Bandwidth and Low Dimensional Embedding

Yair Bartal * Douglas E. Carroll† Adam Meyerson‡ Ofer Neiman§

Abstract

We design an algorithm to embed graph metrics into ℓ_p with dimension and distortion both dependent only upon the bandwidth of the graph. In particular, we show that any graph of bandwidth k embeds with distortion polynomial in k into $\ell_p^{O(\log k)}$, $1 \leq p \leq \infty$. Prior to our result the only known embedding with distortion independent of n was into high dimensional ℓ_1 and had distortion exponential in k. Our low dimensional embedding is based on a general method for reducing the dimension of an ℓ_p embedding. This method requires that the embedding satisfy certain conditions, and the dimension is reduced to the intrinsic dimension of the point set, without substantially increasing the distortion. We observe that the family of graphs with bounded bandwidth are doubling, thus our main result can be viewed as a positive answer to a conjecture of Assouad [Ass83], limited to this family. We also study an extension to graphs of bounded tree-bandwidth.

*Institute of Computer Science, Hebrew University of Jerusalem, Israel. Email: `yair@cs.huji.ac.il`. The work was done in part while the author was at the Center for the Mathematics of Information, Caltech, and the Institute for Pure and Applied Mathematics, UCLA. Supported in part by a grant from the Israeli Science Foundation (195/02) and in part by a grant from the National Science Foundation (NSF CCF-065253).

†NetSeer.com. Email: `dougecarroll@yahoo.com`. Research done while a student at UCLA.

‡Department of Computer Science, University of California, Los Angeles. Email: `awm@cs.ucla.edu`. Research supported by the National Science Foundation under Grant No. CCF- 106540.

§Department of Computer Science, Ben Gurion University of the Negev, Beer-Sheva, Israel. Email: `neimano@cs.bgu.ac.il`. Supported by ISF grant No. (523/12) and by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement $n^\circ 303809$. Part of this work was done while the author was a student at the Hebrew university, and was visiting Caltech and UCLA.

1 Introduction

The problem of embedding graph metrics into normed spaces with low dimension and distortion has attracted much research attention (cf. [LLR95]). In this paper we study the family of graphs with bounded bandwidth. The bandwidth of an unweighted graph $G = (V, E)$ is the minimal k such that there exists an ordering of the vertices in which the end points of every edge are at most k apart. Let d_G be the shortest path metric on the graph G. Let (Y, ρ) be a metric space, we say that an embedding $f : V \to Y$ has distortion $D \geq 1$ if there exists a constant $c > 0$ such that for all $x, y \in V$,

$$d_G(x, y) \leq c\rho(f(x), f(y)) \leq D d_G(x, y) \ .$$

Our main result is the following.

Theorem 1.1. *For any choice of integers $p, k \geq 1$, there exist functions $d = d(k)$ and $D = D(k)$ such that any graph $G = (V, E)$ with bandwidth at most k can be embedded into ℓ_p^d with distortion D. In particular, we have: $D(k) = O(k^2)$ and $d(k) = O(\log^2 k)$. Alternatively, we also get: $D(k) = O(k^{2.001})$ and $d(k) = O(\log k)$.*

Our work is related to a conjecture of Assouad [Ass83]. The doubling constant of a metric space is the minimal α, such that any ball of radius r can be covered by α balls of half the radius. The doubling dimension of a metric space is defined as $\log_2 \alpha$. Assouad proved that for any metric (V, d), the "snow-flake" metric $(V, d^{1-\epsilon})$ embeds into Euclidean space with both distortion and dimension depending only on the doubling constant of (V, d) and ϵ. Assouad conjectured that this is possible even when $\epsilon = 0$, but this was disproved by [Sem96] (a quantitative bound was given by [GKL03]). It is also shown in [GKL03] that Assouad's conjecture holds for the family of doubling tree metrics. As the doubling constant of graphs with bandwidth k can be bounded by $O(k)$, one can view our result as providing a different family of doubling metrics for which Assouad's conjecture holds.

Graphs with low bandwidth play an important role in fast manipulation of matrices, in particular computing Gauss elimination and multiplication [CCDG82]. In his seminal paper Feige [Fei98] showed an approximation algorithm for computing the bandwidth with poly-logarithmic guarantee. The bandwidth of a graph also plays a role in certain biological settings, such as gene clustering problems [ZACS09].

There has been a great deal of previous work on embedding families of graphs into ℓ_p with bounded distortion (for example [CGN+03, GKL03, GNRS99, Rao99, CJLV08]). The problem of embedding graphs of bounded bandwidth has been first tackled by [CGM06]. They show that this family of graphs includes interesting instances which do not fall within any of the cases for which constant distortion embeddings are known. In their paper they show that bounded bandwidth graphs can be embedded into ℓ_1 [CGM06] with distortion independent of the number of vertices n. However, the distortion of their embedding was exponential in the bandwidth k. Also, the dimension of that embedding was dependent on the number of vertices (in fact polynomial in n). We improve the result of [CGM06] for graphs of bandwidth k in several ways: First, our embedding works for any ℓ_p space ($1 \leq p \leq \infty$) as a target space, not just ℓ_1. Second, the distortion obtained is polynomial in k; specifically: $O(k^{2+\theta})$. Finally, we show that the dimension can be independent of n as well, and as low as $O((\log k)/\theta)$ (for any $0 < \theta < 1$).

Note that the fact that a graph has bandwidth k can be viewed as providing an embedding into 1 dimension with expansion bounded by k, but without any control on the contraction. Our result means that by increasing the dimension to $O(\log k)$, one can get a bound not only on the expansion but also on the contraction of the embedding.

The low dimensionality of our embedding follows from a generalization we give for a result of [ABN08], who study embedding metric spaces in their intrinsic dimension. In [ABN08] it is shown that for any n point metric space, with doubling constant α, there exists an embedding into ℓ_p space with distortion $O(\log^{1+\theta} n)$ and dimension $O((\log \alpha)/\theta)$ (for any $0 < \theta < 1$). Here, we extend their method in a way that may be applicable for reducing the dimension of embeddings in other settings. We show sufficient conditions on an embedding of any metric space (V, d) into ℓ_p (possibly high dimensional) with distortion γ, allowing to reduce the dimension to $O((\log \alpha)/\theta)$ with distortion only $O(\gamma^{1+\theta})$.

Our embedding for graphs of bandwidth k is obtained as follows: we first provide an embedding with distortion $O(k^2)$ which satisfies the conditions of the dimension reduction theorem. Our final embedding follows from the fact that the doubling constant of graphs of bandwidth k is $O(k)$.

It is worth noting that our embedding provides bounds independent of n for all $1 \le p \le \infty$. This is unusual: most previous non-trivial results for embedding infinite graph classes into normed spaces with constant distortion (independent of n) have ℓ_1 as a target metric [CGN$^+$03, GNRS99] (and require high dimension). This is because of strong lower bounds indicating that trees have a distortion of $\Omega(\sqrt{\log \log n})$ [Mat99] and tree-width two graphs have a distortion of $\Omega(\sqrt{\log n})$ when embedded into ℓ_2 [NR02]. Since bandwidth k graphs do not include all trees, these lower bounds will not apply and we are able to embed into ℓ_2 with constant (independent of n) distortion. We observe that ℓ_2 is potentially a more natural and useful target metric, as many algorithms are tailored for Euclidean space.

We extend our study to graphs of bounded tree-bandwidth [CGM06] (see Definition 4.1 for precise definition). In their paper, [CGM06] showed an embedding of tree-bandwidth k graphs into ℓ_1 with distortion $O(k \cdot 2^k)$. While this family of graphs includes all trees and thus requires distortion at least $\Omega(\sqrt{\log \log n})$ when embedded into ℓ_2, we are still able to apply our techniques with an additional overhead related to the embeddability of trees. We show the following:

Theorem 1.2. *There is a randomized algorithm to embed tree-bandwidth k graphs into ℓ_p with expected distortion $O(k^3 \log k + k\rho)$ where ρ is the distortion required for embedding a tree into ℓ_p.*

Thus for $p = 2$ the distortion is $O(\mathrm{poly}(k)\sqrt{\log \log n})$ and into ℓ_1 the distortion is polynomial in k. Moreover, when the graph has bounded doubling dimension we can apply our dimension reduction technique to achieve expected distortion and dimension depending solely on the doubling dimension and on k, utilizing the embedding of [GKL03] for trees with bounded doubling dimension.

In general there has been a great deal of work on finding low-distortion embeddings. These embeddings have a wide range of applications in approximation algorithms, and in most cases low dimension is also desirable (for example improving the running time). Our work makes further progress towards achieving low-distortion results with dimension reduced to the intrinsic dimension. In particular, our embeddings imply better bounds in applications such as nearest neighbor search, distance labeling and clustering.

1.1 Summary of Techniques

The result of [ABN08] includes the design of a specific embedding technique (locally padded probabilistic partitions), combined with the careful application of the Lovasz Local Lemma to show that it is possible to randomly merge the coordinates of this embedding in such a way that there is a non-zero probability that no distance is contracted. This can then be combined with constructive versions of the local lemma [Bec91, MT10] to deterministically produce a low-dimensional embedding with no contraction.

We decouple the embedding technique of [ABN08] from the local lemma, showing that any embedding technique which satisfies certain locality properties as well as having a single coordinate which lower bounds each particular distance can be applied in this way. Given any metric space (V, d) with doubling constant α, we give sufficient conditions to reduce the dimension of an embedding of V into ℓ_p with distortion γ to have dimension $O((\log \alpha \log \gamma)/\log(1/\epsilon))$ and distortion $O(\gamma/\epsilon)$ where $\gamma^{-1} < \epsilon < 1$. This approach allows some modularity in defining an embedding – if we are given a low distortion embedding (potentially much lower distortion than $\log n$ for some source metrics) which satisfies the locality properties then we can maintain the low distortion while obtaining low dimension as well.

In order to demonstrate the power of this approach, we apply it to the problem of embedding bounded bandwidth graphs into ℓ_p. We first need to define a low distortion embedding. The embedding of [CGM06] is not useful for our purpose as it does not satisfy the necessary properties (in particular the single coordinate lower bound on distances fails) and because its distortion is undesirably high (exponential in bandwidth). Instead, we define a new embedding. The basis for our embedding is the standard scale based approach [Rao99] using probabilistic partitions of [GKL03, ABN08] as a black box. The problem is that when using this approach we obtain an expansion factor of 1 at each scale of the embedding. The number of scales is logarithmic in the graph diameter, giving us a total expansion of $\Theta(\log n)$. The key innovation of our bandwidth embedding is showing that the number of scales can be reduced to $O(k)$.

Of course, for any scale there may be some point pair whose distance is at that scale (there are n^2 point pairs and only $\log n$ scales after all). We cannot simply remove some scales and expect our distortion to be reasonable. Instead, we compute a set of *active scales* for each graph vertex; these are the scales that represent distances to other points which are nearby in the optimum bandwidth ordering of the graph. We will reduce coordinate values to zero for vertices which do not consider the coordinate's scale to be active. Each vertex has only $O(k)$ active scales; the issue now is that different vertices have different scales and if two adjacent vertices have different active scales we might potentially introduce large expansion. In addition, we need to show that the critical coordinates which maintain the lower bound of $d(x, y)$ (thus preventing contraction) are active at one of the two points (x or y). Instead of applying active coordinates directly, we allow coordinates to decline gracefully by upper-bounding them by the distance to the nearest point where they are inactive, then use the bandwidth ordering to prove that the critical coordinates for preventing contraction are not only active where they need to be, but have not declined by too much to be useful.

A careful analysis of this construction shows that we can obtain distortion of $O(k^2)$. We also show that our modified embedding still possesses the locality properties. Thus we can apply our dimension reduction technique to get dimension $O((\log k)/\theta)$ while maintaining the distortion bound up to a factor $O(k^\theta)$.

2 Embedding in the Doubling Dimension

2.1 Preliminaries and Definitions

Definition 2.1. The **doubling constant** of a metric space (V, d) is the minimal integer α such that for any $r > 0$ and $x \in V$, the ball $B(x, 2r)$ can be covered by α balls of radius r. The **doubling dimension** or **intrinsic dimension**, denoted by $\dim(V)$, is defined as $\log \alpha$.

Suppose we are given a metric space (V, d) along with a randomly selected mapping $\phi : V \to \Re^D$ for some dimension D. For $1 \leq c \leq D$ we denote by $\phi_c(x)$ the c'th coordinate of $\phi(x)$ and thus we have $\phi_c : V \to \Re$. We may assume w.l.o.g that all coordinates of all points in the range of this mapping are non-negative.

Definition 2.2. The mapping ϕ is **single-coordinate** (ϵ, β) **lower-bounded** if for every pair of points $x, y \in V$ there is some coordinate c such that $|\phi_c(x) - \phi_c(y)| \geq \beta d(x, y)$ with probability at least $1 - \epsilon$.

In the metric embedding literature, we often speak of an embedding having contraction β. For ℓ_1 embedding, this means there is a set of coordinates whose sum is lower-bounded by $\beta d(x, y)$. The single-coordinate (ϵ, β) lower-bounded condition is stronger than contraction β, although for ℓ_p norms with large values of p (i.e. as p tends towards infinity) it becomes equivalent.

Definition 2.3. Given a mapping ϕ, the ℓ_1 **expansion** of ϕ is $\delta = \max_{x,y} \frac{\|\phi(x)-\phi(y)\|_1}{d(x,y)}$.

We observe that the expansion of ϕ when viewed as an ℓ_p embedding for $p > 1$ will be at most the ℓ_1 expansion. On the other hand, the single-coordinate (ϵ, β) lower-bound condition will still imply that the embedding has contraction β (for any pair of points with $1 - \epsilon$ probability).

Definition 2.4. A mapping ϕ has the **local property** if for every coordinate c we can assign a *scale* s_c which is a power of two such that the following conditions hold:

1. For every $x, y \in V$ with $d(x, y) > s_c$ we have either $\phi_c(x) = 0$ or $\phi_c(y) = 0$.

2. For every $x, y \in V$, if there is a single-coordinate lower-bound for x, y, it has scale $\Omega(d(x, y)) < s_c < O(d(x, y))$.

We observe that a mapping ϕ can be viewed as an embedding of (V, d) into normed space. Provided that the mapping is single-coordinate (ϵ, β) lower-bounded, we can eliminate contraction by repeatedly (and independently) selecting such mappings many times over and weighting the results by the number of selections, then multiplying all coordinates by $\frac{1}{\beta}$. This provides an embedding into ℓ_1 with distortion upper-bounded by $\frac{\delta}{\beta}$; note that this embedding can also be viewed as into ℓ_p for any $p > 1$ and in fact will have only lower distortion (the single-coordinate lower-bound condition still guarantees non-contraction).

2.2 Low Dimensional Embedding

An embedding ϕ maps (V, d) to potentially high dimensional space, and we are interested in *reducing the dimension* of such an embedding to resemble the doubling dimension of (V, d) without increasing the distortion. While for general ϕ such a result would imply dimension reduction for ℓ_1 (which is impossible in general [BC03]), the additional constraints that ϕ be single-coordinate lower-bounded and local will enable us to reduce the dimension. In Appendix A we prove the following generalization of [ABN08].

Theorem 2.5. *Suppose we are given a metric space (V,d) with doubling constant α and a mapping $\phi:(V,d)\to\Re^D$ where ϕ is single-coordinate (ϵ,β) lower-bounded, local, and has ℓ_1 expansion at most δ for some $\beta/\delta\le\epsilon\le 1/8$. Then for any $1\le p\le\infty$ we can produce in polynomial time an embedding $\tilde{\phi}:(V,d)\to\ell_p^m$ with distortion at most $O(\delta/(\epsilon\beta))$, where $m=O\left(\frac{\log\alpha\log(\delta/\beta)}{\log(1/\epsilon)}\right)$.*

Next we construct an embedding with the local property, which will serve as a basis embedding in Section 3. Recall that a partition $P=\{C_1,\ldots,C_n\}$ of an n-point metric space (V,d) is a pairwise disjoint collection of clusters (possibly some clusters are empty) which covers V, and $P(x)$ denotes the cluster containing $x\in V$. W.l.o.g we may assume that $\min_{x\neq y\in X}\{d(x,y)\}\ge 1$. The following lemma is a generalization of a lemma of [GKL03] and was proven in [ABN08].

Lemma 2.6. *For any metric space (V,d) with doubling constant α, any $0<\Lambda<\mathrm{diam}(V)$ and $0<\epsilon\le 1/2$ there exists a distribution $\hat{\mathcal{P}}$ over a set of partitions $\mathcal{P}$ such that the following conditions hold.*

- *For any $1\le j\le n$, $\mathrm{diam}(C_j)\le\Lambda$.*
- *For any $x\in V$, $\Pr_{P\sim\hat{\mathcal{P}}}[B(x,\epsilon\Lambda/(64\log\alpha))\nsubseteq P(x)]\le\epsilon$.*

For every scale $s\in I=\{2^i\mid -1\le i<\log(\mathrm{diam}(V)), i\in\mathbb{Z}\}$ let $P_s=\{C_1(s),\ldots,C_n(s)\}$ be a random partition sampled from $\hat{\mathcal{P}}$ with $\Lambda=s$, and let $c_1(s),\ldots c_n(s)$ be n coordinates that are assigned to the scale s. The random mapping is defined as

$$\phi_{c_j(s)}(x)=\begin{cases} d(x,V\setminus C_j(s)) & x\in C_j(s)\\ 0 & \text{otherwise}\end{cases}\tag{1}$$

and

$$\phi=\bigoplus_{s\in I,1\le j\le n}\phi_{c_j(s)}\tag{2}$$

Claim 2.7. *For any $0<\epsilon\le 1/2$ the mapping ϕ is single-coordinate $(\epsilon,\epsilon/(128\log\alpha))$ lower-bounded, and its ℓ_1 expansion is at most $O(\log(\mathrm{diam}(V)))$.*

Proof. For any $x,y\in V$ let s be a power of two such that $s<d(x,y)\le 2s$, then in the coordinates assigned to scale s, the first property of Lemma 2.6 suggests that it must be that x,y fall into different clusters of the partition associated with the coordinates. Let j be such that $x\in C_j$, it follows that with probability $1-\epsilon$, $\phi_{c_j(s)}(x)\ge\epsilon s/(64\log\alpha)\ge\epsilon d(x,y)/(128\log\alpha)$ and that with probability 1, $\phi_{c_j(s)}(y)=0$.

To see that the ℓ_1 expansion is at most $2(\log(\mathrm{diam}(V))+2)$, note that the triangle inequality implies that $|\phi_{c_j(s)}(x)-\phi_{c_j(s)}(y)|\le d(x,y)$ for any $x,y\in V$ and $j\in[n]$, and since $\phi_{c_j(s)}(x)$ is non-zero for a single $j\in[n]$ it follows that for any $s\in I$

$$\Sigma_{1\le j\le n}|\phi_{c_j(s)}(x)-\phi_{c_j(s)}(y)|\le 2d(x,y),\tag{3}$$

and hence

$$\Sigma_{s\in I,1\le j\le n}|\phi_{c_j(s)}(x)-\phi_{c_j(s)}(y)|\le\Sigma_{s\in I}2d(x,y)=2(\log(\mathrm{diam}(V))+2)d(x,y)\ .$$

□

Claim 2.8. *The mapping ϕ has the local property.*

Proof. The first local property is immediate by the first property of Lemma 2.6 and by (1). The second local property follows from the choice of s in the proof of Claim 2.7. □

3 Low Distortion ℓ_p-embeddings of Low Bandwidth Graphs

3.1 Preliminaries and Definitions

Definition 3.1. Given graph $G = (V, E)$ and linear ordering $f : V \to \{1, 2, ..., |V|\}$ the *bandwidth* of f is $max\{|f(v) - f(w)| \mid (v, w) \in E\}$. The *bandwidth* of G is the minimum bandwidth over all linear orderings f. Given an optimal bandwidth ordering f, the index of u is simply $f(u)$.

Definition 3.2. Define $\lambda(x, y) = |f(x) - f(y)|$ which is the distance between x, y in the bandwidth ordering f of G.

In what follows we are given a graph G of bandwidth k, the metric space associated with G is the usual shortest-path metric, and we assume we are given the optimal ordering f obtaining this bandwidth. This ordering is computable in time exponential in k, and since our embedding only improves upon previous work (for example Bourgain [Bou85]) when k is quite small, it may be reasonable to assume that the ordering is given. In general computing the best bandwidth ordering is NP-Hard, and the best approximations are poly-logarithmic in n [Fei98].

Claim 3.3. *Let G be a graph of bandwidth k. Then there exists an ordering where for any $x, y \in G$, $\lambda(x, y) \leq k \cdot d(x, y)$.*

Proof. Assume $d(x, y) = r$, and let $P_{xy} = (x = v_0, v_1, \ldots, v_r = y)$ be a shortest path in G connecting x and y, then by the triangle inequality $\lambda(x, y) \leq \Sigma_{i=1}^{r} \lambda(v_{i-1}, v_i)$. By the definition of bandwidth for all $1 \leq i \leq r$, $\lambda(v_{i-1}, v_i) \leq k$, hence the claim follows. □

Proposition 3.4. *If $G = (V, E)$ has bandwidth k, then the doubling constant α of G is at most $4k + 1$.*[1]

Proof. Consider the ball of radius $2r$ about some point $x \in V$. We must show that this ball can be covered by at most $4k + 1$ balls of radius r.

Consider any integer $0 < a \leq r$. Let Y_a be the set of points y such that $d(x, y) = a$; similarly let Y_{a+r} be the set of points y such that $d(x, y) = a + r$. We claim that the set of balls of radius r centered at points in $\{x\} \cup Y_a \cup Y_{a+r}$ covers the ball of radius $2r$ around x. In particular, consider any point z in this ball. If $d(x, z) \leq r$ then $z \in B(x, r)$. If $a \leq d(x, z) < a + r$ then there is some shortest path from x to z of length $d(x, z)$ which must include a point y with $d(x, y) = a$ and $d(y, z) = d(x, z) - a < r$. It follows that $z \in B(y, r)$ and that $y \in Y_a$. If $a + r \leq d(x, z) < 2r$ then again there is a shortest path from x to z of length $d(x, z)$ which must include a point y with $d(x, y) = a + r$ and $d(y, z) = d(x, z) - a - r < r$. It follows that $z \in B(y, r)$ and $y \in Y_{a+r}$.

Now consider the various sets $\{x\} \cup Y_a \cup Y_{a+r}$ as we allow a to range from 1 to r. With the exception of x, these sets are disjoint for distinct values of r. So every point in $B(x, 2r)$ other than x appears exactly once. It follows that there must be some choice of a such that the size of this set is only $1 + \frac{1}{r}|B(x, 2r)|$. Since the graph G has bandwidth k, it follows that any pair of adjacent nodes are within k of each other in the bandwidth ordering. So all points in $B(x, 2r)$ are within $2rk$ of x in the ordering, and thus there are at most $4rk$ such points. From this it follows that we need only $4k + 1$ balls to cover $B(x, 2r)$. □

[1] A somewhat simpler argument could be applied to give an $O(k)$ bound on the doubling constant, which would suffice for our application. The argument presented here seems to give a better estimate on the constant.

The remainder of this section will be devoted to proving our main theorem, that graphs of bounded bandwidth embed into ℓ_p with low dimension and distortion.

Theorem 3.5. *Let G be a graph with bandwidth k and let $0 < \theta < 1$, then for any $p \geq 1$, there exists an embedding of G into ℓ_p with distortion $O(k^{2+\theta})$ and dimension $O((\log k)/\theta)$.*

3.2 Proof of Theorem 3.5

Consider the mapping ϕ defined in (2). By Claim 2.7 combined with Theorem 2.5 (noting that for unweighted graphs we get $O(\log n)$ ℓ_1 expansion) we can transform it into an embedding of a graph with bandwidth k into any ℓ_p space of dimension $O((\log k)/\theta)$ with distortion $O(\log^{1+\theta} n)$ for any $0 < \theta \leq 1$. Our main innovation is to reduce the number of scales effecting each of the points, thereby reducing the overall distortion to $O(k^2)$.

Let G be a graph with bandwidth k and f be the optimal ordering obtaining this bandwidth. Let $\alpha \leq 4k + 1$ be the doubling constant of G. For each scale s, we will say that scale s is *active* at point x if there exists a y such that $\lambda(x, y) \leq k$ and $s/8 \leq d(x, y) \leq 4s$. We define $h_s(x)$ to be the distance from x to the nearest point z for which s is not active (note that $h_s(x) = 0$ if s is not active at x). We then define a mapping $\hat{\phi}$ as follows:

$$\hat{\phi}_c(x) = \min(\phi_c(x), h_{s_c}(x))$$

We add an extra coordinate f which is the location of the points in the bandwidth ordering. We will claim that this $\hat{\phi}$ is single-coordinate $\frac{1}{k}$ lower-bounded for all point pairs x, y with $|f(x)-f(y)| \leq \frac{1}{4}d(x, y)$ and local, and that it has ℓ_1 expansion bounded by $O(k)$. This will allow us to apply Theorem 2.5 without the f coordinate, then add in the f coordinate to get our final embedding.

Lemma 3.6. *The mapping $\hat{\phi}$ has ℓ_1 expansion at most $O(k)$.*

Proof. Consider any pair of points x, y. We observe that the total number of scales which are *active* for these two points is at most $O(k)$. In total there are at most $O(k)$ non-zero coordinates for these two points. So the ℓ_1 expansion expression has only $O(k)$ non-zero terms. Let c be a non-zero coordinate. Because each coordinate produces expansion of at most 1 in ϕ (see proof of Claim 2.7), we have:

$$\phi_c(x) - \phi_c(y) \leq d(x, y)$$

If $\hat{\phi}_c(y) = \phi_c(y)$ then since $\hat{\phi}_c(x) \leq \phi_c(x)$, we can write:

$$\hat{\phi}_c(x) - \hat{\phi}_c(y) \leq \phi_c(x) - \phi_c(y) \leq d(x, y)$$

On the other hand, suppose that $\hat{\phi}_c(y) = h_s(y)$ where $s = s_c$. Then there is some z where scale s is inactive, such that $h_s(y) = d(y, z)$. Now

$$\hat{\phi}_c(x) - \hat{\phi}_c(y) \leq h_s(x) - h_s(y) \leq d(x, z) - d(y, z) \leq d(x, y)$$

From this we conclude that each non-zero coordinate produces expansion at most $O(1)$, and when we total this over $O(k)$ non-zero coordinates we get total ℓ_1 expansion at most $O(k)$. □

We note that adding the coordinate f does not increase the expansion by much. In particular, for any point pair x, y we have $|f(x) - f(y)| \leq kd(x, y)$ by Claim 3.3. So the extra coordinate increases expansion by at most an additive k.

The tricky part is proving that the mapping is single-coordinate $(\epsilon, \frac{1}{k})$ lower-bounded. Given some pair of points x, y, one might imagine that the critical coordinates were deemed *inactive* for x and y, and thus the single-coordinate lower-bound will no longer hold. We will prove that this is not the case.

Lemma 3.7. *For any $k^{-1/2} \leq \epsilon \leq 1/2$ the mapping $\hat{\phi}$ is single-coordinate $(\epsilon, \Omega(\frac{1}{k}))$ lower-bounded for any pair of points x, y with $|f(x) - f(y)| \leq \frac{1}{4}d(x, y)$.*

Proof. Consider any pair of points x, y with $|f(x) - f(y)| \leq \frac{1}{4}d(x, y)$. Let $x' \in B(x, \frac{d(x,y)}{8k})$ and $y' \in B(y, \frac{d(x,y)}{8k})$. Let s be the scale such that $d(x, y)/2 \leq s < d(x, y)$. We will show that scale s must be *active* at x' or at y'. But this holds for *any pair* of points x', y' from the appropriate balls around x, y. It follows that for one of these two balls it must be the case that scale s is active at *all points in the ball*. Suppose without loss of generality that this is $B(x, \frac{d(x,y)}{8k})$. Then since all points in this ball have scale s active, we conclude that $h_s(x) \geq \frac{d(x,y)}{8k}$. By Claim 2.7 and the local property of ϕ, there is a coordinate c assigned to scale s, which with $1 - \epsilon$ probability, has $\phi_c(x) \geq \Omega(\frac{\epsilon}{\log \alpha})d(x, y)$ and by the first local property of ϕ also $\phi_c(y) = 0$. If this event occurs, then since $\Omega\left(\frac{\epsilon}{\log \alpha}\right) \geq \Omega\left(\frac{k^{-1/2}}{2 \log k}\right) \geq \Omega(1/k)$, we get that $\hat{\phi}_c(x) \geq d(x, y) \min(\Omega(\frac{\epsilon}{\log \alpha}), \frac{1}{8k}) \geq \Omega(\frac{d(x,y)}{k})$, and of course $\hat{\phi}_c(y) = 0$. We conclude that x, y are $(\epsilon, \Omega(1/k))$ lower bounded.

In the remainder of proof we show that indeed scale s must be active at either x' or y'. Since $d(x, x') \leq d(x, y)/(8k)$ and $d(y, y') \leq d(x, y)/(8k)$ it follows that $d(x', y') \geq d(x, y)(1 - \frac{1}{4k}) \geq \frac{3}{4}d(x, y)$. On the other hand, $|f(x) - f(x')| \leq \frac{d(x,y)}{8}$ and similarly for $|f(y) - f(y')|$ from which we can conclude that $|f(x') - f(y')| \leq \frac{1}{2}d(x, y)$. Now consider a fixed shortest path from x' to y'. Assume without loss of generality that $f(x') < f(y')$. We define two special points along this path as follows:

- $\tilde{x}$ is the first point on the path from x' to y' such that for all points z subsequent to or equal to $\tilde{x}$ on the path, we have $f(z) \geq f(x')$.
- $\tilde{y}$ is the first point on the path from $\tilde{x}$ to y' with $f(\tilde{y}) \geq f(y')$

These points will be auxiliary points showing that scale s is active at either x' or y'. For instance to show that scale s is active at x' it is enough to show that $\lambda(x', \tilde{x}) \leq k$ and that $s/8 \leq d(x', \tilde{x}) \leq 4s$. We observe that because any pair of consecutive vertices on a path are at most k apart in the bandwidth ordering, it must be that $|f(x') - f(\tilde{x})| \leq k$ and $|f(y') - f(\tilde{y})| \leq k$ (because the point x'' just before $\tilde{x}$ on the path has $f(x'') \leq f(x')$, and similarly for y). Note that for every point z on the path from $\tilde{x}$ to $\tilde{y}$, the value of $f(z)$ is a unique point between $f(x')$ and $f(y')$. We conclude that $d(\tilde{x}, \tilde{y}) \leq |f(x') - f(y')| \leq \frac{1}{2}d(x, y)$. Since these points are on the shortest path, we know that $d(x', y') = d(x', \tilde{x}) + d(\tilde{x}, \tilde{y}) + d(\tilde{y}, y')$. It follows that either $d(x', \tilde{x}) \geq \frac{1}{8}d(x, y)$ or $d(\tilde{y}, y') \geq \frac{1}{8}d(x, y)$.

On the other hand, it is not hard to see that $d(x', \tilde{x}) \leq d(x', y') \leq d(x, y) + \frac{1}{4k}d(x, y) \leq 2d(x, y)$ and similarly for $d(y', \tilde{y})$. We conclude that indeed scale s must be active for one of x', y'. □

Lemma 3.8. *The mapping $\hat{\phi}$ is local.*

Proof. The first condition follows immediately from the fact that ϕ is local and $\hat{\phi}_c(x) \leq \phi_c(x)$ for all c and x. The second condition follow from Lemma 3.7. ☐

We now combine the lemmas and apply Theorem 2.5 to $\hat{\phi}$. This guarantees bounded contraction for point pairs with $|f(x) - f(y)| \leq \frac{1}{4}d(x,y)$. We add the single additional coordinate f, and this guarantees bounded contraction for points with $|f(x) - f(y)| \geq \frac{1}{4}d(x,y)$. Choosing for any $0 < \theta < 1$, $\epsilon = k^{-\theta}$ will give distortion $O(k^{2+\theta})$ and dimension $O((\log k)/\theta)$.

4 Tree Bandwidth

4.1 Definitions and Preliminaries

We will give an embedding of a graph of low tree-bandwidth [CGM06] into ℓ_p. The distortion will be polynomial in k, with a multiplicative $O(\sqrt{\log\log n})$ term for $p > 1$ [Bou86]. This improves upon the result of [CGM06] by reducing the distortion and extending to ℓ_p. We first define tree bandwidth and valid partitionings:

Definition 4.1. [[CGM06]] Given a graph $G = (V, E)$, we say that it has **tree-bandwidth** k if there is a rooted tree $T = (I, F)$ and a collection of sets $\{X_i \subset V | i \in I\}$ such that: $\forall i, |X_i| \leq k$, $V = \bigcup X_i$, the X_i are disjoint, $\forall (u,v) \in E$, u and v lie in the same set X_i or $u \in X_i$ and $v \in X_j$ and $(i,j) \in F$, and if i has parent $p(i)$ in T, then $\forall v \in X_i, \exists u \in X_{p(i)}$ such that $d(u,v) \leq k$.

Definition 4.2. Given a metric space (V, d), a partitioning C_s at scale s is **valid** if for any pair of points $x, y \in V$ with $d(x,y) > s$, C_j partitions x, y into separate clusters, and for any pair of points $x, y \in V$ with $d(x,y) \leq \frac{s}{|V|}$, C_j places x, y in the same cluster.

Definition 4.3. Given a metric space (V, d), a partitioning C_s at scale s is **almost-valid** if for any pair of points $x, y \in V$ with $d(x,y) > 2s$, C_j partitions x, y into separate clusters, and for any pair of points $x, y \in V$ with $d(x,y) \leq \frac{s}{2|V|}$, C_j places x, y in the same cluster.

A family of partitions contains a partition for each scale s which is a power of two, up to the diameter of (V, d). We will say a family of partitions is valid (almost-valid) if every partition in the family is valid (almost-valid). We first show that it is always possible to construct a valid partition.

Theorem 4.4. *A valid C_s can be constructed for any metric (V, d) and scale s.*

Proof. To build the clusters at scale s connect all pairs of points x, y where $d(x,y) \leq \frac{s}{|V|}$. Let clusters correspond to the resulting connected components.

Each connected component has diameter at most $|V|\frac{s}{|V|} = s$. Furthermore, distinct clusters are separated by distance $> \frac{s}{|V|}$. ☐

We will now define an embedding technique based upon almost-valid partition families. When we embed tree-bandwidth, we apply this technique separately to each of the node sets in the tree bandwidth decomposition.

Theorem 4.5. *Suppose we are given an almost-valid family of partitions. We define a coordinate for each cluster of C_s, such that this coordinate has value $4s$ for points in the cluster and 0 for other points. After creating such a coordinate for each cluster and scale s, the resulting mapping of (V, d) into high dimensional space does not contract any distance for any ℓ_p norm, and does not expand any distance by more than $32|V|$.*

Proof. Consider some x, y with $s < d(x,y) \leq 2s$.

We observe that x, y will be in the same cluster for $C_{4s|V|}$ since $d(x,y) \leq \frac{4s|V|}{2|V|}$. Thus the embedded distance between x and y will be at most $2\Sigma_{m<4s|V|}4m \leq 32s|V| \leq 32|V|d(x,y)$, which bounds the expansion (the sum is over m which are powers of two and thus is geometric).

On the other hand, x, y will be partitioned into different clusters by $C_{s/2}$ (and C_m for $m \leq s/2$) since $d(x,y) > 2\frac{s}{2}$. Thus the embedded distance between x and y will be at least $4\frac{s}{2} \geq 2s \geq d(x,y)$, implying no contraction. □

The following theorem is our main result of this section, its proof will be given below in Section 4.2, using almost-valid families of partitions to produce a low-distortion embedding of a graph of tree-bandwidth k into ℓ_p.

Theorem 4.6. *There is a randomized algorithm to embed tree-bandwidth k graphs into ℓ_p with expected distortion $O(k^3 \log k + k\rho)$ where ρ is the distortion for embedding the tree decomposition into ℓ_p.*

In the case of ℓ_1, there is a simple embedding of a tree with $\rho = 1$. For ℓ_2, the bound of [Bou86] ensures $\rho = O(\sqrt{\log \log n})$. We can also apply Theorem 2.5 to reduce the dimension of the embedding of Theorem 4.6. To do this we need to bound the dimension in which the tree can be embedded.

Lemma 4.7. *Let G be a graph with tree bandwidth k, and let α be the doubling constant of G, then the doubling dimension of the decomposition tree T for G is $\log \alpha_T = O((\log \alpha)(\log k))$.*

Proof. Let $T = (I, F)$ be the tree decomposition of G. Assume that T has doubling constant α_T, then there must be $r > 0$ and α_T nodes $s_1, \ldots, s_{\alpha_T} \in I$ such that $d(s_i, s_j) > r$ for all $i \neq j$ but there is $y \in I$ with $s_i \in B(y, 2r)$ for all i (otherwise the doubling constant of T would be less than α_T). Let z be the least common ancestor of $s_1, \ldots, s_{\alpha_T}$, since T is a tree it must be that $d_T(z, s_i) \leq 2r$. By the definition of tree bandwidth, for every $u \in X_{s_i}$ there is a node $v \in X_z$ such that $d_G(u,v) \leq 2rk$, and since X_z contains at most k points there must be a node $w \in X_z$ which is the ancestor for at least α_T/k nodes $u_i \in X_{s_i}$ for different i. Consider two such nodes $u_i \in X_{s_i}$ and $u_j \in X_{s_j}$, $i \neq j$, then since $d_T(s_i, s_j) > r$ it must be that $d_G(u_i, u_j) > r$ for any (it can be checked that tree distances lower bound the graph distances), and also for all i: $u_i \in B(w, 2rk)$. This implies that in the graph G there are α_T/k points which are r apart but all within a ball of radius $2rk$. Since the graph G has doubling constant α there can be at most $\alpha^{\lceil \log(2k) \rceil}$ such points, which yields $\alpha_T/k \leq \alpha^{2\log(2k)}$, and hence: $\log \alpha_T \leq O(\log \alpha \log k)$. □

It follows that we can use an embedding for the decomposition tree T of G where the distortion and dimension are function of the doubling dimension of T, and therefore function of α and k. In particular, we can use the tree embedding of [GKL03].

Corollary 4.8. *Suppose that we are given a tree-bandwidth k graph along with its tree decomposition. Let the doubling constant of this graph be α. Let α_T be the doubling constant of T, given by Lemma 4.7. Further, suppose that there exists an embedding of the tree decomposition into $d(\alpha_T)$ dimensional ℓ_p with distortion $\rho(\alpha_T)$. Then for any $0 < \theta < 1$ there is an embedding of the graph into ℓ_p with expected distortion $O(k^{3+\theta} \log k + k\rho(\alpha_T))$ and dimension $O((\log \alpha)/\theta + d(\alpha_T))$.*

4.2 Proof of Theorem 4.6

Proof overview: We construct an almost-valid partition at each scale for each of the sets of nodes X_i in the bandwidth decomposition. We start with the root and construct a valid partition for each scale s. We also define a timer τ_s which is initialized to a value selected uniformly from $[0, \frac{s}{4k^2} - 1]$. We now traverse the tree, moving downwards from the root. For each set of nodes X_i, we copy the values of τ_s for each scale from the parent set $X_{p(i)}$. We then increment the τ_s which correspond to a scale such that there are two points $x, y \in X_i$ with $\frac{s}{2k} \leq d(x, y) \leq 2s$. If the timer has $\tau_s < \frac{s}{4k^2}$ then we produce partition C_s by using the same clusters which we defined for the parent, and adding each point $x \in X_i$ to the cluster containing its nearest neighbor in $X_{p(i)}$. Otherwise we have $\tau_s = \frac{s}{4k^2}$ and we construct a valid partition C_s from scratch using Theorem 4.4. We then reset τ_s to zero.

We now have many partitions corresponding to each scale. We create a coordinate for each cluster in each partition and give that coordinate value $4s$ for points in that cluster and 0 for other points. We additionally create coordinates which embed the tree-bandwidth decomposition tree into ℓ_p with distortion ρ; for each point in X_i we assign values to these coordinates corresponding to X_i. We then multiply all of these tree-bandwidth decomposition based coordinates by k. We claim that the coordinates described here will satisfy Theorem 4.6.

The proof will be based upon proving that all the partitions we use are almost-valid, that not too many timers are incremented at any step, and that therefore the expected expansion between a parent and child cannot be too large. We observe that this construction will be single-coordinate lower bounded except for pairs obtaining contribution from the embedding of the tree, since the proof proceeds by showing that points x, y should belong to different clusters of an almost-valid partition at an appropriate scale. Similarly, the local property will hold which implies Corollary 4.8. We now provide the detailed proof.

Lemma 4.9. *The clusterings we use for X_i are almost-valid.*

Proof. Consider some clustering C_s that is not *almost-valid*. It follows that there is some pair $x, y \in X_i$ such that $d(x, y) \leq \frac{s}{2k}$ but where x, y are in different clusters, or that there is some pair $x, y \in X_i$ such that $d(x, y) > 2s$ but where x, y are in the same cluster. We will consider the first case (the proof of contradiction for the second case is similar). We backtrack to the most recent time when we replaced cluster C_s. At that time, there were some ancestors of x, y which we will call $a(x), a(y)$. Since we have simply copied the clusters from the ancestors, and at that time the partition was valid, it must be that $d(a(x), a(y)) > \frac{s}{k}$. By triangle inequality, it must be that either $d(x, a(x)) > \frac{s}{4k}$ or $d(y, a(y)) > \frac{s}{4k}$. Since at each step the parent of a node is within k of the child, we conclude that there are at least $\frac{s}{4k^2}$ super-nodes in the intervening time, each of which includes an ancestor of x and an ancestor of y which are at least $\frac{s}{2k}$ and at most $\frac{s}{k}$ apart. It follows that the timer τ_s must have been updated at each of these super-nodes and thus was updated at least $\frac{s}{4k^2}$ times without generating a new partition C_s. This contradicts the manner in which the timers work. □

Lemma 4.10. *When we update timers for X_i, at most $O(k \log k)$ timers will be incremented.*

Proof. We first observe that a single distance can increment at most $O(\log k)$ timers. Consider building the minimum spanning tree on the nodes of X_i, where the weight of an edge equals the distance between its endpoints. There are $k - 1$ edges in this spanning tree. Now consider any pair of points x, y in X_i. Consider the path $P(x, y)$ through the spanning tree. If the distance between

x, y is less than the length of the longest spanning tree edge in $P(x, y)$, then we could produce a better spanning tree by removing that longest edge and replacing it with (x, y). On the other hand, the distance between x, y cannot exceed the length of $P(x, y)$ which is at most $(k-1)$ times the length of the longest spanning tree edge in $P(x, y)$. Thus every distance is within a factor of k of the length of some spanning tree edge. Distances within a factor of k of a particular distance can increment only $O(\log k)$ timers, so we get at most $O(k \log k)$ timers incremented in total. □

Lemma 4.11. *For any timer τ_s which is incremented, the probability of defining a new partition C_s for X_i is exactly $\frac{4k^2}{s}$; for any timer τ_s which is not incremented, the probability of defining a new partition C_s for X_i is zero.*

Proof. We always maintain that timers have value at most $\frac{s}{4k^2} - 1$, so we can only define a new clustering C_s if we increment timer τ_s. The number of times τ_s has been incremented since the root node is deterministic, so there is exactly one initial value for τ_s such that we will define a new clustering C_s. Since the initial values were determined uniformly at random, the lemma follows. □

Theorem 4.12. *The expected embedded distance between $x \in X_i$ and $p(x) \in X_{p(i)}$ is no more than $O(k^3 \log k + k\rho)$ and at least k.*

Proof. Consider nodes x and $p(x)$. The coordinates which differ for these nodes are the coordinates corresponding to the tree-bandwidth decomposition X_i and $X_{p(i)}$ as well as possibly some coordinates corresponding to partitions. Since the embedding of the tree-bandwidth decomposition has distortion ρ and X_i and $X_{p(i)}$ are adjacent, using these coordinates and multiplying them by k gives us a distance of at most $k\rho$ and at least k. This immediately gives us the lower bound claimed; we will prove the upper bound.

Inherited partitions don't create any distance between parent and child – the same coordinate will be given the same value. The coordinates where x and $p(x)$ differ are those corresponding to clusters of partitions C_s where we defined a new valid partition for X_i. Whenever this happens, we will have x and $p(x)$ in different clusters, so we will have a pair of coordinates one of which is $4s$ and the other of which is zero for x, with $p(x)$ being exactly the opposite. For any C_s where we incremented the timer, the probability of this mismatch will be $\frac{4k^2}{s}$ by 4.11. Applying linearity of expectations and the fact that at most $k \log k$ timers were incremented by lemma 4.10, we can conclude that the expected ℓ_1 distance between parent and child is at most $O(k^3 \log k + k\rho)$. Since the ℓ_p distance can never exceed the ℓ_1 distance (for $p > 1$) we have the same bound for ℓ_p. □

We will now complete the proof of Theorem 4.6. We will define $d_E(x, y)$ to be the embedded distance between x and y, and attempt to relate it to $d(x, y)$. We consider three cases:

1. If x, y are in the same set X_i in the tree decomposition, then we simply combine Theorem 4.5 with Lemma 4.9 to see that their embedded distance via the clusterings does not contract the real distance and does not expand the real distance by more than $O(k)$. We observe that the coordinates corresponding to the tree decomposition will be identical for x and y and will therefore have no effect.

2. If x, y are in adjacent sets $x \in X_i$ and $y \in X_j$, then assume without loss of generality that $j = p(i)$. Abusing notation slightly, let $p(x)$ be the nearest node to x in X_j. Now $E[d_E(y, x)] \leq E[d_E(y, p(x))] + E[d_E(p(x), x)]$ by applying triangle inequality and linearity of expectation to the embedded distances. Since y and $p(x)$ are both in X_j, we can apply the previous case to

this distance. For the distance between x and its parent, we apply Theorem 4.12. Combining these yields $E[d_E(y,x)] \leq O(k)d(y,p(x)) + O(k\rho + k^3 \log k) \leq O(k)(d(y,x)+k) + O(k\rho + k^3 \log k) \leq O(k)d(x,y) + O(k\rho + k^3 \log k) \leq d(x,y)O(k\rho + k^3 \log k)$. On the other hand, we also guarantee that $d_E(y,x) \geq (d_E(y,p(x))^p + k^p)^{1/p}$ where the first term comes from the almost-valid partition coordinates and the second from the coordinates corresponding to the tree-bandwidth decomposition. This gives us $d_E(y,x) \geq \frac{1}{2}d(y,x)$ since if $d(y,x) < 2k$ we have $d_E(y,x) \geq k$ and otherwise we have $d(y,p(x)) \geq \frac{1}{2}d(y,x)$ and $d_E(y,p(x)) \geq d(y,p(x))$ since $y, p(x)$ are in the same set. This bounds the contraction.

3. If $x \in X_i$ and $y \in X_j$ for distinct, non-adjacent sets X_i, X_j, then there is some path Q through the tree $T = (I, F)$ from i to j. We will prove our bound by induction on the length of this path, with the base case being covered in the previous case where $|Q| = 1$ and X_i, X_j are thus adjacent. Inductively, we find the shortest path in the original metric between x and y, and observe that it must visit some node z in each set Z lying on the path Q. Since z is closer in the tree-bandwidth decomposition to both x and y, we can inductively write $d_E(x,z) \leq O(k^3 \log k + k\rho)d(x,z)$ and similarly for $d_E(z,y)$. Applying triangle inequality along with the fact that $d(x,y) = d(x,z) + d(y,z)$ completes the induction and gives us the desired bound on the expansion.

   We now need to bound the contraction. We consider W to be the common ancestor of the super-nodes containing x and y, and let x', y' be closest nodes in W to x and y respectively. The coordinates representing partitions guarantee that $d_E(x,y) \geq d_E(x',y')$ since x and y inherit clusterings from their ancestors (i.e. x and y cannot be in the same cluster at a particular scale unless x', y' are also in the same cluster). The tree decomposition coordinates give $d_E(x,y) \geq k|Q|$. Since x', y' are in the same super-node we have $d_E(x',y') \geq d(x',y')$, and because of the parent-child distance of at most k we can guarantee that $d(x',y') \geq d(x,y) - k|Q|$. Combining these yields:

$$d_E(x,y) \geq \max[d(x,y) - k|Q|, k|Q|] \geq \frac{1}{2}d(x,y)$$

Thus we have expansion by at most $O(k^3 \log k + k\rho)$ and contraction by at most a factor of two. We can eliminate contraction by simply doubling the values of all coordinates (which increases expansion by a factor of two).

In general we will have $\rho \leq O(\sqrt{\log \log n})$ as this is the bound for embedding trees into ℓ_p. Of course, for ℓ_1 we have $\rho = 1$ and for specific trees we may have smaller ρ values in ℓ_2.

References

[ABN08] I. Abraham, Y. Bartal, and O. Neiman. Embedding metric spaces in their intrinsic dimension. In *SODA '08: Proceedings of the 18th ann. ACM-SIAM sym. on Discrete algorithms*, 2008.

[Ass83] P. Assouad. Plongements lipschitziens dans $\mathbb{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.

[BC03] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in l_1. In *FOCS*, pages 514–523, 2003.

[Bec91] J. Beck. An algorithmic approach to the lovasz local lemma. *Random Struct. Algorithms*, 2:343–365, 1991.

[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.

[CCDG82] P. Z. Chinn, J. Chvtalov, A. K. Dewdney, and N. E. Gibbs. The bandwidth problem for graphs and matrices - a survey. *Journal of Graph Theory*, (6):223–254, 1982.

[CGM06] D. E. Carroll, A. Goel, and A. Meyerson. Embedding bounded bandwidth graphs into ℓ_1. *Proceedings of the 33rd Ann. ICALP*, pages 27–37, 2006.

[CGN+03] C. Chekuri, A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Embedding k-outerplanar graphs into ℓ_1. In *SODA '03: Proceedings of the 14th ann. ACM-SIAM sym. on Discrete algorithms*, pages 527–536, 2003.

[CJLV08] A. Chakrabarti, A. Jaffe, J. R. Lee, and J. Vincent. Embeddings of topological graphs: Lossy invariants, linearization, and 2-sums. In *Proceedings of the 49th Ann. IEEE Sym. on Foundations of Computer Science*, pages 761–770, 2008.

[Fei98] U. Feige. Approximating the bandwidth via volume respecting embeddings. In *Proceedings of the 30th ann. ACM sym. on Theory of computing*, STOC '98, pages 90–99, New York, NY, USA, 1998. ACM.

[GKL03] A. Gupta, R. Krauthgamer, and J.s R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *Proceedings of the 44th Ann. IEEE Sym. on Foundations of Computer Science*, page 534, 2003.

[GNRS99] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and ℓ_1-embeddings of graphs. In *Proceedings of the 40th Ann. Sym. on Foundations of Computer Science*, pages 399–409, 1999.

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[Mat99] J. Matousek. On embedding trees into uniformly convex banach spaces. *Israel Journal of Mathematics*, 114:221–237, 1999.

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

[NR02] Ilan Newman and Yuri Rabinovich. A lower bound on the distortion of embedding planar metrics into euclidean space. In *SCG '02: Proceedings of the 18th ann. sym. on Computational geometry*, pages 94–96. ACM Press, 2002.

[Rao99] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the 15th Ann. Sym. on Computational Geometry*, pages 300–306, New York, 1999. ACM.

[Sem96] S. Semmes. On the nonexistence of bilipschitz parameterizations and geometric problems about a_∞ weights. *Revista Matemática Iberoamericana*, 12:337–410, 1996.

[ZACS09] Q. Zhu, Z. Adam, V. Choi, and D. Sankoff. Generalized gene adjacencies, graph bandwidth, and clusters in yeast evolution. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 6(2):213–220, 2009.

A Proof of Theorem 2.5

Consider the following transformation of ϕ. The first step is to unite coordinates assigned to the same scale, and make sure the value of any scale s is bounded. Let $\phi^{(s)} = \min\{\Sigma_{c:s_c=s}\phi_c, \beta s\}$. Note that for any $x, y \in V$, if $\phi^{(s)}(y) = \beta s$ then $\phi^{(s)}(x) - \phi^{(s)}(y) \le \beta s - \beta s = 0$ and if $\phi^{(s)}(y) = \Sigma_{c:s_c=s}\phi_c(y)$ then $\phi^{(s)}(x) - \phi^{(s)}(y) \le \Sigma_{c:s_c=s}(\phi_c(x) - \phi_c(y))$. In any case it holds by symmetry of x, y that $|\phi^{(s)}(x) - \phi^{(s)}(y)| \le \Sigma_{c:s_c=s}|\phi_c(x) - \phi_c(y)|$.

Let I be the set of new coordinates, its exact size will be determined later. For each scale s and $i \in I$ we select σ_s^i independently and uniformly from $[0, 1]$. We now define:

$$\phi^i(x) \leftarrow \Sigma_s \sigma_s^i \phi^{(s)}(x)$$

This replaces the D coordinates generated by ϕ with a single coordinate. We observe that the ℓ_1 expansion remains bounded because $|\phi^i(x) - \phi^i(y)| \le \Sigma_c|\phi_c(x) - \phi_c(y)|$. The problem is that this transformation may introduce contraction. We will first show that for any pair of points, the contraction is bounded with $1 - 2\epsilon$ probability. In what follows we assume w.l.o.g that the lower bound single coordinate for any pair $x, y \in X$ has scale s such that $s < d(x, y) \le 2s$ (this will only affect the constant factors of the following analysis).

Lemma A.1. *For any pair of points $x, y \in V$, we have $|\phi^i(x) - \phi^i(y)| \ge \epsilon\beta d(x, y)/4$ with probability at least $1 - 2\epsilon$.*

Proof. This follows from the single-coordinate (ϵ, β) lower-bound condition. Let $\mathcal{E}$ be the event that there is a lower bounding coordinate, and let z be that coordinate, *i.e.* $|\phi_z(x) - \phi_z(y)| \ge \beta d(x, y)$. Recall that $\Pr[\mathcal{E}] \ge 1 - \epsilon$. Let $s = s_z$ and note that since $s < d(x, y)$ the second local condition implies that either $\phi^{(s)}(x)$ or $\phi^{(s)}(y)$ is zero. Assume w.l.o.g that $\phi^{(s)}(y) = 0$, then assuming event $\mathcal{E}$ we get that $\phi_z(x) \ge \beta d(x, y) > \beta s$ and so $\phi^{(s)}(x) = \beta s$ (recall that we assumed that the range of ϕ is non-negative).

Let $f(x) = \Sigma_{t \ne s}\sigma_t^i \phi^{(t)}(x)$, and denote $A = \beta s$. Assuming that event $\mathcal{E}$ occurred, the random variable $\sigma_s^i \phi^{(s)}(x)$ takes values uniformly on the interval $[0, A]$, independently from the random choices determining $f(x)$ and $f(y)$. Therefore for any value of $|f(x) - f(y)|$ there is probability at least $1 - \epsilon$ that the distance of $\sigma_s^i \phi^{(s)}(x)$ from $|f(x) - f(y)|$ is at least $\epsilon A/2$. In such a case we get that

$$|\phi^i(x) - \phi^i(y)| = |f(x) - f(y) + \sigma_s^i\phi^{(s)}(x) - \sigma_s^i\phi^{(s)}(y)| \ge \Big||f(x) - f(y)| - \sigma_s^i\phi^{(s)}(x)\Big| \ge \epsilon\beta s/2 \ .$$

The probability of both these independent events happening is at least $(1 - \epsilon)^2 \ge 1 - 2\epsilon$. The proof is completed recalling that $s \ge d(x, y)/2$.

□

Of course, this contraction bound is not independent from one pair of points to another. This is why we have the superscript i; we will in fact generate many functions ϕ^i, each of which will be one coordinate in our normed space. We can use Chernoff bounds to show that $O(\log n)$ coordinates are sufficient; however we would like to reduce the dimension to resemble the doubling dimension of (V, d) which may be substantially smaller than $\log n$. The idea is to use locality along with the Lovasz Local Lemma. We want to establish that for a sufficient number of dimensions, there is a non-zero probability that *no distance contracts.*

For any pair x, y, let z be the single-coordinate lower-bound for (x, y). We define $\mathcal{Z}^i(x, y)$ to be the event that if σ_s^i were zero for all $s < Qs_z$ (for $Q < 1$ to be defined later), then we would contract by at most a factor of $\epsilon\beta/4$. In other words, it is the event that:

$$|\Sigma_{s \geq Qs_z} \sigma_s^i(\phi^{(s)}(x) - \phi^{(s)}(y))| \geq \epsilon\beta d(x, y)/4$$

Lemma A.2. *Provided Q is sufficiently small ($O(\epsilon)$), if $\mathcal{Z}^i(x, y) = 1$ then $|\phi^i(x) - \phi^i(y)| \geq \epsilon\beta d(x, y)/8$.*

Proof. We can split ϕ^i by considering scales separately and conclude that:

$$|\phi^i(x) - \phi^i(y)| \geq |\Sigma_{s \geq Qs_z} \sigma_s^i(\phi^{(s)}(x) - \phi^{(s)}(y))| - |\Sigma_{s < Qs_z} \sigma_s^i(\phi^{(s)}(x) - \phi^{(s)}(y))|$$

Because $\mathcal{Z}^i(x, y) = 1$ the first expression is at least $\epsilon\beta d(x, y)/4$. For the second expression, we observe that this includes only scales smaller than Qs_z. By definition $|\phi^{(s)}(x) - \phi^{(s)}(y)| \leq \beta s$, so

$$|\Sigma_{s < Qs_z} \sigma_s^i(\phi^{(s)}(x) - \phi^{(s)}(y))| \leq \beta \Sigma_{s < Qs_z} s \leq \beta Q s_z \ . \tag{4}$$

The second locality property implies that $s_z < cd(x, y)$ for some constant c, and by selecting sufficiently small $Q = \epsilon/(8c)$ we can make (4) at most $\epsilon\beta d(x, y)/8$, which enables us to complete the proof. □

We define $\mathcal{G}(x, y)$ to be the event that $\mathcal{Z}^i(x, y) = 1$ for at least $1/2$ of the i values. We will establish that if $\mathcal{G}(x, y)$ holds for every pair x, y in a net, then this will be sufficient to imply that no contraction occurs.

Lemma A.3. *For each scale s, let N_s be an $\frac{\epsilon\beta s}{64\delta}$-net; in other words, a set of points each pair of which are at least $\frac{\epsilon\beta s}{64\delta}$ apart such that all points in V are within $\frac{\epsilon\beta s}{32\delta}$ of the nearest member of N_s.*[2] *If for every s and every $x, y \in N_s$ we have $\mathcal{G}(x, y)$, then for all pairs of points our contraction is bounded by a constant times $\epsilon\beta$.*

Proof. Consider any pair of points u, v. Let $d(u, v)/2 \leq s < d(u, v)$. Let x, y be the nearest neighbors of u, v respectively in the net N_s. Then $d(x, u) \leq \frac{\epsilon\beta s}{64\delta}$ and $d(y, v) \leq \frac{\epsilon\beta s}{64\delta}$. This suggests that $d(u, v)/2 \leq d(x, y) \leq 2d(u, v)$. Since we have $x, y \in N_s$, the event $\mathcal{G}(x, y)$ must hold, and therefore there are a constant fraction of $i \in I$ for which $\mathcal{Z}^i(x, y) = 1$. Applying lemma A.2 we conclude that the weighted sum of ϕ^i gives us an embedded distance between x and y of at least $\epsilon\beta d(x, y)/8$. Let $\tilde{d}(u, v)$ be the embedded distance between u and v using the weighted combination of ϕ^i coordinates. Then:

$$\tilde{d}(u, v) \geq \tilde{d}(x, y) - \tilde{d}(x, u) - \tilde{d}(y, v) \geq \epsilon\beta d(x, y)/16$$

The first inequality follows from triangle inequality; the second from the expansion bound of δ on $\tilde{d}(x, u)$ and $\tilde{d}(y, v)$. □

We now need to argue that the probability of all $\mathcal{G}(x, y)$ events holding simultaneously is non-zero. We will do this by applying Lemma A.6, combined with the following lemmas about the probabilities.

[2]We may assume that the nets are hierarchical, *i.e.* a net for scale s will contain all the net points of scales $t > s$

Lemma A.4. *Let $u, v \in N_s$ and $d(u,v)/2 \leq s < d(u,v)$. Define T as the set of u', v' pairs such that $u', v' \in N_s$ and additionally $d(u,u') < 4\frac{s}{Q}$, $d(u,v') < 4\frac{s}{Q}$. Then:*

$$\Pr[\neg \mathcal{Z}^i(u,v) \mid \bigwedge_{u',v' \in (N_s - T): d(u',v') \geq s} \mathcal{G}(u',v')] \leq 2\epsilon$$

Proof. We note that from Lemma A.1, the probability of $\mathcal{Z}^i(u,v)$ is $1-2\epsilon$. This is the case regardless of the variables σ_t^i corresponding to larger scales. Consider any $u', v' \in N_s - T$ with $d(u',v') \geq s$. By definition of T, we must have $d(u,u') \geq 4\frac{s}{Q}$. If the scale of $d(u',v')$ is such that $d(u',v') > \frac{s}{Q}$ then event $\mathcal{G}(u',v')$ is independent of any coordinates corresponding to scale s or smaller, so will not effect the probability of $\mathcal{Z}^i(u,v)$. On the other hand, if $d(u',v') \leq \frac{s}{Q}$ then in order for there to be some coordinate which effects both $\mathcal{G}(u',v')$ and $\mathcal{Z}^i(u,v)$ that coordinate would need to be of scale at most s and non-zero for one of u', v' and also one of u, v. It follows from the first local property of ϕ that some pair of these nodes is within s; say $d(v,u') \leq s$. But then $d(u,u') \leq s + d(u,v) \leq 3s$ and $d(u,v') \leq d(u,u') + d(u',v') \leq 3s + \frac{s}{Q}$. Thus $(u',v') \in T$ and will not be included in the union of events. We conclude that $\mathcal{G}(u',v')$ (which is completely determined by coordinates of zero value for u and v combined with coordinates at larger scale than $d(u,v)$) does not effect the probability of $\mathcal{Z}^i(u,v)$.

□

Lemma A.5. *Again, for $u, v \in N_s$ define T as the set of u', v' pairs such that $u', v' \in N_s$ and $d(u,u') < 4\frac{s}{Q}$, $d(u,v') < 4\frac{s}{Q}$ for $Q = \Theta(\epsilon)$. Then $|T| \leq 2^{O(\log \alpha \log \frac{\delta}{\beta})}$ where α is the doubling constant (and $\log \alpha$ thus the doubling dimension) of (V,d) and δ is the expansion of ϕ.*

Proof. Consider the ball of radius $4\frac{s}{Q}$ about u. Both u', v' must lie within this ball. By the definition of the doubling constant, this ball can be covered by $\alpha^{\log(512\delta/(Q\epsilon\beta)}$ balls of radius $\epsilon\beta s/(128\delta)$. Each of these balls contains at most one point from N_s because points in N_s must be at least $\epsilon\beta s/(64\delta)$ apart. By bounding the number of candidate points, we bound the size of $|T|$ by the square of that value, which is still at most the required bound (recalling that $Q\epsilon = O(\epsilon^2)$ and that $\epsilon \geq \beta/\delta$ so $\log(1/\epsilon^2) \leq O(\log(\delta/\beta))$. □

We now state a variation on the Lovasz Local Lemma proven in [ABN08].

Lemma A.6 (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 $c(\mathcal{A}_i) \geq c(\mathcal{A}_j)$. Assume that for any $i = 1, \ldots, n$*

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

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

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

We will apply this lemma to complete the proof of Theorem 2.5. The "rating function" referred to in the local lemma will simply order the coordinates by their scales (decreasing order). The dependency graph connects the event $\mathcal{G}(u, v)$ to the events $\mathcal{G}(u', v')$ for the pairs $u', v' \in T$. We can bound the value of $d \leq 2^{O(\log \alpha \log \frac{\delta}{\beta})}$, which is the number of dependencies, according to Lemma A.5.

Let $m = \Omega\left(\log \alpha \frac{\log(\delta/\beta)}{\log(1/\epsilon)}\right)$ be the number of coordinates of ϕ^i, and let $\bar{Z} = \bar{Z}(x, y) = \Sigma_{i=1}^{m} \neg Z^i(x, y)$ be the number of failed coordinates, then by Lemma A.4 it follows that $\mathbb{E}[\bar{Z}] \leq 2\epsilon m$, and by Chernoff bound

$$\Pr[\bar{Z} > m/2] = \Pr\left[\bar{Z} > \frac{\mathbb{E}[\bar{Z}]}{4\epsilon}\right] \leq \left(\frac{e^{1/(4\epsilon)-1}}{1/(4\epsilon)^{1/(4\epsilon)}}\right)^{2\epsilon m},$$

since $\epsilon < 1/8$ it follows that the above expression is bounded by $\exp\{-1/(8\epsilon) \cdot \ln(1/(4\epsilon)) \cdot 2\epsilon m\} \leq \exp\{-\Omega(\log \alpha \log(\delta/\beta))\}$ as required. As Lemma A.4 suggests, this bound still holds even if we condition on events $\mathcal{G}(u', v')$ for some $u', v' \in N_s \setminus T$.

It will follow that the probability of simultaneously having all $\mathcal{G}$ events occur for pairs satisfying the condition of Lemma A.3 will be nonzero. We can find the choice of randomness necessary for this to occur in polynomial time using a technique similar to [Bec91, MT10] as described in [ABN08]. We can make this embedding non-contracting by simply multiplying all coordinates by $O(1/(\epsilon^2\beta))$.