A tight bound on approximating arbitrary metrics by tree metrics.* Proceedings of the 35th ACM symposium on theory of computing, pp. 448–455, 2003.

[15] E. Helly. *Über Mengen konvexer Körper mit gemeinschaftlichen Punkten.* Jber. Deutsch. Math. Vereinig, vol. 32, pp. 175–176, 1923.

[16] N. Linial, E. London, and Yu. Rabinovich. *The geometry of graphs and some of its algorithmic applications.* Combinatorica, vol. 15, pp. 215–245, 1995.

[17] J. Lee, M. Mendel, and A. Naor. *Metric structures in L_1: Dimension, snowflakes, and average distortion,* European Journal of Combinatorics, vol. 26(8), pp. 1180–1190.

[18] L. Lovász and A. Schrijver. *Cones of matrices and set-functions and 0-1 optimization.* SIAM J. Optim., vol. 1, pp. 166-190, 1991.

[19] J. Matoušek. *On embedding expanders into ℓ_p spaces.* Israel J. Math., vol. 102, pp. 189–197, 1997.

[20] J. Matoušek. *On embedding trees into uniformly convex Banach spaces.* Israel J. Math., vol. 114, pp. 221–237, 1999.

[21] J. Matoušek. *Lectures on Discrete Geometry.* Graduate Texts in Mathematics, vol. 212, 2002.

[22] M. Mendel and A. Naor. *Euclidean quotients of finite metric spaces.* Advances in Mathematics, vol. 189, pp. 451–494, 2004.

[23] K. Menger. *Untersuchungen über allgemeine Metrik.* Mathematische Annalen, vol. 100, pp. 75-163, 1928.

[24] S. Rao. *Small distortion and volume preserving embeddings for planar and Euclidean metrics.* Proceedings of the 15th symposium on Computational geometry, pp. 300–306, 1999.

[25] H. D. Sherali and W. P. Adams. *A Hierarchy of Relaxations between the Continuous and Convex Hull Representations for Zero-One Programming Problems.* SIAM J. Discrete Math., vol. 3, pp. 411–430, 1990.

[26] G. Schoenebeck, L. Trevisan, and M. Tulsiani. *Tight Integrality Gaps for Lovász-Schrijver LP Relaxations of Vertex Cover and Max Cut.* Proceedings of the 39th symposium on ACM symposium on Theory of computing , pp. 302–310, 2007.

[27] N. Tomczak-Jaegermann. *Banach–Mazur distances and finite-dimensional operator ideals.* Pitman monographs and surveys in pure and applied mathematics, 1989.