Metric Embeddings with Relaxed Guarantees

Ittai Abraham* Yair Bartal* T-H. Hubert Chan† Kedar Dhamdhere† Anupam Gupta†
Jon Kleinberg‡ Ofer Neiman* Aleksandrs Slivkins‡

Abstract

We consider the problem of embedding finite metrics with slack*: we seek to produce embeddings with small dimension and distortion while allowing a (small) constant fraction of all distances to be arbitrarily distorted. This definition is motivated by recent research in the networking community, which achieved striking empirical success at embedding Internet latencies with low distortion into low-dimensional Euclidean space, provided that some small slack is allowed.*

Answering an open question of Kleinberg, Slivkins, and Wexler [29], we show that provable guarantees of this type can in fact be achieved in general: any finite metric can be embedded, with constant slack and constant distortion, into constant-dimensional Euclidean space. We then show that there exist stronger embeddings into ℓ_1 which exhibit gracefully degrading *distortion: these is a single embedding into ℓ_1 that achieves distortion at most $O(\log \frac{1}{\epsilon})$ on all but at most an ϵ fraction of distances,* simultaneously *for all $\epsilon > 0$. We extend this with distortion $O(\log \frac{1}{\epsilon})^{1/p}$ to maps into general ℓ_p, $p \geq 1$ for several classes of metrics, including those with bounded doubling dimension and those arising from the shortest-path metric of a graph with an excluded minor. Finally, we show that many of our constructions are tight, and give a general technique to obtain lower bounds for ϵ-slack embeddings from lower bounds for low-distortion embeddings.*

*School of Computer Science and Engineering, Hebrew University of Jerusalem, Jerusalem, Israel. Email: {*ittaia, yair, neiman*} *at cs.huji.ac.il*. Bartal and Neiman were supported in part by a grant from the Israeli Science Foundation (195/02).

†Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: { *hubert, kedar, anupamg*} *at cs.cmu.edu*. Chan and Dhamdhere were supported in part by NSF ITR CCR-0122581 (The ALADDIN project), and Chan and Gupta were supported by an NSF CAREER award CCF-0448095, and by an Alfred P. Sloan Fellowship.

‡Department of Computer Science, Cornell University, Ithaca, NY 14853, USA. Email: {*kleinber, slivkins*} *at cs.cornell.edu*. Kleinberg was supported by a David and Lucile Packard Foundation Fellowship and NSF grants CCF-0325453, IIS-0329064, and CCR-0122381; work done in part while on sabbatical leave at Carnegie Mellon University. Slivkins was supported by the Packard Fellowship of Jon Kleinberg.

1 Introduction

Over the past decade, the field of metric embeddings has gained much importance in algorithm design. The central genre of problem in this area is the mapping of a given metric space into a "simpler" one, in such a way that the distances between points do not change too much. More formally, an *embedding* of a finite metric space (V, d) into a *target* metric space (V', d') is a map $\varphi : V \to V'$. Recent work on embeddings has used *distortion* as the fundamental measure of quality; the distortion of an embedding is the worst multiplicative factor by which distances are increased by the embedding[1]. The popularity of distortion has been driven by its applicability to approximation algorithms: if the embedding $\varphi : V \to V'$ has a distortion of D, then the cost of solutions to some optimization problems on (V, d) and on $(\varphi(V), d')$ can only differ by some function of D; this idea has led to numerous approximation algorithms [25].

In parallel with theoretical work on embeddings, there has been a surge of interest in the networking community on *network embedding* problems closely related to the framework above (see e.g. [15, 36, 41]). This work is motivated by different applications: one takes the point-to-point latencies among nodes in a network such as the Internet, treats this as a distance matrix,[2] and embeds the nodes into a low-dimensional space so as to approximately preserve the distances. In this way, each node is assigned a short sequence of virtual "coordinates," and distances between nodes can be approximated simply by looking up their coordinates and computing the distance, rather than having to interact with the relevant nodes themselves. As location-aware applications in networks become increasingly prevalent — for ex-

[1]Formally, for an embedding $\varphi : V \to V'$, the *distortion* is the smallest D so that $\exists\, \alpha, \beta \geq 1$ with $\alpha \cdot \beta \leq D$ such that $\frac{1}{\alpha}\, d(x, y) \leq d'(\varphi(x), \varphi(y)) \leq \beta\, d(x, y)$ for all pairs $x, y \in V \times V$. Note that this definition of distortion is slightly non-standard—since $\alpha, \beta \geq 1$, it is no longer invariant under arbitrary scaling; however, this is merely for notational convenience, and all our results can be cast in the usual definitions of distortion.

[2]While the triangle inequality can be violated by network latencies, empirical evidence suggests that these violations are small and/or infrequent enough to make metric methods a useful approach.

ample, finding the nearest server in a distributed application with replicated services, or finding the nearest copy of a file or resource in a peer-to-peer system — having such distance information in a compact and easily usable form is an issue of growing importance (see e.g. the discussion in [15]).

In the context of these networking applications, however, distortion as defined above has turned out to be too demanding an objective function — many metrics cannot be embedded into Euclidean space with constant distortion; many of those that can be so embedded require a very large number of dimensions; and the algorithms to achieve these guarantees require a type of centralized coordination (and extensive measurement of distances) that is generally not feasible in Internet settings. Instead, the recent networking work has provided *empirical* guarantees of the following form: if we allow a small fraction of all distances to be *arbitrarily* distorted, we can embed the remainder with (apparently) constant distortion in constant-dimensional Euclidean space. Such guarantees are natural for the underlying networking applications; essentially, a very small fraction of the location-based lookups may yield poor performance (due to the arbitrary distortion), but for the rest the quality of the embedding will be very good.

These types of results form a suggestive contrast with the theoretical work on embeddings. In particular, are the strong empirical guarantees for Internet latencies the result of fortuitous artifacts of this particular set of distances, or is something more general going on? To address this, Kleinberg, Slivkins, and Wexler [29] defined the notion of embeddings with *slack*: in addition to the metrics (V, d) and (V', d') in the initial formulation above, we are also given a *slack parameter* ϵ, and we want to find a map φ whose distortion is bounded by some quantity $D(\epsilon)$ on all but an ϵ fraction of the pairs of points in $V \times V$. (Note that we allow the distortion on the remaining ϵn^2 pairs of points to be arbitrarily large.) This question can be viewed as a natural variant of metric Ramsey theory [10]. Roughly, Kleinberg et. al. [29] showed that any metric of bounded doubling dimension — in which every ball can be covered by a constant number of balls of half the radius — can be embedded with constant distortion into constant-dimensional Euclidean space, allowing a constant slack ϵ. Such metrics, which have been extensively studied in their own right, have also been proposed on several occasions as candidates for tractable abstractions of the set of Internet latencies (see e.g. [19, 27, 36, 38]).

There were two main open questions posed in [29].

(1) There was no evidence that the main embedding result of [29] needed to be restricted to metrics of bounded doubling dimension. Could it be the case that for *every* finite metric, and every $\epsilon > 0$, there is an embedding of the metric with distortion $f(\epsilon)$ into Euclidean space?

(2) Rather than have the embedding depend on the given slack parameter ϵ, a much more flexible and powerful alternative would be to have a *single* embedding of the metric with the property that, for some (slowly growing) function $D(\cdot)$, it achieved distortion $D(\epsilon)$ on all but an ϵ fraction of distance pairs, for all $\epsilon > 0$. We call such an embedding *gracefully degrading* [29], and ask whether such an embedding (with a polylogarithmic function $D(\cdot)$) could exist for all metrics.

In this paper, we resolve the first of these questions in the affirmative, showing constant distortion with constant slack for all metrics. Moreover, the embedding we design to achieve this guarantee is *beacon-based*, requiring only the measurement of distances involving a small set of distinguished "beacon nodes"; see Section 2. Approaches that measure only a small number of distances are crucial in networking applications, where the full set of distances can be enormous; see, e.g., [23, 20, 30, 36, 37, 42] for beacon-based approaches and further discussions. We then resolve the second question in the affirmative for metrics that admit an $O(1)$-padded decomposition (a notion from previous work on embeddings that we specify precisely in Section 1.1); this includes several well-studied classes of metrics including those with bounded doubling dimension and those arising from the shortest-path metric of a graph with an excluded minor. We further show that gracefully degrading distortion can be achieved in the ℓ_1 norm for all metrics. The second question has been subsequently solved in full in [1] (see also the bibliographic notes in the sequel), providing an embeddings with gracefully degrading distortion for all metrics in ℓ_p for every $p \geq 1$. Finally, we show that many of our constructions are tight, and give a general technique to obtain lower bounds for ϵ-slack embeddings from lower bounds for low-distortion embeddings.

Basic Definitions. Before we formally present our results, let us present some of the notions that will be used throughout the paper. We will assume that the metric (V, d) is also represented as a graph on the nodes V, with the length of edge uv being $d(u, v) = d_{uv}$. We imagine this graph as having n^2 edges, one for each pair $u, v \in V \times V$; this makes the exposition cleaner and does not change the results in any significant way. For a map $\varphi : V \to V'$ let us define the notion of the *distortion of a set S of edges* under embedding φ as the smallest $D \geq 1$ such that for some positive constant K and all edges $(u, v) \in S$ we have

$$d(u, v) \leq d'(\varphi(u), \varphi(v))/K \leq D \cdot d(u, v).$$

Note that the distortion of φ (as given in Footnote 1) is the same as the distortion of the set of all edges.

Definition 1.1 (ϵ-slack distortion) *Given ϵ, an embedding $\varphi : V \to V'$ has* distortion D with ϵ-slack *if a set of all but an ϵ-fraction of edges has distortion at most D under φ.*

We will also consider a stronger notion of slack, for which we need the following definition. Let $\rho_u(\epsilon)$ be the radius of the smallest ball around u that contains at least ϵn nodes. Call an edge uv *ϵ-long* if $d_{uv} \geq \min(\rho_u(\epsilon), \rho_v(\epsilon))$. Then there are at least $(1-\epsilon)\, n^2$ edges that are ϵ-long. For any such edge uv, at least one endpoint u is at least as far from the other endpoint v as the (ϵn)-th closest neighbor of v.

Definition 1.2 (ϵ-uniform slack distortion) *Given ϵ, an embedding $\varphi : V \to V'$ has* distortion D with ϵ-uniform slack *if the set of all ϵ-long edges has distortion at most D.*

While the above notions of embeddings with slack allow the map φ to depend on the slack ϵ, the following notion asks for a single map that is *good for all ϵ simultaneously*.

Definition 1.3 (gracefully degrading distortion) *An embedding $\psi : V \to V'$ has a* gracefully degrading distortion $D(e)$ *if for each $\epsilon > 0$, the distortion of the set of all ϵ-long edges is at most $D(\epsilon)$.*

Our Results. We now make precise the main results described above, and also describe some further results in the paper. Our first result shows that if we are allowed constant slack, we can indeed embed *any* metric into constant dimensions with constant distortion.

Theorem 1.4 *For any source metric (V, d), any target metric ℓ_p, $p \geq 1$ and any parameter $\epsilon > 0$, we give the following two $O(\log \frac{1}{\epsilon})$-distortion embeddings:*

(a) with ϵ-slack into $O(\log^2 \frac{1}{\epsilon})$ dimensions, and

(b) with ϵ-uniform slack into $O(\log n \log \frac{1}{\epsilon})$ dimensions.

Both embeddings can be computed with high probability by randomized beacon-based algorithms.

These results extend Bourgain's theorem on embedding arbitrary metrics into ℓ_p, $p \geq 1$ with distortion $O(\log n)$ [11], and are proved in a similar manner.

Note that the bounds on both the distortion as well as the dimension in Theorem 1.4(a) are independent of the number of nodes n, which suggests that they could be extended to infinite metrics; this is further discussed in Section 2. In part (b), the dimension is proportional to $\log n$; we show that, for arbitrary metrics, this dependence on n is indeed inevitable. As an aside, let us mention that metrics of bounded doubling dimension do not need such a dependence on n: in Slivkins [42], these metrics are embedded into any ℓ_p, $p \geq 1$ with ϵ-uniform slack, distortion $O(\log \frac{1}{\epsilon} \log\log \frac{1}{\epsilon})$ and dimension $(\log \frac{1}{\epsilon})^{O(\log \frac{1}{\epsilon})}$.

In section 3 we generalize these results by a theorem that converts practically any classical embedding into ℓ_p, into ϵ-slack embedding or ϵ-uniform slack embedding, in the latter case there is an inevitable increase of $O(\log n)$ factor in the dimension. The following corollary gives some main examples:

Corollary 1.5
1. *Any finite metric space has a ϵ-slack embedding into ℓ_p with distortion $O(\log \frac{1}{\epsilon})$ and dimension $O(\log \frac{1}{\epsilon})$.*
2. *Any finite metric space has a ϵ-slack embedding into ℓ_p with distortion $O(\lceil (\log \frac{1}{\epsilon})/p \rceil)$ and dimension $e^{O(p)} \log \frac{1}{\epsilon}$.*
3. *Any decomposable metric [3] has a ϵ-slack embedding into ℓ_p with distortion $O((\log \frac{1}{\epsilon})^{1/p})$ and dimension $O(\log^2 \frac{1}{\epsilon})$.*
4. *Any negative type metric (in particular l_1 metrics) has a ϵ-slack embedding into ℓ_2 with distortion $O\left(\sqrt{\log \frac{1}{\epsilon}} \log\log \frac{1}{\epsilon}\right)$ and dimension $O(\log^2 \frac{1}{\epsilon})$.*
5. *Any tree metric has a ϵ-slack embedding into ℓ_2 with distortion $O\left(\sqrt{\log\log \frac{1}{\epsilon}}\right)$ and dimension $O(\log \frac{1}{\epsilon})$.*

Where the improved dimension in the first two results follows from a recent improvement of the dimension in Bourgain's theorem due to [8].

We then study *embeddings into trees*. We extend the known results of probabilistic embedding into trees [5, 6, 17, 7] to obtain embeddings with slack. In particular, we use the technique of Fakcharoenphol et al. [17] to obtain the following two results:

Theorem 1.6 *For any input metric (V, d) and any parameter $\epsilon > 0$ there exists an embedding into a tree metric with ϵ-uniform slack and distortion $O(\frac{1}{\epsilon} \log \frac{1}{\epsilon})$.*

In fact, the tree metric in Theorem 1.6 is induced by a *Hierarchically Separated Tree* (*HST*) [5], which is a rooted tree with edge-weights w_e such that $w_e < w_{e'}/2$ whenever edge e' is on the path from the root to edge e.

Theorem 1.7 *For any input metric (V, d), the randomized embedding of [17] into tree metrics has expected gracefully degrading distortion $D(\epsilon) = O(\log \frac{1}{\epsilon})$. [4] Since tree metrics are isometrically embeddable into L_1, this immediately implies that we can embed any metric into L_1 with gracefully degrading distortion $D(\epsilon) = O(\log \frac{1}{\epsilon})$.*

---

[3] A metric is called *decomposable* if it admits a β-padded decomposition for some constant β, see Section 1.1 for details.

[4] More formally, we show that if an edge uv is ϵ-long, then $d_{uv} \leq E_T[d_T(u,v)] \leq O(\log \frac{1}{\epsilon})\, d_{uv}$, where d_T is the tree metric generated by the randomized algorithm in [17].

However, the dimension of the above embedding into L_1 may be prohibitively large. To overcome this hurdle, and to extend this embedding to ℓ_p, $p > 1$, we explore a different approach:

Theorem 1.8 *Consider a metric (V, d) which admits β-padded decompositions. Then it can be embedded into ℓ_p, $p \geq 1$ with $O(\log^2 n)$ dimensions and gracefully degrading distortion $D(\epsilon) = O(\beta)(\log \frac{1}{\epsilon})^{1/p}$.*

For the reader unfamiliar with padded decompositions, let us mention that $\beta \leq O(\dim_V)$, the doubling dimension of the metric, which in turn is always bounded above by $O(\log n)$. Moreover, doubling metrics, and metrics induced by planar graphs have $\beta = O(1)$; hence Theorem 1.8 implies that such metrics admit embeddings into ℓ_p, $p \geq 1$ with gracefully degrading distortion $O(\log \frac{1}{\epsilon})^{1/p}$. Note that for $p > 1$ this result can be seen as a strengthening of Theorem 1.4(b) on embeddings with ϵ-uniform slack.

The proof of Theorem 1.8 is technically the most involved part of the paper; at a high level, we develop a set of scale-based embeddings which are then combined together (as in most previous embeddings)—however, since the existing ways to perform this do not seem to guarantee gracefully degrading distortion, we construct new ways of defining distance scales.

Finally, we prove *lower bounds* on embeddings with slack: we give a very general theorem that allows us to convert lower bounds on the distortion and dimension of embeddings that depend only on $n = |V|$ into lower bounds in terms of the slack parameter ϵ. This result works under very mild conditions, and allows us to prove matching or nearly matching lower bounds for all of our results on ϵ-slack embeddings. These lower bounds are summarized in Corollary 6.5 of Section 6.

Related Work. This work is closely related to the large body of work on metric embeddings in theoretical computer science; see the surveys [25, 26] for a general overview of the area. Our results build on much of the previous work on embeddings into ℓ_p, including [11, 33, 40, 34, 21, 31, 32], and on embeddings of metrics into distributions of trees [2, 5, 6, 22, 17, 7]. Among the special classes of metrics we consider are *doubling metrics* [4, 21, 43, 35]; the book by Heinonen [24] gives more background on the analysis of metric spaces.

All of these papers consider low-distortion embeddings *without slack*. Note that an embedding with $(\epsilon = 1/2n^2)$-slack or $(\epsilon = 1/2n)$-uniform-slack is the same as an embedding with no slack; for many of our results, plugging in these values of ϵ gives us the best known slackless results—hence our results can be viewed as extensions of these previous results.

The notion of embedding with slack can be viewed as a natural variant of metric Ramsey theory. The first work on metric Ramsey-type problems was by Bourgain, Figiel and Milman [13] and a comprehensive study was more recently developed by Bartal et. al. [9, 10]. In the original metric Ramsey problem we seek a large subset of the points in the metric space which admit a low distortion embedding, whereas an embedding with slack provides low distortion for a subset of the pairs of points.

Bibliographic note. The results in this paper have been obtained independently by two groups: by I. Abraham, Y. Bartal and O. Neiman (whom we will refer to as G1 below), and by T-H.H. Chan, K. Dhamdhere, A. Gupta, J. Kleinberg and A. Slivkins (referred to as G2). The present paper combines the original write-ups of these two groups into a single presentation; as part of this, we briefly discuss the relation between the work of the two groups.

For embeddings with slack, Theorem 1.4 is due to (G2), and Theorem 3.1 is due to (G1). The results on lower bounds (Theorem 6.3) and embedding into distributions of trees (Theorem 1.7) were proved independently and similarly by both groups; the presentation in this paper of the proof of Theorem 4.1 follows that of (G2) while the presentation of the proof of Theorem 6.3 is a combination of the approaches of the two groups. The presentation in Section 5 and the result presented here on embedding into a single tree metric (Theorem 4.2) are due to (G2).

Abraham, Bartal, and Neiman (G1) independently proved several important extensions which will be published as a separate paper [1]. In particular, they resolve the second of the main open questions discussed above, i.e., showing a gracefully degrading embedding of any metric space into ℓ_p with $O(\log \frac{1}{\epsilon})$ distortion and dimension $O(\log n)$. This result is based on a new type of metric decompositions developed in [8, 1]. They also observe that the result implies a constant *average distortion* for embedding any metric into ℓ_p. In their paper they study more general notions of average distortion and in particular show this notion is applicable - and under certain conditions they show that Theorem 4.1 combined with their techniques yields improved approximation algorithms for problems such as sparsest cut, minimum linear arrangement, uncapacitated quadratic assignment etc. Among other results, they show a tight result of $O(\frac{1}{\sqrt{\epsilon}})$ distortion for ϵ-slack embedding into a tree metric, and improve the distortion in Theorem 1.8 to $O\left(\min\{\beta^{1-\frac{1}{p}}(\log \frac{1}{\epsilon})^{\frac{1}{p}}, \log \frac{1}{\epsilon}\}\right)$.

1.1 Notation and Preliminaries

Throughout the paper (V, d) is the metric to be embedded, and $d_{uv} = d(u, v)$ is the distance between nodes $u, v \in V$. Define the closed ball $\mathbf{B}_u(r) = \{v \in V \mid d_{uv} \leq r\}$.

The distance between a node u and set $S \subseteq V$ is denoted $d(u, S) = \min_{v \in S} d_{uv}$, and hence $d(u, V \setminus \mathbf{B}_u(r)) > r$. We will assume that the smallest distance in the metric is 1, and the largest distance (or the diameter) is Φ_d.

A *coordinate map* f is a function from V to $\mathbb{R}$; for an edge uv define $f(uv) = |f(u) - f(v)|$. Call such map *1-Lipschitz* if for every edge $f(uv) \leq d_{uv}$. For $k \in \mathbb{N}$ define $[k]$ as the set $\{0, 1, \ldots, k-1\}$.

Doubling. A metric (V, d) is *s-doubling* if every set $S \subseteq V$ of diameter Δ can be covered by s sets of diameter $\Delta/2$; the *doubling dimension* of such a metric is $\lceil \log s \rceil$ [24, 21]. A doubling metric is one whose doubling dimension is bounded. A *measure* is *s-doubling* if the measure of any ball $B_u(r)$ is at most s times larger than the measure of $B_u(r/2)$. It is known that for any s-doubling metric there exists an s-doubling measure; moreover, such measure can be efficiently computed [24, 35].

Padded Decompositions. Let us recall the definition of a *padded decomposition* (see e.g. [21, 31]). Given a finite metric space (V, d), a positive parameter $\Delta > 0$ and $\beta : V \to \mathbb{R}$, a Δ-bounded *β-padded decomposition* is a distribution Π over partitions of V such that the following conditions hold.

(a) For each partition P in the support of Π, the diameter of every cluster in P is at most Δ.

(b) If P is sampled from Π, then each ball $\mathbf{B}_x(\frac{\Delta}{\beta(x)})$ is partitioned by P with probability at most $\frac{1}{2}$.

For simplicity, say that a metric *admits β-padded decompositions* (where β is a number) if for every $\Delta > 0$ it admits a Δ-bounded β-padded decomposition. It is known that any finite metric space admits $O(\log n)$-padded decomposition [5]. Moreover, metrics of doubling dimension $\dim_V$ admit $O(\dim_V)$-padded decompositions [21]; furthermore, if a graph G excludes a K_r-minor (e.g., if it has treewidth $\leq r$), then its shortest-path metric admits $O(r^2)$-padded decompositions [28, 40, 18].

2 Embeddings with slack into ℓ_p

In this section we show that for any $\epsilon > 0$ any metric can be embedded into ℓ_p for $p \geq 1$ with ϵ-slack and distortion $O(\log \frac{1}{\epsilon})$, thus resolving one of the two main questions left open by [29].

Let us fix $\epsilon > 0$ and write $\rho_u = \rho_u(\epsilon)$. Recall that an edge uv is *ϵ-long* if $d_{uv} \geq \min(\rho_u, \rho_v)$; call it *$\epsilon$-good* if $d_{uv} \geq 4 \min(\rho_u, \rho_v)$. We partition all the ϵ-long edges into two groups, namely those which are ϵ-good and those which are not, and use a separate embedding (i.e. a separate block of coordinates) to handle each of the groups. Specifically, we handle ϵ-good edges using a Bourgain-style embedding from [29], and for the rest of the ϵ-long edges we use an auxiliary embedding such that for any edge uv the embedded uv-distance is $\Theta(\rho_u + \rho_v)$. The combined embedding has dimension $O(\log^2 \frac{1}{\epsilon})$ and achieves distortion $O(\log \frac{1}{\epsilon})$ on a set of all but an ϵ-fraction of edges.

There are several ways in which this result can be refined. Firstly, we can ask for low *ϵ-uniform-slack* distortion, and require distortion $O(\log \frac{1}{\epsilon})$ on the set of all *ϵ-long* edges; we can indeed get this, but have to boost the number of dimensions to $O(\log n \log \frac{1}{\epsilon})$. As Theorem 2.2 shows, this increase is indeed required. We note that this logarithmic increase in the number of dimensions is not the case for doubling metrics: Slivkins [42] shows how these metrics are embedded into any ℓ_p, $p \geq 1$ with ϵ-uniform slack, distortion $O(\log \frac{1}{\epsilon} \log \log \frac{1}{\epsilon})$ and dimension $(\log \frac{1}{\epsilon})^{O(\log \frac{1}{\epsilon})}$

Secondly, this embedding can be computed in a distributed *beacon-based* framework. Here a small number of nodes are selected independently and uniformly at random, and designated as *beacons*. Then the coordinates of each node are computed as a (possibly randomized) function of its distances to the beacons.

Thirdly, note that for the ϵ-slack result, the target dimension is independent of n, which suggests that this result can be extended to infinite metrics. To state such extension, let us modify the notion of slack accordingly. Following [42], let us assume that an infinite metric is equipped with a probability measure μ on nodes. This measure induces a *product measure* $\mu \times \mu$ on edges. We say that a given embedding ϕ has *distortion D with (ϵ, μ)-slack* if some set of edges of product measure at least $1 - \epsilon$ incurs distortion at most D under ϕ. Note that in the finite case, ϵ-slack coincides with (ϵ, μ)-slack when μ is the counting measure, i.e. when all nodes are weighted equally.

In the embedding algorithm, instead of selecting beacons uniformly at random (i.e. with respect to the counting measure) we select them with respect to measure μ. The proof carries over without much modification; we omit it from this version of the paper.

Theorem 2.1 *For any source metric (V, d), any target metric ℓ_p, $p \geq 1$ and any parameter $\epsilon > 0$, we give the following two $O(\log \frac{1}{\epsilon})$-distortion embeddings:*

(a) with ϵ-slack into $O(\log^2 \frac{1}{\epsilon})$ dimensions, and

(b) with ϵ-uniform slack into $O(\log n \log \frac{1}{\epsilon})$ dimensions.

These embeddings can be computed with high probability by randomized beacon-based algorithms that use, respectively, only $O(\frac{1}{\epsilon} \log \frac{1}{\epsilon})$ and $O(\frac{1}{\epsilon} \log n)$ beacons.

Proof: Let $\delta > 0$ be the desired total failure probability. The embedding algorithm is essentially the same for both parts, with one difference: we let $k = O(\log \frac{1}{\delta} + \log \frac{1}{\epsilon})$ for part (a), and $k = O(\log \frac{1}{\delta} + \log n)$ for part (b). We describe a centralized algorithm first, and prove that it indeed constructs the desired embedding. Then we show how to make this algorithm beacon-based.

We use two blocks of coordinates, of size kt and k, respectively, where $t = \lceil \log \frac{1}{\epsilon} \rceil$. The first block comes from a Bourgain-style embedding without the smaller distance scales. For each $i \in [t]$ choose k independent random subsets of V of size 2^i each, call them S_{ij}, $j \in [k]$. The first-block coordinates of a given node u are

$$f_{ij}(u) = (kt)^{-1/p}\, d(u, S_{ij}), \text{ where } i \in [t],\ j \in [k]. \quad (1)$$

For every node u and every $j \in [k]$, choose a number $\beta_{ui} \in \{-1, 1\}$ independently and uniformly at random. The second-block coordinates of u are $g_j(u) = k^{-1/p}\, \rho_u\, \beta_{uj}$, where $j \in [k]$. This completes the embedding.

For an edge uv, let $f(uv)$ and $g(uv)$ denote the ℓ_p-distance between u and v in the first and the second block of coordinates, respectively. By construction, $f(uv) \leq d_{uv}$ and $g(uv) \leq \rho_u + \rho_v$. Moreover,

$$\begin{array}{c}\text{for every } \epsilon\text{-good edge } uv,\ f(uv) \geq \Omega(d_{uv}/t) \\ \text{with failure probability at most } t/2^{\Omega(k)}. \end{array} \quad (2)$$

Indeed, fix an ϵ-good edge uv and let $d = d_{uv}$. Let α_i be the minimum of the following three quantities: $\rho_u(2^{-i})$, $\rho_v(2^{-i})$ and $d/2$. The numbers α_i are non-increasing; $\alpha_0 = d/2$. Moreover, since uv is ϵ-good we have $\alpha_t \leq \min(\rho_u, \rho_v, d/2) \leq d/4$. By a standard Bourgain-style argument it follows that for each i the event

$$\sum_j |d(u, S_{ij}) - d(v, S_{ij})| \geq \Omega(k)(\alpha_i - \alpha_{i+1})$$

happens with failure probability at most $1/2^{\Omega(k)}$. (We omit the details from this version of the paper.) Therefore, with failure probability at most $t/2^{\Omega(k)}$, this event happens for all $i \in [t]$ simultaneously, in which case

$$\begin{aligned}\sum_{ij} |d(u, S_{ij}) - d(v, S_{ij})| &\geq \sum_{i \in [t]} \Omega(k)(\alpha_i - \alpha_{i+1}) \\ &= \Omega(k)(\alpha_0 - \alpha_t) \geq \Omega(kd),\end{aligned}$$

so $f(uv) \geq \Omega(d/t)$ for the case $p = 1$. It is easy to extend this to $p > 1$ using standard inequalities. This proves the claim (2).

Furthermore, we claim that for each edge uv, $g(uv) = \Omega(\rho_u + \rho_v)$ with failure probability at most $1/2^{\Omega(k)}$. Indeed, let N_j be the indicator random variable for the event $\beta_{uj} \neq \beta_{vj}$. Since N_j's are independent and their sum N has expectation $k/2$, by Chernoff Bounds (Lemma A.1a) $N \geq k/4$ with the desired failure probability. This completes the proof of the claim.

Now fix an ϵ-long edge uv and let $d = d_{uv}$. Without loss of generality assume $\rho_u \leq \rho_v$; note that $\rho_u \leq d$. Since $B_u(\rho_u) \subset B_v(\rho_u + d)$, the cardinality of the latter ball is at least ϵn. It follows that $\rho_v \leq \rho_u + d$, so $g(uv) \leq \rho_u + \rho_v \leq 3d$. Since $f(uv) \leq d$, the embedded uv-distance is $O(d)$.

To lower-bound the embedded uv-distance, note that with failure probability at most $t/2^{\Omega(k)}$ the following happens: if edge uv is ϵ-good then this distance is $\Omega(d/t)$ due to $f(uv)$; else it is $\Omega(d)$ due to $g(uv)$. For part (a) we use Markov inequality to show that with failure probability at most δ this happens for all but an ϵ-fraction of ϵ-long edges. For part (b) we take a Union Bound to show that with failure probability at most δ this happens for *all* ϵ-long edges. This completes the proof of correctness for the centralized embedding.

It remains to provide the beacon-based version of the algorithm. Let S be the union of all sets S_{ij}. The Bourgain-style part of the algorithm depends only on distances to the $\Theta(k/\epsilon)$ nodes in S, so it can be seen as beacon-based, with all nodes in S acting as beacons. To define the second block of coordinates we need to know the ρ_u's, which we do not. However, we will estimate them using the same set S of beacons.

Fix a node u. Let B be the open ball around u of radius ρ_u, i.e. the set of all nodes v such that $d_{uv} < \rho_u$. Let B' be the smallest ball around u that contains at least $4\epsilon n$ nodes. Note that S is a set of ck/ϵ beacons chosen independently and uniformly at random, for some constant c.

In expectation at most ck beacons land in B, and at least $4ck$ beacons land in B'. By Chernoff Bounds (Lemma A.1ab) with failure probability at most $1/2^{\Omega(k)}$ the following event E_u happens: at most $2ck$ beacons land in B, and at least $2ck$ beacons land in B'. Rank the beacons according to its distance from u, and let w be the $(2ck)$-th closest beacon. Define our estimate of ρ_u as $\rho'_u = d_{uw}$. Note that if event E_u happens, then ρ'_u lies between ρ_u and $\rho_u(4\epsilon)$.

Consider a 4ϵ-good edge uv such that both E_u and E_v happen. Then (as in the non-beacon-based proof) we can upper-bound the embedded uv-distance by $O(d_{uv})$, and lower-bound it by $\Omega(d_{uv}/t)$ with high probability. For part (a) we use Markov inequality to show that with failure probability at most δ event E_u happens for all but an ϵ-fraction of nodes. For part (b) we take a Union Bound to show that with failure probability at most δ this event happens for all nodes. □

The following theorem lower-bounds the target dimension required for ϵ-uniform slack, essentially showing that in part (b) of the above theorem the dependence on $\log n$ is indeed necessary.

Theorem 2.2 *For any $\epsilon < \frac{1}{2}$ there is a metric (V, d) such that any ϵ-uniform slack embedding into l_p, $p \geq 1$ with distortion D requires $\Omega(\log_D n)$ dimensions.*

Proof: Take a clique on $\epsilon\, n$ red and $(1-\epsilon)n$ blue nodes, assign length two to each of the blue-blue edges, and assign unit lengths to all the remaining edges. Consider the metric generated by this graph. Now all the blue-blue edges are ϵ-long, and thus any distortion-D ϵ-uniform-slack embedding must maintain all the distances between the blue vertices. But this is just a uniform metric on $(1-\epsilon)n$ nodes, and the lower bound follows by a simple volume argument. $\square$

3 Embeddings with slack: a general theorem

In this section, we generalize the results of the previous section. We formulate and prove a general theorem which takes a result on classic (distortion-minimizing) embeddings of finite metrics into ℓ_p, and converts it into results on embeddings with ϵ-slack and embeddings with ϵ-uniform slack.

For embeddings with ϵ-slack, the idea is to choose a small set of nodes (*beacons*) uniformly at random, embed the beacons using the result on classic embeddings, then embed all the other points according to the nearest beacon, and add some auxiliary coordinates. To obtain embeddings with ϵ-uniform slack, for each non-beacon node instead of choosing the nearest beacon we choose the "best" beacon *in each coordinate*. In both cases, we apply the result on classic embeddings to a *subset* of the original metric. Therefore our results are only about families of metrics that are subset-closed: a family $\mathcal{X}$ of metrics is *subset-closed* if any metric in $\mathcal{X}$ restricted to any subset of nodes is also in $\mathcal{X}$. The auxiliary coordinates are similar to those in Section 2.

For the ϵ-uniform slack result we will need a technical restriction that the original classic embedding is *strongly non-expansive*. An embedding f from (V,d) into ℓ_p^k is *strongly non-expansive* if it is a contraction and of the form

$$f = (\eta_1 f_1,\, \dots\,,\, \eta_k f_k) \;\text{ and }\; \textstyle\sum_{i=1}^k \eta_i^p = 1,$$

where for any two nodes $u,v \in V$ and any coordinate i we have $|f_i(u) - f_i(v)| \le d(u,v)$.

Note that the above requirement is not so restricting, since almost every known embedding can be converted to a strongly non-expansive one. In particular, it is easy to check that any generalized Fréchet embedding (i.e., an embedding where each coordinate i is associated with a set S_i such that $f_i(u) = d(S_i,u)$) is strongly non-expansive.

Theorem 3.1 *Consider a fixed space ℓ_p, $p \ge 1$. Let $\mathcal{X}$ be a subset-closed family of finite metric spaces such that for any $n \ge 1$ and any n-point metric space $X \in \mathcal{X}$ there exists an embedding $\phi_X : X \hookrightarrow \ell_p$ with distortion $\alpha(n)$ and dimension $\beta(n)$.*

Then there exists a universal constant $C > 0$ such that for any metric space $X \in \mathcal{X}$ and any $\epsilon > 0$ we have

(a) an embedding into ℓ_p with ϵ-slack, distortion $\alpha(\frac{C}{\epsilon}\log\frac{1}{\epsilon})$ and dimension $\beta(\frac{C}{\epsilon}\log\frac{1}{\epsilon}) + C\log\frac{1}{\epsilon}$.

(b) an embedding into ℓ_p with ϵ-uniform slack, distortion $\alpha(C/\epsilon)$ and dimension $C\log(n)\,\beta(C/\epsilon)$.

In part (b) we need to assume that for all $Y \in \mathcal{X}$ the original embedding ϕ_Y is strongly non-expansive.

The most notable application of the above theorem is for of arbitrary metrics into ℓ_p, $p \ge 1$. Using Bourgain's embedding [11] of distortion $O(\log n)$ and dimension $O(\log^2 n)$, we obtain another proof of Theorem 2.1. Using a recent result of Bartal [8] which improves of the dimension in Bourgain's theorem to $O(\log n)$ we obtain an improved dimension in Thorem 2.1. Corollary 1.5 states several additional applications of Theorem 3.1. The proof follows from known upper bounds: (1) from [11], (2) from [34], where the improved dimension of (1), (2) follow from [8], (3) from [31], (4) from [3], (5) from [12].

Both embeddings in Theorem 3.1 can be cast in a distributed beacon-based framework which is similar to that in Section 2. Specifically, we have two phases. In the first phase, beacons measure distances to each other and compute an embedding ϕ for the set of beacons. In the second phase, each node computes its coordinates as a (possibly randomized) function of ϕ and its distances to beacons. We need $O(\frac{1}{\epsilon}\log\frac{1}{\epsilon})$ and $O(\frac{1}{\epsilon}\log n)$ beacons for parts (a) and (b), respectively. The necessary modifications are similar to those in Section 2; we defer the details to the full version.

We prove part (a) here and defer part (b) to Appendix B.

Proof of Theorem 3.1(a): We will choose a constant set of beacons, embed them, then embed all the other points according to the nearest beacon, and add some auxiliary coordinates.

Formally, consider some metric $X = (V,d) \in \mathcal{X}$, where V is a set of n nodes. Given $\epsilon > 0$ let $\hat{\epsilon} = \epsilon/20$, and $t = 100\log\left(\frac{1}{\hat{\epsilon}}\right)$. Let B be a uniformly distributed random set of $\frac{t}{\hat{\epsilon}}$ points in V (the beacons). Let g be a contracting embedding from B into ℓ_p with distortion $\alpha(\frac{t}{\hat{\epsilon}})$ and dimension $\beta(\frac{t}{\hat{\epsilon}})$. Let

$$\{\sigma_j(u) \mid u \in V, 1 \le j \le t\}$$

be i.i.d symmetric $\{0,1\}$-valued Bernoulli random variables. Define the following functions:

$$\begin{aligned} h_j(u) &= \sigma_j(u)\;\rho_u(\hat{\epsilon})\;t^{-1/p} \\ &\qquad \text{for all } u \in V \text{ and } j \text{ such that } 1 \le j \le t. \\ f(u) &= g(b) \quad \text{for all } u \in V, \text{ where } b \in B \\ &\qquad \text{is the beacon that is closest to } u. \end{aligned}$$

The embedding will be $\varphi = f \oplus h$, where h is the t-vector with j-th coordinate equal to $h_j(u)$. Let E be the set of all unordered node pairs, and let $G' = E \setminus (D_1 \cup D_2)$, where

$$D_1 = \{(u,v) \mid d(u,v) \le \max\{\rho_u(\hat\epsilon), \rho_v(\hat\epsilon)\}\}$$
$$D_2 = \{(u,v) \mid d(u,B) \ge \rho_u(\hat\epsilon) \vee d(v,B) \ge \rho_v(\hat\epsilon)\}.$$

Observe that $|D_1| \le \hat\epsilon n^2$. For any $u \in V$ we have

$$\Pr[\,d(u,B) \ge \rho_u(\hat\epsilon)\,] \le (1 - t/(n\hat\epsilon))^{\hat\epsilon n} \le e^{-t} \le \hat\epsilon,$$

so by Markov inequality $|D_2| \le 2\hat\epsilon n^2$ w.p. at least $1/2$. We begin with an upper bound on φ for all $(u,v) \in G'$:

$$\begin{aligned}
&\|\varphi(u) - \varphi(v)\|_p^p \\
&\quad= \|f(u) - f(v)\|_p^p + \textstyle\sum_{j=1}^t |h_j(u) - h_j(v)|^p \\
&\quad\le (3d(u,v))^p + \textstyle\sum_{j=1}^t \left|t^{-1/p}\max\{\rho_u(\hat\epsilon), \rho_v(\hat\epsilon)\} - 0\right|^p \\
&\quad\le (3^p + 1)\,(d(u,v))^p
\end{aligned}$$

We now partition G' into two sets:

$$G_1 = \{(u,v) \in G' :\ \max\{\rho_u(\hat\epsilon), \rho_v(\hat\epsilon)\} \ge d(u,v)/4\}$$

and $G_2 = G' \setminus G_1$.

Consider an edge $(u,v) \in G_1$. Without loss of generality assume $\rho_u(\hat\epsilon) \ge \rho_v(\hat\epsilon)$. Let $\mathcal{E}_j(u,v)$ be the event that $h_j(v) = 0$ and $h_j(u) = \rho_u(\hat\epsilon)t^{-1/p}$. This event happens with probability $\frac{1}{4}$. Let $A(u,v) = \sum_{j=1}^t \mathbf{1}_{\mathcal{E}_j(u,v)}$. Then $\mathbb{E}[A(u,v)] = t/4$, so using Chernoff's bound we can bound the probability that $A(u,v)$ is smaller than half it's expectation:

$$\Pr[A(u,v) \le t/8] \le e^{-t/50} \le \hat\epsilon.$$

Let $D_3 = \{(u,v) \in G_1 \mid A(u,v) \le t/8\}$ so by Markov inequality with probability at least $1/2$, $|D_3| \le 2\hat\epsilon n^2$.

Therefore, for any $(u,v) \in G_1 \setminus D_3$ we lower bound the contribution.

$$\begin{aligned}
\|\varphi(u) - \varphi(v)\|_p^p &\ge \textstyle\sum_{j=1}^t |h_j(u) - h_j(v)|^p \\
&\ge \tfrac{t}{8}\left(\rho_u(\hat\epsilon)\cdot t^{-1/p}\right)^p \ge \tfrac{1}{8}\cdot(\tfrac{1}{4}d(u,v))^p
\end{aligned}$$

For any $(u,v) \in G_2$ let b_u, b_v be the beacons such that $f(u) = g(b_u), f(v) = g(b_v)$. Due to the definition of D_2 and G_2 and from the triangle inequality it follows that

$$\begin{aligned}
d(b_u, b_v) &\ge d(u,v) - d(u,b_u) - d(v,b_v) \\
&\ge d(u,v) - d(u,v)/2 = d(u,v)/2.
\end{aligned}$$

Therefore, we lower bound the contribution of $(u,v) \in G_2$.

$$\begin{aligned}
\|\varphi(u) - \varphi(v)\|_p^p &\ge \|f(u) - f(v)\|_p^p = \|g(b_u) - g(b_v)\|_p^p \\
&\ge \frac{1}{\alpha(t/\hat\epsilon)}\cdot d(b_u,b_v) \ge \frac{d(u,v)}{2\alpha(t/\hat\epsilon)}
\end{aligned}$$

Finally note that D_2, D_3 are independent and $G = E\setminus(D_1 \cup D_2 \cup D_3)$ is the set of edges suffering the desired distortion. So with probability at least $1/4$ we have

$$|G| \ge \binom{n}{2} - 5\hat\epsilon n^2 \ge \binom{n}{2} - \epsilon n^2/4 \ge (1-\epsilon)\binom{n}{2}$$

as required. □

4 Embeddings into Trees

Probabilistic embedding of finite metric space into trees was introduced in [5]. Fakcharoenphol et al. [17] proved that finite metric space embeds into a distribution of dominating trees with distortion $O(\log n)$ (slightly improving the result of[6], other proofs can be found in [7]). In this section we exploit the technique of [17] to obtain embeddings with slack. First we show that it gives a probabilistic embedding of arbitrary metrics into tree metrics with *expected* gracefully degrading distortion $D(\epsilon) = O(\log 1/\epsilon)$. For technical convenience, we will treat n-point metrics as functions from $[n] \times [n]$ to reals. Note that all metrics d_T generated by the algorithm in [17] are *dominating*, i.e. for any edge uv we have $d(u,v) \le d_T(u,v)$.

Theorem 4.1 *For any input metric (V,d), let d_T be the dominating HST metric on V constructed by the randomized algorithm in Fakcharoenphol et al. [17]. Then the embedding from (V,d) to (V,d_T) has* expected *gracefully degrading distortion $D(\epsilon) = O(\log 1/\epsilon)$. Specifically, for any parameter $\epsilon > 0$ and any ϵ-long edge uv we have*

$$d_{uv} \le E_\varphi[d_T(u,v)] \le O(\log 1/\epsilon)\, d_{uv}. \tag{3}$$

Since tree metrics are isometrically embeddable into L_1, it follows that we can embed any metric into L_1 with gracefully degrading distortion $D(\epsilon) = O(\log \frac{1}{\epsilon})$.

Proof: For simplicity let us assume that all distances in (V,d) are distinct; otherwise we can perturb them a little bit and make them distinct, without violating the triangle inequality; see the full version of this paper for details. In what follows we will assume a working knowledge of the decomposition scheme in [17].

Let us fix the parameter $\epsilon > 0$ and an ϵ-long edge uv, and let $d = d(u,v)$. Let us assume without loss of generality that $\rho_u(\epsilon) \le \rho_v(\epsilon)$. Then $\rho_u(\epsilon) \le d$, so $|\mathbf{B}_u(d)| \le \epsilon n$.

Run the randomized algorithm of [17] to build a tree T and the associated tree metric d_T. The decomposition scheme will separate u and v at some distance scale $2^i \ge d/2$. Let Δ be the maximum distance in the input metric. Under the distribution over tree metrics d_T that is induced by the algorithm, the expected distance $E[d_T(u,v)]$ between u and v in tree T is equal to the sum

$$\textstyle\sum_{i \ge \log d - 1}^{\log\Delta} 4\cdot 2^i \times Pr[(u,v) \text{ first separated at level } 2^i].$$

Look at the sum for i such that $d/2 \le 2^i < 4d$: this is at most $48d$. By the analysis of [17], the rest of the sum, i.e. the sum for $i \ge \log 4d$, is at most

$$\textstyle\sum_{i \ge \log 4d}^{\log\Delta} 4\cdot 2^i \times \frac{2d}{2^i}\log\frac{|\mathbf{B}_u, 2^i)|}{|\mathbf{B}_u, 2^{i-2})|}$$

Since the above sum telescopes, it is at most

$$8d \cdot 2 \log (n/|\mathbf{B}_u(d)|) \leq O(d \log 1/\epsilon),$$

which proves the second inequality in (3). The first inequality in (3) holds trivially because all metrics d_T generated by the algorithm in [17] are dominating. □

The above embedding into ℓ_1 can be made algorithmic by sampling from the distribution and embedding each sampled tree into ℓ_1 using a fresh set of coordinates; however, the number of trees now needed to give a small distortion may be as large as $\Omega(n \log n)$. We will see how to obtain gracefully degrading distortion with a smaller number of dimensions in the next section.

A slightly modified analysis yields an embedding into a *single tree*; we omit the details from this version.

Theorem 4.2 *For any source metric (V, d) and any parameter $\epsilon > 0$ there exists an embedding into a dominating HST metric with ϵ-uniform slack and distortion $O(\frac{1}{\epsilon} \log \frac{1}{\epsilon})$.*

5 Low-dimensional Embeddings with Gracefully Degrading Distortion

In this section we prove our result on embeddings into ℓ_p, $p \geq 1$ with gracefully degrading distortion:

Theorem 5.1 *Consider a metric (V, d) which admits β-padded decompositions. Then it can be embedded into ℓ_p, $p \geq 1$ with $O(\log^2 n)$ dimensions and gracefully degrading distortion $O(\beta)(\log \frac{1}{\epsilon})^{1/p}$. The embedding procedure is given as a randomized algorithm which succeeds with high probability.*

The proof of this theorem builds on the well-known embedding algorithms of Bourgain [11] and Linial et al. [33], and combines ideas given in [40, 21, 29, 42, 31] with some novel ones. To the best of our understanding, the embeddings given in the previous papers do not directly give us gracefully degrading distortion, and hence the additional machinery indeed seems to be required.

Let us fix $k = O(\log n)$, where the constant will be specified later. We will construct an embedding $\varphi : V \to \ell_p$ with $7k^2$ dimensions; the coordinates of φ will be indexed by triples $(i, j, l) \in [k] \times [k] \times [7]$.

We will show how to construct the map φ in rest of this section, which has the following conceptual steps. We first define a concrete notion of "distance scales" in Section 5.1, in terms of which we can cast many previous embeddings, and specify the desired properties for the distance scales in our embedding. We then show how to construct the distance scales as well as the claimed embedding φ in Section 5.2, and show that it has gracefully degrading distortion in Section 5.3.

5.1 Distance Scales and Scale Bundles

Our algorithm, just like the algorithms in [11, 33, 40, 21, 29, 31, 32], operates on distance scales that start around the diameter of the metric, and go all the way down to the smallest distance in the metric. Informally, the embedding φ has block of coordinates for each distance scale, such that if the true uv-distance for some edge uv is within this scale, then the uv-distance in these coordinates of φ is roughly equal to the true distance. These blocks of coordinates are then combined into an embedding that works for all scales simultaneously.

Different embeddings use very different notions of distance scales; in cases like the Rao-style embeddings, there are clear coordinates for each distance that is a power of 2—but in Bourgain-style embeddings, this is not the case. To be able to give a unified picture, let us formally define a *distance scale* f to be a coordinate map $f : V \to \mathbb{R}$. A *scale bundle* $\{f_{ij}\}$ is then a collection of coordinate maps f_{ij}, such that for every fixed index j and node u, the values $f_{ij}(u)$ are *decreasing* with i.

We can now cast and interpret previous embeddings in this language: in the Bourgain-style embeddings [11, 33], $f_{ij}(u)$ is the radius of the smallest ball around u containing 2^{n-i} nodes, and hence the cardinality of $\mathbf{B}_u(f_{ij}(u))$ halves as we increase i. In the Rao-style embeddings [40, 21], the scales are $f_{ij}(u) = \text{diameter}(V)/2^i$, and hence the distance scales halve as we increase i. The measured descent embedding in [31] essentially ensures a judicious mixture of the above two properties: as we increase i, the ball $\mathbf{B}_u(f_{ij}(u))$ either halves in radius, or halves in cardinality, whichever comes first.

For our embedding, we need *both* the radius and the cardinality of $\mathbf{B}_u(f_{ij}(u))$ to halve—and hence have to define the scale-bundles accordingly. This would be easy to achieve by itself; however, to give good upper bounds on the embedded distance, we also need each distance scale to be sufficiently smooth, by which we mean that all the distance scales f_{ij} must themselves be 1-Lipschitz. In other words, we want that $|f_{ij}(u) - f_{ij}(v)| \leq d(u, v)$. The construction of the scale bundle $\{f_{ij}\}$ with both halving and smoothness properties turns out to be a bit non-trivial, the details of which are given in the next section.

5.2 The Embedding Algorithm

Let us construct the embedding for Theorem 5.1. We have not attempted to optimize the multiplicative constant for distortion, having chosen the constants for ease of exposition whilst ensuring that the proofs work.

First we will construct a *scale bundle* $\{f_{ij} : i, j \in [k]\}$. For a fixed j, the maps f_{ij} are constructed by an independent random process, inductively from $i = 0$ to $i = k - 1$.

We start with $f_{(0,j)}(\cdot)$ equal to the diameter Φ_d of the metric. Given f_{ij}, we construct $f_{(i+1,j)}$ as follows. Let U_{ij} be a random set such that each node u is included independently with probability $1/|\mathbf{B}_u(4f_{ij}(u))|$. Claim 5.8.) Define $f_{(i+1,j)}(u)$ as the minimum of $d(u, U_{ij})$ and $f_{ij}(u)/2$. This completes the construction of the scale bundle.

To proceed, let us state a lemma that captures, for our purposes, the structure of the metric.

Lemma 5.2 *Consider a source metric (V, d) which admits β-padded decompositions. Then for any 1-Lipschitz coordinate map f there is a randomized embedding g into ℓ_p, $p \geq 1$ with $t = 6$ dimensions so that*

(a) each coordinate of g is 1-Lipschitz and upper-bounded by f; and

(b) if $f(u)/d_{uv} \in [\frac{1}{4}; 4]$ for some edge uv then, with probability $\Omega(1)$, $\|g(u) - g(v)\|_p \geq \Omega(d_{uv}\, t^{1/p}/\beta)$.

Section 5.4 and Appendix C contain two different proofs of this lemma; the first one uses padded decomposition techniques from [21, 31], and the other uses some Bourgain-style ideas [11, 33] which we believe are novel and possibly of independent interest.[5]

Fix a pair $i, j \in [k]$. Apply Lemma 5.2 to the map f_{ij} and obtain a 6-dimensional embedding; denote these 6 coordinates as $g_{(i,\,j,\,l)}$, $1 \leq l \leq 6$. Let W_{ij} be a random set such that each node u is included independently with probability $1/|\mathbf{B}_u(f_{ij}(u)/2)|$. Define $g_{(i,\,j,\,0)}(u)$ as the minimum of $f_{ij}(u)$ and $d(u, W_{ij})$. Finally, we set $\varphi_{(i,\,j,\,l)} = k^{-1/p}\, g_{(i,\,j,\,l)}$.

Lemma 5.3 *The maps f_{ij}, g_{ij} and $\varphi_{(i,\,j,\,l)}$ are 1-Lipschitz.*

Proof: Indeed, $f_{(0,j)}$ is 1-Lipschitz by definition, and the inductive step follows since the min of two 1-Lipschitz maps is 1-Lipschitz. For the same reason, the maps $g_{(i,\,j,\,l)}$ are 1-Lipschitz as well, and therefore so are the maps $\varphi_{(i,\,j,\,l)}$. □

Since $k = O(\log n)$, it immediately follows that the embedded distance is at most $O(\log n)$ times the true distance. In the next section, we will prove a sharper upper bound of $O(d_{uv})(\log \frac{1}{\epsilon})^{1/p}$ for any ϵ-long edge uv, and a lower bound $\Omega(d_{uv}/\beta)$ for any edge.

5.3 Analysis

In this section, we complete the proof of Theorem 5.1 by giving bounds on the stretch and contraction of the embedding φ. The following definition will be useful: for a node u, an interval $[a, b]$ is *u-broad* if a or b is equal to d_{uv} for some v, $a \leq b/4$ and $|\mathbf{B}_u(a)| \leq \frac{1}{32}|\mathbf{B}_u(b)|$.

Let us state two lemmas that capture the useful properties of the maps f_{ij} and $g_{(i,\,j,\,0)}$, respectively: note that these properties are independent of the doubling dimension. The proofs are deferred to Section 5.5.)

Lemma 5.4 *With high probability it is the case that:*

(a) for any 1-Lipschitz maps $f'_{ij} \leq f_{ij}$ and any ϵ-long edge uv $\sum_{ij} f'_{ij}(uv) \leq O(kd_{uv} \log \frac{1}{\epsilon})$.

(b) for each node u, each u-broad interval contains values f_{ij} for $\Omega(k)$ different j's.

Lemma 5.5 *Fix edge uv and indices ij; let $R = f_{ij}(u)$ and $d = d_{uv}$. Given that $R \geq 4d$ and $|\mathbf{B}_u(d/4)| = c\,|\mathbf{B}_u(R)|$, the event $g_{(i,\,j,\,0)}(uv) \geq \Omega(d)$ happens with conditional probability $\Omega(c)$.*

Proof of Theorem 5.1: Fix an ϵ-long edge uv and let $d = d_{uv}$. Since $g_{(i,\,j,\,l)} \leq f_{ij}$ for each l, by Lemma 5.4a the embedded uv-distance is upper-bounded by $O(d \log \frac{1}{\epsilon})$ for $p = 1$; the same argument gives an upper bound of $O(d)(\log \frac{1}{\epsilon})^{1/p}$ for $p > 1$.

It remains to lower-bound the embedded uv-distance by $\Omega(d/\beta)$, where β is the parameter in Theorem 5.1 and Lemma 5.2. Denote by $g_{ij}(uv)$ the total ℓ_p-distance between u and v in the coordinates $g_{(i,\,j,\,l)}$, $l \geq 1$. Denote by $\mathcal{E}_{ij}$ the event that $g_{(i,\,j,\,0)}(uv)$ or $g_{ij}(uv)$ is at least $\Omega(d/\beta)$. It suffices to prove that with high probability events $\mathcal{E}_{ij}$ happen for at least $\Omega(k)$ (i, j)-pairs. We consider two cases, depending on whether $\rho_u(\epsilon/32) \geq d/4$.

Case (a). If $\rho_u(\epsilon/32) \geq d/4$ then the interval $I = [d/4; d]$ is u-broad, so by Lemma 5.4b there are $\Omega(k)$ different j's such that $f_{ij}(u) \in I$ for some i. By Lemma 5.2 and Chernoff bounds (Lemma A.1a) for $\Omega(k)$ of these ij pairs we have $g_{ij}(uv) \geq \Omega(d/\beta)$, case (a) complete.

Case (b). Assume $\rho_u(\epsilon/32) < d/4$; consider the interval $I = [d;\ \max[4d,\ \rho_u(32\epsilon)]]$. We claim that

$$\Pr\left[\mathcal{E}_{ij} \mid f_{ij}(u) \in I\right] \geq \Omega(1), \text{ for each } (i, j)\text{-pair.} \quad (4)$$

Indeed, fix ij and suppose $f = f_{ij}(u) \in I$. There are two cases, $f \in [d; 4d]$ and $f \in (4d; \rho_u(32\epsilon)]$. In the first case by Lemma 5.2 $g_{ij}(uv) \geq \Omega(d/\beta)$ with conditional probability at least $\Omega(1)$. In the second case

$$|\mathbf{B}_u(d/4)| \geq \epsilon n/32 \geq 2^{-10}\,(32\epsilon n) \geq 2^{-10}\,|\mathbf{B}_u(f)|,$$

so by Lemma 5.5 $g_{(i,\,j,\,0)}(uv) \geq \Omega(d)$ with conditional probability $\Omega(1)$. This proves (4). Since the interval I is u-broad, by Lemma 5.4b there are $\Omega(k)$ different j's such that $f_{ij}(u) \in I$ for some i. Since for different j's the events in (4) are independent, case (b) follows by Chernoff bounds (Lemma A.1a). □

[5] More precisely, the second proof is for the important special case when β is the doubling dimension. In this proof the target dimension becomes $t = O(\beta \log \beta)$, which results in target dimension $O(\log^2 n)(\beta \log \beta)$ in Theorem 5.1.

5.4 Analysis: proof of Lemma 5.2

In this section we use padded decomposition techniques from [21, 31] to prove Lemma 5.2. Let us recall the definitions of a *padded decomposition* and a *decomposition bundle* [21, 31].

Definition 5.6 *Given a finite metric space (V, d), a positive parameter $\Delta > 0$ and $\beta : V \to \mathbb{R}$, a Δ-bounded β-*padded decomposition *is a distribution Π over partitions of V such that the following conditions hold.*

(a) For each partition P in the support of Π, the diameter of every cluster in P is at most Δ.

(b) If P is sampled from Π, then each ball $\mathbf{B}_x(\frac{\Delta}{\beta(x)})$ is partitioned by P with probability at most $\frac{1}{2}$.

*Given a function $\beta : V \times \mathbb{Z} \to \mathbb{R}$, a β-*padded decomposition bundle *on V is a set of padded decompositions $\{\eta(i) : i \in \mathbb{Z}\}$ such that each $\eta(i)$ is a 2^i-bounded $\beta(\cdot, i)$-padded decomposition of V.*

If a metric admits a β-padded decomposition bundle such that β is constant, we simply say that this metric *admits β-padded decompositions*.

The randomized construction. Let η be a β-padded decomposition bundle. For each $u \in \mathbb{Z}$, let the decomposition P_u be chosen according to the distribution $\eta(u)$. We denote $P_u(x)$ to be the unique cluster in P_u containing x.

Moreover, for $u \in \mathbb{Z}$, let $\{\sigma_u(C) : C \subseteq V\}$ be i.i.d. unbiased $\{0,1\}$-random variables. Let $T = \{0, 1, \ldots, 5\}$. Let $u(x) := \lceil \log_2 f(x) \rceil$. For each $t \in T$, we define a (random) subset

$$W^t := \{x \in V : \sigma_{u(x)-t}(P_{u(x)-t}(x)) = 0\}, \qquad (5)$$

from which we obtain $g_t(\cdot) = \min\{d(\cdot, W^t), f(\cdot)\}$.

Bounding the contraction of the embedding. We fix vertices $x, y \in V$ and let $d = d(x, y)$. Consider the embedded distance between them. The aim is to show that under some condition, there exists t such that $|g_t(x) - g_t(y)| \geq \rho d$ happens with constant probability, where ρ depends on the β-padded decomposition bundle.

Lemma 5.7 *Suppose $f(x) \in [\frac{d}{4}, 4d]$ and $t \in T$ is the integer such that $\hat{u} := u(x) - t$ satisfies $2^{\hat{u}} \in [d/8, d/4)$. Let $J := \{-1, 0, 1\}$ and $\rho := \min\{\frac{1}{32\beta(x,u)} : u \in \hat{u} + J\}$. Then the event $|g_t(x) - g_t(y)| \geq \rho d$ happens with probability at least 1/64.*

Proof: Consider the random process that determine the coordinate g_t. We like to show that the union of the following two disjoint events happens with constant probability, which implies our goal. There are two cases:

Case 1 The set W^t contains x but is disjoint with $B_y(\rho d)$.

Case 2 The set W^t contains no points from $B_x(2\rho d)$ but at least one point from $B_y(\rho d)$.

Let us define the following auxiliary events.

- Event $\mathcal{E}_1$ occurs when x is contained in W^t.
- Event $\mathcal{E}_2$ occurs when W^t is disjoint with $B_y(\rho d)$.
- Event $\mathcal{E}_3$ occurs when for all $z \in B_x(2\rho d)$ and $u \in \hat{u} + J$, x and z are in the same cluster in $\eta(u)$.
- Event $\mathcal{E}_4$ occurs if for all $u \in \hat{u} + J$, $\sigma_u(P_u(x)) = 1$.

Observe that the event $\mathcal{E}_1 \cap \mathcal{E}_2$ implies the event in Case 1. Note that given a decomposition $\eta(\hat{u})$, the point x lies in a cluster different from those intersecting $B_y(\rho d)$, because $2^{\hat{u}} < \frac{d}{4} < (1 - \rho)d$. Hence the events $\mathcal{E}_1$ and $\mathcal{E}_2$ are conditionally independent, given $\eta(\hat{u})$; this in turn implies that

$$\begin{aligned} Pr\left[\mathcal{E}_1 \cap \mathcal{E}_2 | \, \eta(\hat{u})\right] &= Pr\left[\mathcal{E}_1 | \, \eta(\hat{u})\right] \; Pr\left[\mathcal{E}_2 | \, \eta(\hat{u})\right] \\ &= \frac{1}{2} Pr\left[\mathcal{E}_2 | \, \eta(\hat{u})\right]. \end{aligned}$$

Since this fact holds for all decompositions $\eta(\hat{u})$, it follows that $Pr[\mathcal{E}_1 \cap \mathcal{E}_2] = \frac{1}{2} Pr[\mathcal{E}_2]$

Observe that the event $\mathcal{E}_3 \cap \mathcal{E}_4 \cap \overline{\mathcal{E}_2}$ implies the event in Case 2. This follows from the fact that $|u(x) - u(z)| \in J$. Since $f(x) \geq \frac{d}{4}$, f is 1-Lipschitz and $d(x, z) \leq 2\rho d \leq \frac{d}{8}$, it follows $f(x)$ and $f(z)$ are within a multiplicative factor of 2 from each other. Hence $u(x)$ and $u(z)$ differ by at most one. Again, given the decompositions $\eta(u)$, $u \in \hat{u} + J$, the event $\mathcal{E}_4$ is independent of the event $\mathcal{E}_3 \cap \overline{\mathcal{E}_2}$. Hence,

$$\begin{aligned} Pr\left[\mathcal{E}_3 \cap \mathcal{E}_4 \cap \overline{\mathcal{E}_2}\right] &= Pr\left[\mathcal{E}_4\right] \; Pr\left[\mathcal{E}_3 \cap \overline{\mathcal{E}_2}\right] \\ &= \tfrac{1}{8} Pr\left[\mathcal{E}_3 \cap \overline{\mathcal{E}_2}\right]. \end{aligned}$$

Finally, it follows that the union of the events in cases 1 and 2 happens with probability at least

$$\begin{aligned} &\tfrac{1}{2} Pr[\mathcal{E}_2] + \tfrac{1}{8} Pr[\mathcal{E}_3 \cap \overline{\mathcal{E}_2}] \\ &\geq \tfrac{1}{8} Pr[\mathcal{E}_3 \cap \mathcal{E}_2] + \tfrac{1}{8} Pr[\mathcal{E}_3 \cap \overline{\mathcal{E}_2}] = \tfrac{1}{8} Pr[\mathcal{E}_3]. \end{aligned}$$

In order to show that $\mathcal{E}_3$ happens with constant probability, we make use of the properties of β-padded decomposition bundle. Since for all $u \in \hat{u} + J$ we have

$$2\rho d \leq 2/32\beta(x, u) \cdot d \leq 2^u/\beta(x, u),$$

it follows that $\mathcal{E}_3$ happens with probability at least 1/8. Therefore, it follows the desired event happens with probability at least 1/64. $\square$

5.5 Analysis: maps f_{ij} and $g_{(i,j,0)}$

Here we prove Lemma 5.4 and Lemma 5.5. First we prove part (a) of Lemma 5.4, which is essentially the upper bound on the embedded distance for the case $p = 1$. We start with a local smoothness property of the sets U_{ij}.

Claim 5.8 *Fix $i, j \in [k]$ and an edge uv. Condition on the map f_{ij}, i.e. pause our embedding algorithm right after f_{ij} is constructed; let $r = f_{ij}(u)$. If $d_{uv} \leq r/4$ then*

$$\Pr[v \in U_{ij}] \leq 1/|\mathbf{B}_u(r)| \leq \Pr\left[v \in U_{(i+3,j)}\right].$$

Proof: Let $B = \mathbf{B}_u(r)$. For the RHS inequality, letting $r' = f_{(i+3,j)}(v)$ we have

$$4r' \leq f_{ij}(v)/2 \leq (r + d_{uv})/2 \leq 17r/32,$$

so $d_{uv} + 4r' < r$. It follows that $\mathbf{B}_v(r') \subset B$, so $v \in U_{(i+3,j)}$ with probability $1/|\mathbf{B}_v(4r')| \geq |B|$.

For the LHS inequality, letting $r' = f_{ij}(v)$ we have

$$4r' \geq 4(r - d_{uv}) \geq r + d_{uv},$$

so $B \subset \mathbf{B}_v(4r')$. Therefore $v \in U_{ij}$ with probability $1/|\mathbf{B}_v(4r')| \leq 1/|B|$. □

Fix a node u; for simplicity assume $k = 4k_0 + 1$. Let $B_{ij} = \mathbf{B}_u(f_{ij})$ and let X_{ij} be the indicator random variable for the event that $|B_{(4i+4,\,j)}| \leq |B_{(4i,\,j)}|/2$. Note that for a fixed j, the random variables X_{ij} are not independent. However, we can show that given all previous history, the ij-th event happens with at least a constant probability.

Claim 5.9 *For each $i \in [k_0]$, $j \in [k]$ and $q = 1 - e^{-1/2}$ we have $Pr[X_{ij} = 1 \mid f_{lj}, l < i] \geq q$.*

Proof: Indeed, fix ij, let $f = f_{(4i,j)}(u)$ and $f' = f_{(4i+4,j)}(u)$, and let $B = \mathbf{B}_u(r)$ be the smallest ball around u that contains at least $|B_{(4i,\,j)}|/2$ nodes. Clearly, $X_{ij} = 1$ if and only if $f' \leq r$. By definition of f_{ij}'s we have $f' \leq f/16$, so we are done if $r \geq f/16$. Else by Claim 5.8 any node $v \in B$ included into the set $U_{(4i+3,j)}$ with probability at least $1/2|B|$, so the probability of including at least one node in B into this set (in which case $f' \leq r$) is at least $1 - (1 - 1/2|B|)^{|B|} \geq q$. □

For a random variable X define the *distribution function* $F_X(t) = \Pr[X < t]$. For two random variables X and Y, say Y *stochastically dominates* X (written as $Y \succeq X$, or $X \preceq Y$) if $F_Y(t) \leq F_X(t)$ for all $t \in \mathbb{R}$. Note that if $X \geq Y$ then $X \succeq Y$. Consider a sequence of i.i.d. Bernoulli random variables $\{Y_i\}$ with success probability q. By Claim 5.9 and Lemma A.3 (proved in Section A) we have the following:

$$\sum_{i=0}^{t} X_{ij} \succeq \sum_{i=0}^{t} Y_i, \text{ for any } t \in [k_0] \text{ and each } j \in [k]. \quad (6)$$

We'll use (6) to prove the following crucial claim:

Claim 5.10 *Fix $\epsilon > 0$; for each j let T_j be the smallest i such that $f_{ij}(u) \leq \rho_u(\epsilon)$, or k if no such i exists. Then $\sum_j T_j = O(k \log \frac{1}{\epsilon})$ with high probability.*

Proof: Let $\alpha = \lceil \log \frac{1}{\epsilon} \rceil$. Let L_j be the smallest t such that $\sum_{i=0}^{t} X_{ij} \geq \alpha$, or k_0 if such t does not exist; note that $T_j \leq 4L_j$. For the sequence $\{Y_i\}$, let Z_r be the number of trials between the $(r-1)$-th success and the r-th success. Let $A_j = \sum_{r=(j-1)\alpha+1}^{j\alpha} Z_r$ and $Z = \sum_{r=1}^{k\alpha} Z_r$. By (6) for any integer $t \in [k_0]$

$$\begin{aligned} \Pr[L_j > t] &= \Pr\left[\textstyle\sum_{i=0}^{t} X_{ij} < \alpha\right] \leq \Pr\left[\textstyle\sum_{i=0}^{t} Y_i < \alpha\right] \\ &= \Pr\left[\textstyle\sum_{r=1}^{\alpha} Z_r > t\right] = \Pr[A_1 > t] \end{aligned} \quad (7)$$

Since $\{A_j\}$ are i.i.d., by (7) and Lemma A.2 it follows that $\sum_j L_j \succeq \sum_j A_j = Z$. Therefore by Lemma A.4

$$\begin{aligned} \Pr\left[\sum T_j > 8k\alpha/q\right] &\leq \Pr\left[\sum L_j > 2k\alpha/q\right] \\ &\leq \Pr[Z > 2k\alpha/q] < (0.782)^{k\alpha}, \end{aligned}$$

which is at most $1/n^3$ when $k = O(\log n)$ with large enough constant. □

Now we have all tools to prove Lemma 5.4a.

Proof of Lemma 5.4a: Use $T_j = T_j(u)$ from Claim 5.10. Fix some ϵ-long edge uv and let $d = d_{uv}$. Let $t_j = \max(T_j(u), T_j(v))$. Then since by the 1-Lipschitz property $f'_{ij}(uv) \leq d$ for all ij; moreover, for any ij such that $i \geq t_j$ both $f_{ij}(u)$ and $f_{ij}(v)$ are at most $d/2^{i-t_j}$. Then $f'_{ij}(uv)$ is at most twice that much (since $f'_{ij} \leq f_{ij}$), so taking the sum of the geometric series we see that

$$\begin{aligned} \sum_{ij} f'_{ij}(uv) &\leq \textstyle\sum_j \left(dt_j + \sum_{i \geq t_j} d/2^{i-t_j}\right) \\ &\leq \textstyle\sum_j O(dt_j) = O\left(kd \log \frac{1}{\epsilon}\right), \end{aligned}$$

where the last inequality follows by Claim 5.10. □

To prove part (b) Lemma 5.4, let us recall the definition of a u-broad interval: for a node u, an interval $[a, b]$ is *u-broad* if a or b is equal to d_{uv} for some v, $a \leq b/4$ and $|\mathbf{B}_u(a)| \leq \frac{1}{32}|\mathbf{B}_u(b)|$.

Proof of Lemma 5.4b: It suffices to consider the u-broad intervals $[a, b]$ such that one of the endpoints is equal to d_{uv} for some v, and the other is the largest b or the smallest a, respectively, such that the interval is u-broad. Call these intervals *u-interesting*; note that there are at most $2n$ such intervals for each u.

Fix node u and a u-broad interval $I = [a, b]$, fix j and let $r_i = f_{ij}(u)$. It suffices to show that with constant probability some r_i lands in I. Indeed, then we can use Chernoff bounds (Lemma A.1a), and then we can take the Union Bound over all nodes u and all u-interesting intervals.

Denote by $\mathcal{E}_i$ the event that $r_i > b$ and $r_{i+1} < a$; note that these events are disjoint. Since some r_i lands in I if and only if none of the $\mathcal{E}_i$'s happen, we need to bound the probability of $\cup\mathcal{E}_i$ away from 1.

For each integer $l \geq 0$ define the interval

$$I_l = \left[\rho_u\left(\epsilon\, 2^l\right),\, \rho_u\left(\epsilon\, 2^{l+1}\right)\right), \text{ where } \epsilon n = |\mathbf{B}_u(b)|.$$

For each $\alpha \in \{0, 1, 2, 3\}$ let $N_{(l,\alpha)}$ be the number of i's such that $r_{4i+\alpha} \in I_l$. We claim that $E[N_{(l,\alpha)}] \leq 1/q$.

Consider the case $\alpha = 0$; other cases are similar. Let $N_l = N_{(l,\alpha)}$ and suppose $N_l \geq 1$. Let i_0 be the smallest i such that $r_{4i} \leq I_l$. Then $N_l \geq t$ implies $X_{ij} = 0$ for each $i \in [i_0;\, i_0 + t - 2]$. Recall that the construction of the maps f_{ij} starts with $f_{(0,j)}$. Given the specific map $f = f_{(i_0,j)}$, the construction of the maps f_{ij}, $i > i_0$ is equivalent to a similarly defined construction that starts with $f_{(i_0,j)} = f$. Therefore, by (6) (applied to this modified construction) we have

$$\begin{aligned}
\Pr[N_l \geq t] &\leq \Pr\left[\textstyle\sum_{\beta=0}^{t-2} X_{(i_0+\beta,\, j)} = 0\right] \\
&\leq \Pr\left[\textstyle\sum_{\beta=0}^{t-2} Y_\beta = 0\right] = (1-q)^{t-1}; \\
E[N_l] &= \textstyle\sum_{t=1}^{\infty} \Pr\left[N_l \geq t\right] \leq \sum_{t=1}^{\infty} (1-q)^{t-1} = \frac{1}{q},
\end{aligned}$$

claim proved. For simplicity assume $k = 4k_0 + 1$; it follows that

$$\begin{aligned}
\textstyle\sum_{i=0}^{k-1} \Pr[r_i \in I_l] &= \textstyle\sum_{\alpha=0}^{3} \sum_{i=0}^{k_0-1} \Pr[r_{4i+\alpha} \in I_l] \\
&= \textstyle\sum_{\alpha=0}^{3} E\left[N_{(l,\alpha)}\right] \leq 4/q \qquad (8)
\end{aligned}$$

By Claim 5.8 if $r_i \in I_l$ then $r_{i+1} \leq a$ with conditional probability at most $|\mathbf{B}_u(a)|/|\mathbf{B}_u(r_u)| \leq 2^{-l}/32$. Therefore, $\Pr[\mathcal{E}_i \,|\, r_i \in I_l] \leq 2^{-l}/32$. By (8) it follows that

$$\begin{aligned}
\Pr[\cup\mathcal{E}_i] &= \textstyle\sum_{i=0}^{k-1} \Pr[\mathcal{E}_i] = \sum_{i=0}^{k-1} \sum_{l=0}^{\infty} \Pr\left[r_i \in I_l \text{ and } \mathcal{E}_i\right] \\
&\leq \textstyle\sum_{i=0}^{k-1} \sum_{l=0}^{\infty} \Pr[r_i \in I_l] \times 2^{-l}/32 \\
&= \textstyle\frac{1}{32} \sum_{l=0}^{\infty} 2^{-l} \sum_{i=0}^{k-1} \Pr[r_i \in I_l] \\
&\leq \textstyle\frac{1}{8q} \sum_{l=0}^{\infty} 2^{-l} = \frac{1}{4q} < 1,
\end{aligned}$$

so some r_i lands in I with at least a constant probability. □

It remains to prove Lemma 5.5 about the maps $g_{(i,\,j,\,0)}$.

Proof of Lemma 5.5: Let's pause our embedding algorithm right after the map f_{ij} is chosen, and consider the probability space induced by the forthcoming random choices. Let $X_w = f_{ij}(w)$. First we claim that

$$\Pr\left[g_{(i,\,j,\,0)}(u) \leq r \,|\, r \leq X/8\right] \geq \Omega(\beta_r), \qquad (9)$$

where $\beta_r = |\mathbf{B}_u(r)|/|\mathbf{B}_u(X)|$. Indeed, suppose $r \leq X/8$, let $B = \mathbf{B}_u(r)$ and consider any $w \in B$. Then by (12):

$$\begin{aligned}
\Pr[w \in W_{ij}] &= 1/|\mathbf{B}_w(X_w/2)| \\
&\geq 1/|\mathbf{B}_u(X)| \geq \beta_r |B| \\
\Pr\left[\, g_{(i,\,j,\,0)}(u) \leq r\,\right] &= \Pr[W_{ij} \text{ hits } B] \\
&\geq 1 - (1 - \beta_r |B|)^{|B|} \\
&\geq 1 - e^{-\beta_r} \geq \Omega(\beta_r),
\end{aligned}$$

proving (9). Now let $B = \mathbf{B}_v(X_v/8)$; then by (12) any $w \in B$ is included into the set W_{ij} with probability at most $1/B$, so

$$\begin{aligned}
&\Pr\left[\, g_{(i,\,j,\,0)}(v) \geq X_v/8\right] = \Pr[W_{ij} \text{ misses } B] \\
&\quad \geq (1 - 1/|B|)^{|B|} \geq 1/4. \qquad (10)
\end{aligned}$$

Finally, let's combine (9) and (10) to prove the claim. Let $r = d/4$ and suppose $X \geq 4d$. Since $X_v \geq X - d_{uv} \geq 3d$, by (10) event $g_{(i,\,j,\,0)}(v) \geq 3d/8$ happens with probability at least $1/4$. This event and the one in (9) are independent since they depend only on what happens in the balls $\mathbf{B}_u(d/4)$ and $\mathbf{B}_v(3d/8)$, respectively, which are disjoint. Therefore with probability at least $\Omega(\beta_r)$ both events happen, in which case $g_{(i,j,0)}(uv) \geq d/8$. □

6 Lower Bounds on Embeddings with Slack

In this section, we describe a general technique to derive lower bounds for ϵ-slack embeddings from lower bounds for ordinary embeddings. The bounds obtained by this technique are given in Corollary 6.5, the most notable of which is the lower bound of $\Omega\left(\log(\frac{1}{\epsilon})/p\right)$ for embedding any metric into ℓ_p.

We make use of the following definitions from [10]

Definition 6.1 *Let H be a metric space, assume we have a collection of disjoint metric spaces C_x associated with the elements x of H, and let $\mathcal{C} = \{C_x\}_{x \in H}$. The β-composition of H and $\mathcal{C}$, for $\beta \geq \frac{1}{2}$, denoted $\hat{H} = \mathcal{C}_\beta[H]$, is a metric space on the disjoint union $\dot{\bigcup}_x C_x$. Distances in $\mathcal{C}$ are defined as follows: let $x, y \in H$ and $u \in C_x, v \in C_y$, then:*

$$d_{\hat{H}}(u, v) = \begin{cases} \gamma^{-1} d_{C_x}(u, v) & x = y \\ \beta d_H(x, y) & x \neq y \end{cases}$$

where $\gamma = \frac{\max_{x \in H} diam(C_x)}{\min_{x,y \in H} d_H(x,y)}$, guarantees that $\hat{H}$ is indeed a metric space.

Definition 6.2 *Given a family $\mathcal{H}$ of metric spaces, consider $comp_\beta(\mathcal{H})$, its closure under β'-composition for $\beta' \geq \beta$. $\mathcal{H}$ is called* nearly closed under composition *if for every $\delta > 0$ there exists some $\beta \geq 1/2$, such that for every $M \in comp_\beta(\mathcal{H})$ there is $\tilde{M} \in \mathcal{H}$ and an embedding of M into $\tilde{M}$ with distortion at most $1 + \delta$.*

Remark Among the families nearly closed under composition we can find the following [10]: Trees, planar graphs, minor-excluded graphs, normed spaces. In the special case that each of the composed metrics C_x is of equal size, also doubling metrics are closed under composition.

Theorem 6.3 *Let $\mathcal{X}$ be a family of target metric spaces. Suppose $\mathcal{H}$ is a family of metrics nearly closed under composition such that for each k, $H_k \in \mathcal{H}$ has k points and any embedding of H_k into $\mathcal{X}$ has distortion at least $D(k)$. Then for arbitrarily small $\epsilon > 0$, there exists $H' \in \mathcal{H}$ such that the embedding of H' into $\mathcal{X}$ has ϵ-slack distortion at least $\Omega(D(\frac{1}{3\sqrt{\epsilon}}))$.*

Remark We can take $\mathcal{X}$ to be a family of metrics in ℓ_p with limited dimension, thus obtaining ϵ-slack lower bounds on the dimension in terms of ϵ as well.
Note that this result can be used to translate, e.g., the Brinkman and Charikar [14] lower bound for dimensionality reduction in ℓ_1 into the realm of ϵ-slack as well.

Let us now prove Theorem 6.3, first we show how to construct a family of metric spaces with the desired properties. Suppose $H \in \mathcal{H}$ is a metric such that $|H| = k$. Moreover, H embeds into $\mathcal{X}$ with distortion at least D. For any n that is a multiple of $3k$, we can define a metric $\hat{H}$ with n points in the following way.

Let $\mathcal{C} = \{C_x\}_{x\in H}$ where each $C_x \in \mathcal{H}$ is in size $\frac{n}{k}$, and let $\hat{H} = \mathcal{C}_\beta[H]$ be its β-composition for $\beta \geq D$ satisfying that $\hat{H}$ can be embedded into $\mathcal{H}$ with distortion 2.

We now proceed to the proof; indeed, the following lemma implies Theorem 6.3.

Lemma 6.4 *Let $\hat{H}$ be the metric defined as above. Let $\epsilon := 1/9k^2$. Then, $\hat{H}$ embeds into $\mathcal{X}$ with ϵ-slack distortion at least $D/4$.*

Remark. If we wanted to obtain lower bounds for ϵ-*uniform slack* embeddings instead of just for ϵ-slack embeddings, we would set $\epsilon = 1/3k$, since the number of ignored edges incident on any node is at most ϵn by the very definition; the rest of the proof remains essentially unchanged.

Proof: Suppose, on the contrary, φ is an embedding of $\hat{H}$ into $\mathcal{X}$ with ϵ-slack distortion $R < D/4$ that ignores the set E of edges. Then consider a subset $T \subseteq \hat{H}$ containing all $v \in \hat{H}$ such that v intersects at most $\sqrt{\epsilon}n$ edges in E, by a simple counting argument $|T| \geq (1 - \sqrt{\epsilon})n$.

For each $x \in H$, the set C_x contains $\frac{n}{k} = 3\sqrt{\epsilon}n$ points and hence there exists some point in $T \cap C_x$, which we call v_x.

Let $x, y \in H$. Since v_x and v_y are in T, each of them has at most $\sqrt{\epsilon}n$ neighbors. Observing that $|C_y| = 3\sqrt{\epsilon}n$, it follows that there exists a point $t \in C_y$ such that neither $\{v_x, t\}$ nor $\{v_y, t\}$ is contained in E. We can assume that for $\{u, v\} \notin E$, $d_{\hat{H}}(u, v) \leq ||\varphi(u) - \varphi(v)|| \leq R d_{\hat{H}}(u, v)$.

Hence, it follows that

$$\begin{aligned}
\|\varphi(v_x) - \varphi(v_y)\| \leq& \|\varphi(v_x) - \varphi(t)\| + \|\varphi(t) - \varphi(v_y)\| \\
\leq& R(d_{\hat{H}}(v_x, t) + d_{\hat{H}}(t, v_y)) \\
=& R\left(\beta d_H(x, y) + \frac{d_{C_y}(t, v_y)}{\gamma}\right) \\
\leq& R\left(\beta d_H(x, y) + \min_{u,v\in H} d_H(u, v)\right) \\
\leq& 2R\beta d_H(x, y) = 2R d_{\hat{H}}(v_x, v_y)
\end{aligned}$$

since $\frac{d_{C_y}(t,v_y)}{\gamma} \leq \min_{u,v\in H} d_H(u, v)$. Similarly,

$$\begin{aligned}
\|\varphi(v_x) - \varphi(v_y)\| \geq& \|\varphi(v_x) - \varphi(t)\| - \|\varphi(t) - \varphi(v_y)\| \\
\geq& d_{\hat{H}}(v_x, t) - R d_{\hat{H}}(t, v_y) \\
=& \beta d_H(x, y) - R\frac{d_{C_y}(t, v_y)}{\gamma} \\
\geq& \beta d_H(x, y) - \frac{\beta}{4}\min_{u,v\in H} d_H(u, v) \\
\geq& \frac{\beta}{2} d_H(x, y) = \frac{1}{2} d_{\hat{H}}(v_x, v_y)
\end{aligned}$$

using $R < D/4 \leq \beta/4$. Notice that the metric induced on the set $\{v_x\}_{x\in H}$ is isomorphic (up to scaling) to H, therefore φ embeds H into $\mathcal{X}$ with distortion at most $4R < D$, and we obtain the desired contradiction. □

To finish the proof of Theorem 6.3 it remains to notice that $\hat{H}$ 2-embeds into some $H' \in \mathcal{H}$, therefore H' embeds into $\mathcal{X}$ with ϵ-slack distortion at least $D/8 = \Omega(D(\frac{1}{3\sqrt{\epsilon}}))$.

Contracting Embeddings with Slack. Let us mention that allowing arbitrary *expansions* is crucial to our results: if we insisted that *none of the pairwise distances should increase*, the lower bound of $\Omega(\frac{1}{p}\log n)$ distortion [34] for embeddings into ℓ_p holds even with ϵ-slack; the simple details are deferred to the full version of this paper.

Corollary 6.5 *For any $1/n < \epsilon < 1$*

1. $\Omega\left(\frac{\log\left(\frac{1}{\epsilon}\right)}{p}\right)$ *distortion for ϵ-slack embedding into ℓ_p.*
2. *Any ϵ-slack embedding with distortion α into ℓ_p requires dimension $\Omega(\log_\alpha \frac{1}{\epsilon})$.*
3. $\Omega(\frac{1}{\sqrt{\epsilon}})$ *distortion for ϵ-slack embedding into trees.*

4. *$\Omega(\frac{1}{\epsilon})$ distortion for uniform ϵ-slack embedding into trees.*

5. *$\Omega\left(\log\left(\frac{1}{\epsilon}\right)\right)$ distortion in randomized ϵ-slack embedding into distribution of trees.*

6. *$\Omega(\sqrt{\log(1/\epsilon)})$ distortion for ϵ-slack embedding of doubling metrics into l_2.*

7. *$\Omega(\sqrt{\log(1/\epsilon)})$ distortion for ϵ-slack embedding of l_1 into l_2.*

8. *$\Omega(\sqrt{\log\log(1/\epsilon)})$ distortion for ϵ-slack embedding of trees into l_2.*

This follows from known lower bounds: (1) from [34], (2) from equilateral dimension considerations, (3) and (4) from [39], (5) from [5], (6) from [21], (7) from [16] and (8) from [12].

7 Extensions and Further Directions

The main question left open by this work is whether every metric admits a low-dimensional embedding into ℓ_p, $p \geq 1$ with gracefully degrading distortion $D(\epsilon)$. This has been answered affirmatively in Abraham, Bartal and Neiman [1], with $D(\epsilon) = O(\log \frac{1}{\epsilon})$ and dimension $O(\log n)$, using a new type of more advanced metric decompositions.

For specific families of metrics it is still interesting to provide embeddings into ℓ_p with gracefully degrading distortion $D(\epsilon) = o(\log \frac{1}{\epsilon})$; recall that Theorem 5.1 gives such embedding for decomposable metrics. In particular, we would like to ask this question for embedding arbitrary subsets of ℓ_1 into ℓ_2.

References

[1] I. Abraham, Y. Bartal, and O. Neiman. On embedding of finite metric spaces into Hilbert space, 2005. Manuscript.

[2] N. Alon, R. M. Karp, D. Peleg, and D. West. A graph-theoretic game and its application to the k-server problem. *SIAM J. Comput.*, 24(1), 1995.

[3] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut. In *37th STOC*, 2005.

[4] P. Assouad. Plongements lipschitziens dans $\mathbf{R}^n$. *Bull. Soc. Math. France*, 111(4), 1983.

[5] Y. Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *37th FOCS*, 1996.

[6] Y. Bartal. On approximating arbitrary metrics by tree metrics. In *30th STOC*, pages 183–193, 1998.

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

[8] Y. Bartal. On embedding finite metric spaces into low-dimensional normed spaces. Manuscript, 2005.

[9] Y. Bartal, B. Bollobás, and M. Mendel. Ramsey-type theorems for metric spaces with applications to online problems, 2002. To appear in Special issue of Journal of Computer and System Science.

[10] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric ramsey-type phenomena. *Annals Math*, 2003. To appear.

[11] J. Bourgain. On Lipschitz embeddings of finite metric spaces in Hilbert space. *Israel J. of Mathematics*, 52(1-2), 1985.

[12] J. Bourgain. The metrical interpretation of superreflexivity in Banach spaces. *Israel J. Math.*, 56(2):222–230, 1986.

[13] J. Bourgain, T. Figiel, and V. Milman. On Hilbertian subsets of finite metric spaces. *Israel J. Math.*, 55(2):147–152, 1986.

[14] B. Brinkman and M. S. Charikar. On the impossibility of dimension reduction in l1. In *44th FOCS*, 2003.

[15] F. Dabek, R. Cox, F. Kaashoek, and R. Morris. Vivaldi: A decentralized network coordinate system. In *ACM SIGCOMM*, 2004.

[16] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8:103–105 (1969), 1969.

[17] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *STOC*, 2003.

[18] J. Fakcharoenphol and K. Talwar. An improved decomposition theorem for graphs excluding a fixed minor. *RANDOM-APPROX*, 2003.

[19] M. Fomenkov, k. claffy, B. Huffaker, and D. Moore. Macroscopic Internet topology and performance measurements from the DNS root name servers. In *Usenix LISA*, 2001.

[20] P. Francis, S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt, and L. Zhang. IDMaps: A global Internet host distance estimation service. *IEEE/ACM Transactions on Networking*, 9:525–540, October 2001.

[21] A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low–distortion embeddings. In *44th FOCS*, 2003.

[22] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and ℓ_1-embeddings of graphs. *Combinatorica*, 24(2), 2004.

[23] J. Guyton and M. Schwartz. Locating nearby copies of replicated Internet servers. In *ACM SIGCOMM*, 1995.

[24] J. Heinonen. *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, 2001.

[25] P. Indyk. Algorithmic aspects of geometric embeddings. In *42nd FOCS*, 2001.

[26] P. Indyk and J. Matoušek. Low-distortion embeddings of finite metric spaces. In J. E. Goodman and J. O'Rourke, editors, *Handbook of Discrete and Computational Geometry*, Discrete Mathematics and its Applications (Boca Raton), chapter 8. Chapman & Hall/CRC, Boca Raton, FL, second edition, 2004.

[27] D. R. Karger and M. Ruhl. Finding nearest neighbors in growth-restricted metrics. In *34th Annual ACM Symposium on the Theory of Computing*, 2002.

[28] P. Klein, S. A. Plotkin, and S. B. Rao. Excluded minors, network decomposition, and multicommodity flow. In *25th STOC*, 1993.

[29] J. M. Kleinberg, A. Slivkins, and T. Wexler. Triangulation and embedding using small sets of beacons. In *45th FOCS*, 2004.

[30] C. Kommareddy, N. Shankar, and B. Bhattacharjee. Finding close friends on the Internet. In *12th IEEE Intl. Conf. on Network Protocols (INCP)*, 2001.

[31] R. Krauthgamer, J. Lee, M. Mendel, and A. Naor. Measured descent: A new embedding method for finite metrics. In *45th FOCS*, 2004.

[32] J. R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *16th SODA*, 2005.

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

[34] J. Matoušek. On embedding expanders into l_p spaces. *Israel J. Math.*, 102, 1997.

[35] M. Mendel and S. Har-Peled. Fast construction of nets in low dimensional metrics, and their applications. In *21st SoCG*, 2005.

[36] T. Ng and H. Zhang. Predicting Internet network distance with coordinates-based approaches. In *IEEE INFOCOM*, 2002.

[37] T. Ng and H. Zhang. A network positioning system for the Internet. In *USENIX*, 2004.

[38] C. G. Plaxton, R. Rajaraman, and A. W. Richa. Accessing nearby copies of replicated objects in a distributed environment. *Theory Comput. Syst.*, 32(3), 1999.

[39] Y. Rabinovich and R. Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *Discrete Comput. Geom.*, 19(1), 1998.

[40] S. B. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *15th SoCG*, 1999.

[41] Y. Shavitt and T. Tankel. Big-bang simulation for embedding network distances in euclidean space. In *IEEE INFOCOM*, 2003.

[42] A. Slivkins. Distributed approaches to triangulation and embedding. In *16th SODA*, 2005.

[43] K. Talwar. Bypassing the embedding: Algorithms for low-dimensional metrics. In *36th STOC*, 2004.

A Tools from Probability Theory

Here we state some tools from Probability Theory that we used in Section 5. We prove some of these results in the full version.

Lemma A.1 (Chernoff Bounds) *Consider the sum X of n independent random variables on $[0, \Delta]$.*

(a) for any $\mu \leq E(X)$ and any $\epsilon \in (0,1)$ we have $\Pr[X < (1-\epsilon)\mu] \leq exp(-\epsilon^2\mu/2\Delta)$.

(b) for any $\mu \geq E(X)$ and any $\beta \geq 1$ we have $\Pr[X > \beta\mu] \leq \left[\frac{1}{e}(e/\beta)^\beta\right]^{\mu/\Delta}$.

Lemma A.2 *Consider two sequences of independent random variables, $\{X_i\}$ and $\{Y_i\}$, such that all X_i and Y_i have finite domains and $X_i \preceq Y_i$ for each i. Then for each k we have $\sum_{i=1}^k X_i \preceq \sum_{i=1}^k Y_i$.*

Lemma A.3 *Consider two sequences of Bernoulli random variables, $\{X_i\}$ and $\{Y_i\}$, such that variables $\{Y_i\}$ are independent and $\Pr[X_i = 1 \mid X_j, j < i] \geq \Pr[Y_i = 1]$ for each i. Then $\sum_{i=1}^k X_i \succeq \sum_{i=1}^k Y_i$ for each k.*

Lemma A.4 *Consider a sequence of i.i.d. Bernoulli random variables $\{Y_i\}$ with success probability q. Let Z_r be the number of trials between the $(r-1)$-th success and the r-th success. Then* $\Pr\left[\sum_{r=1}^k Z_r > 2k/q\right] \leq (0.782)^k$.

B Proofs from Section 3

Proof of Theorem 3.1(b): The idea of the proof is to choose a constant set of beacons and embed them, then for all the other points, choose the "best" beacon *in each coordinate*, and then add some auxiliary coordinates. Formally, let $\tau = \lceil 100 \log n \rceil$ and denote $T = \{t \in \mathbb{N} \mid 1 \leq t \leq \tau\}$. Let $m = \lceil \frac{1}{\epsilon} \rceil$. For each $t \in T$, let B_t be a uniformly distributed random set of m points in X.

For each $t \in T$ let $\vec{g}^{(t)} = (\eta_1^{(t)} g_1^{(t)}, \ldots, \eta_{\beta(m)}^{(t)} g_{\beta(m)}^{(t)})$ be a strongly non-expansive embedding from B_t into ℓ_p with distortion $\alpha(m)$ and dimension $\beta(m)$. Let $I = \{i \in \mathbb{N} \mid 1 \leq i \leq \beta(m)\}$. When clear from the context we omit the $\vec{g}^{(t)}$ superscript and simply write $\vec{g}$. Again, let $\{\sigma_t(u) \mid u \in X, t \in T\}$ be i.i.d symmetric $\{0,1\}$-valued Bernoulli random variables. Define the following functions:

$$\begin{aligned} h_t(u) &= \sigma_t(u)\rho_u(\epsilon) \\ &\quad \text{for all } u \in X \text{ and } t \in T. \\ f_i^t(u) &= \eta_i^{(t)} \min_{b \in B_t}\{d(u,b) + g_i^{(t)}(b)\} \\ &\quad \text{for all } u \in X, i \in I, t \in T. \end{aligned}$$

Let $f^t = (f_1^t, \ldots, f_{\beta(m)}^t)$, $f = (f^1, \ldots, f^\tau)$, and $h = (h^1, \ldots, h^\tau)$, the final embedding will be $\varphi = f \oplus h$. Let $D = \{(u,v) \mid d(u,v) \leq \min\{\rho_u(\epsilon), \rho_v(\epsilon)\}\}$ and by definition $G = \binom{X}{2} \setminus D$.

We begin by an upper bound for all $(u,v) \in G$. Fix $t \in T, i \in I$, and w.l.o.g. assume $f_i^t(u) \geq f_i^t(v)$ and let $b_i^t \in B_t$ be the beacon that minimizes $f_i^t(v)$. Hence, $|\min_{b\in B_t}\{d(u,b)+g_i(b)\} - \min_{b\in B_t}\{d(v,b)+g_i(b)\}| \leq d(u,b_i^t) + g_i(b_i^t) - d(v,b_i^t) - g_i(b_i^t) \leq d(u,v)$.

Also notice that $\max\{\rho_u(\epsilon), \rho_v(\epsilon)\} \leq 2d_X(u,v)$ since $B(u, \rho_v(\epsilon) + d_X(u,v))$ contains at least ϵn points.

$$\begin{aligned} &\|\varphi(u) - \varphi(v)\|_p^p = \|f(u) - f(v)\|_p^p + \|h(u) - h(v)\|_p^p \\ &\leq \textstyle\sum_{t\in T}\sum_{i\in I} |f_i^t(u) - f_i^t(v)|^p + \sum_{t\in T}\max\{\rho_u(\epsilon), \rho_v(\epsilon)\}^p \\ &\leq \textstyle\sum_{t\in T}\sum_{i\in I} |\eta_i^{(t)}\big(\min_{b\in B_t}\{d(u,b)+g_i(b)\} \\ &\qquad - \min_{b\in B_t}\{d(v,b)+g_i(b)\}\big)|^p + \tau(2d(u,v))^p \\ &\leq \textstyle\sum_{t\in T}\sum_{i\in I} |\eta_i^{(t)} d(u,v)|^p + \tau(2d(u,v))^p \end{aligned}$$

$\leq 3\tau d(u,v)^p$.

(Recall that $\sum_{i\in I}\eta_i^p = 1$)

We now partition G into two sets $G_1 = \{(u,v)\in G \mid \max\{\rho_u(\epsilon),\rho_v(\epsilon)\} \geq \frac{d(u,v)}{16\alpha(m)}\}$ and $G_2 = G\setminus G_1$. For any $(u,v)\in G_1, t\in T$, assume w.l.o.g that $\rho_{\hat\epsilon}(u)\geq\rho_{\hat\epsilon}(v)$, and let $\mathcal{E}_t(u,v)$ be the event

$$\mathcal{E}_t(u,v) = \{h^t(u) = \rho_u(\epsilon) \wedge h^t(v) = 0\}$$

Then $\Pr[\mathcal{E}_t(u,v)] = \frac{1}{4}$. Let $A(u,v) = \sum_{t\in T}\mathbf{1}_{\mathcal{E}_t(u,v)}$, then $\mathbb{E}[A(u,v)] = \tau/4$, using Chernoff's bound we can bound the probability that $A(u,v)$ is smaller than half it's expectation:

$$\Pr[A(u,v)\leq\tau/8] \leq e^{-\tau/50} \leq 1/n^2$$

Therefore with probability greater than $1/2$, for any $(u,v)\in G_1$, $A(u,v)\geq\tau/8$. In such a case we can lower bound the contribution for any $(u,v)\in G_1$:

$$\begin{aligned}\|\varphi(u)-\varphi(v)\|_p^p &\geq \textstyle\sum_{t\in T}|h^t(u)-h^t(v)|^p\\ &\geq (\tau/8)\,(\rho_{\hat\epsilon}(u))^p \geq \frac{\tau}{8}\left(\frac{d(u,v)}{16\alpha(m)}\right)^p.\end{aligned}$$

For any $(u,v)\in G_2, t\in T$ let $b_u, b_v\in B_t$ the nearest beacons to u,v respectively. Let

$$\mathcal{F}_t(u,v) = \{b_u\in B\big(u,\rho_u(\epsilon)\big)\wedge b_v\in B\big(v,\rho_v(\epsilon)\big)\}$$

Then $\Pr[\mathcal{F}_t(u,v)] \geq 1-2/e > 1/4$, since for any $u\in X$, $\Pr[d(u,B_t) > \rho_u(\epsilon)] = (1-\hat\epsilon)^{1/\epsilon}\leq e^{-1}$. Let $h(u,v) = \sum_{t\in T}\mathbf{1}_{\mathcal{F}_t(u,v)}$, then $\mathbb{E}[h(u,v)]\geq\tau/4$, using Chernoff's bound we can bound the probability that $h(u,v)$ is smaller than half its expectation:

$$\Pr[h(u,v)\leq\tau/8]\leq e^{-\tau/50}\leq 1/n^2$$

Therefore with probability greater than $1/2$ for any $(u,v)\in G_2$, $h(u,v)\geq\tau/8$.

For any $t\in T$ such that $\mathcal{F}_t(u,v)$ happened we have

$$\max\left\{d(u,b_u),d(v,b_v)\right\}\leq\frac{d(u,v)}{16\alpha(m)}$$

In such a case let $b_i\in B_t$ be the beacon minimizing $f_i^t(u)$; since for every $i\in I$, $g_i(b_u)-g_i(b_i)\leq d(b_u,b_i)$ we get

$$\begin{aligned}f_i^t(u) &= d(u,b_i)+g_i(b_i)\\ &\geq d(u,b_i)+g_i(b_u)-d(b_u,b_i)\\ &\geq g_i(b_u)-d(u,b_u);\end{aligned}$$

moreover

$$f_i^t(v)\leq d(v,b_v)+g_i(b_v)$$

Claim B.1 *Let $J = \{i\in I \mid |g_i(b_u)-g_i(b_v)|\geq\frac{d(u,v)}{4\alpha(m)}\}$. Then $\sum_{i\in J}\eta_i^p|g_i(b_u)-g_i(b_v)|^p\geq\big[\frac{d(u,v)}{4\alpha(m)}\big]^p$.*

Proof: Assume by contradiction that it is not the case, then

$$\begin{aligned}\|\vec g(b_u)-\vec g(b_v)\|_p^p &= \textstyle\sum_{i\in J}\eta_i^p|g_i(b_u)-g_i(b_v)|^p\\ &\quad+\textstyle\sum_{i\notin J}\eta_i^p|g_i(b_u)-g_i(b_v)|^p\\ &< \big[\tfrac{d(u,v)}{4\alpha(m)}\big]^p+\textstyle\sum_{i\notin J}\eta_i^p\big[\tfrac{d(u,v)}{4\alpha(m)}\big]^p\\ &\leq 2\big[\tfrac{d(u,v)}{4\alpha(m)}\big]^p<\big[\tfrac{d(b_u,b_v)}{\alpha(m)}\big]^p\end{aligned}$$

The last inequality follows since $d(b_u,b_v)\geq d(u,v)-2\frac{d(u,v)}{16\alpha(m)}\geq\frac{7}{8}d(u,v)$. This contradicts the fact that $\vec g$ has distortion $\alpha(m)$ on B_t. □

Finally, we can now bound the distortion of the map f^t.

$$\begin{aligned}&\|f^t(u)-f^t(v)\|_p^p = \textstyle\sum_{i\in I}\eta_i^p|f_i^t(u)-f_i^t(v)|^p\\ &\geq \textstyle\sum_{i\in J}\eta_i^p|g_i(b_u)-d(u,b_u)-d(v,b_v)-g_i(b_v)|^p\\ &\geq \textstyle\sum_{i\in J}\eta_i^p\big|\,|g_i(b_u)-g_i(b_v)|\\ &\qquad -2\max\{d(u,b_u),d(v,b_v)\}\big|^p\\ &\geq \textstyle\sum_{i\in J}\eta_i^p\big||g_i(b_u)-g_i(b_v)|-2\frac{1}{4}|g_i(b_u)-g_i(b_v)|\big|^p\\ &\geq \left(\frac{d(u,v)}{8\alpha(m)}\right)^p\end{aligned}$$

Since $\mathcal{F}_t(u,v)$ happened for at least $\tau/8$ indexes from T we have the lower bound

$$\begin{aligned}\|\varphi(u)-\varphi(v)\|_p^p &\geq \textstyle\sum_{t\in T}\|f^t(u)-f^t(v)\|_p^p\\ &\geq \tau/8\left(\frac{d(u,v)}{8\alpha(m)}\right)^p\end{aligned}$$

□

C Bourgain-style proof of Lemma 5.2 for the special case of doubling metrics.

In this section we use the ideas of [11, 33] to derive an alternative proof of Lemma 5.2 for the important special case when β is the doubling dimension.[6] Let us note that in the well-known embedding algorithms of Bourgain [11] and Linial et al. [33] any two nodes are sampled with the same probability, i.e. with respect to the counting measure. Here use a non-trivial extension of the Bourgain's technique where we sample with respect to a doubling measure transformed with respect to a given 1-Lipschitz map.

We state our result as follows:

Lemma C.1 *Consider a finite metric (V,d) equipped with a non-degenerate measure μ and a 1-Lipschitz coordinate map f; write $f_u = f(u)$. For every node u let*

$$\beta_\mu(u) = 2\mu[\,\mathbf{B}_u(f_u)\,]\;/\;\mu[\,\mathbf{B}_u(f_u/16)\,].$$

Then for any $k,t\in\mathbb{N}$ there is a randomized embedding g into ℓ_p, $p\geq 1$ with dimension kt so that:

[6] In this proof the target dimension becomes $t = O(\beta\log\beta)$, which results in target dimension $O(\log^2 n)(\beta\log\beta)$ in Theorem 5.1.

(a) *each coordinate map of g is 1-Lipschitz and upper-bounded by f; and*

(b) $\|g(u) - g(v)\|_p \geq \Omega(d_{uv}/t)(kt)^{1/p}$ *with failure probability at most $t/2^{\Omega(k)}$ for any edge uv such that*

$$f(u)/d_{uv} \in [1/4;\, 4] \text{ and } \max_{w\in\{u,v\}} \beta_\mu(w) \leq 2^t. \tag{11}$$

To prove Lemma 5.2 for a metric of doubling dimension β, recall that for any such metric there exists a 2^β-doubling measure μ. Plug this measure in Lemma C.1, with $t = 4\beta + 1$ and $k = O(\log \beta)$; note that $\beta_\mu(u) \leq 2^t$ for every node u. We get the embedding in ℓ_p with $O(\beta \log \beta)$ dimensions that satisfies the conditions in Lemma 5.2.

We'll need the following simple fact:

$$\text{If} \quad d_{uv} \leq f(u)/8 \text{ for some edge } uv\text{, then } \mathbf{B}_u(f(u)/8) \subset \mathbf{B}_v(f(v)/2) \subset \mathbf{B}_u(f(u)) \tag{12}$$

Indeed, letting $f_u = f(u)$ the first inclusion follows since $f_v/2 \geq (f_u - d_{uv})/2 \geq f_u/8 + d_{uv}$, and the second one holds since $d_{uv} + f_v/2 \leq d_{uv} + (f_u + d_{uv})/2 < f_u$.

Proof of Lemma C.1: Define the transformation of μ with respect to f as $\mu_f(u) = \mu(u)/2\mu(B)$, where $B = \mathbf{B}_u(f_u/2)$. Fix $k = c \log n$ where c is an absolute constants to be specified later. The coordinates are indexed by ij, where $i \in [t]$ and $j \in [k]$. For each (i,j)-pair construct a random set U_{ij} by selecting $\lceil 2^i \mu_f(V) \rceil$ nodes independently according to the probability distribution $\mu_f(\cdot)/\mu_f(V)$. Let us define the ij-th coordinate of u as $g_{ij}(u) = \min(f_u, d(u, U_{ij}))$.

Note that each map g_{ij} is 1-Lipschitz as the minimum of two 1-Lipschitz maps. Therefore part (a) holds trivially. The hard part is part (b). Fix an edge uv; let $d = d_{uv}$. For any node w let $\alpha_w(\epsilon)$ be the smallest radius r such that $\mu_f[\mathbf{B}_w(r)] \geq \epsilon$, and let

$$\begin{aligned} \rho_i &= \max[\psi_u(2^{-i}), \psi_v(2^{-i})], \text{ where} \\ \psi_w(\epsilon) &= \min[\alpha_w(\epsilon), d/2, f_w]. \end{aligned}$$

Claim C.2 *For each $i \geq 1$ and each $j \in [k]$ with probability $\Omega(1)$ we have $g_{ij}(uv) := |g_{ij}(u) - g_{ij}(v)| \geq \rho_i - \rho_{i+1}$.*

Then by Chernoff bounds (Lemma A.1(a)) w.h.p. we have

$$\sum_{ij} g_{ij}(uv) \geq \sum_{i=1}^{t} \Omega(k)(\rho_i - \rho_{i+1}) = \Omega(k)(\rho_1 - \rho_t). \tag{13}$$

Proof of Claim C.2: Fix $i \geq 1$ and j, and note that if $\rho_{i+1} = d/2$ then $\rho_i = d/2$, in which case the claim is trivial. So let's assume $\rho_{i+1} < d/2$ and without loss of generality suppose $\psi_u(2^{-i}) \geq \psi_v(2^{-i})$. Consider the open ball B of radius ρ_i around u. Since $\rho_i = \psi_u(2^{-i}) \leq \alpha_u(2^{-i})$, it follows that $\mu_f(B) \leq 2^{-i}$. Now there are two cases:

- If $\rho_{i+1} = f_v$ then the desired event $g_{ij}(uv) \geq \rho_i - \rho_{i+1}$ happens whenever U_{ij} misses B, which happens with at least a constant probability since $\mu_f(B) \leq 2^{-i}$.
- If $\rho_{i+1} < f_v$ then the desired event happens whenever U_{ij} misses B and hits $B' = \mathbf{B}_v(\rho_{i+1})$. This happens with at least a constant probability by Claim C.4 since $\rho_{i+1} \geq \psi_v(1/2^{i+1}) \geq \alpha_v(1/2^{i+1})$ and therefore $\mu_f(B') \geq 1/2^{i+1}$, and the two balls B and B' are disjoint.

This completes the proof of the claim. □

Claim C.3 *For any node w we have $\alpha_w(\frac{1}{2}) \geq f_w/8$ and $\alpha_w(1/\beta_\mu(w)) \leq f_w/16$.*

Proof: Let $B = \mathbf{B}_w(f_w/8)$. By (12) for any $w' \in B$

$$\mu(w) \,/\, 2\mu[\,\mathbf{B}_w(f_w)\,] \;\leq\; \mu_f(w') \;\leq\; \mu(w)/2\mu(B),$$

so $\mu_f(B) \leq \frac{1}{2}$ and $\mu_f[\,\mathbf{B}_w(f_w/16)\,] \geq 1/\beta_\mu(w)$. □

Suppose that (11) holds; let $x = \max(f_u, f_v)$. Then using Claim C.3 and the definitions of ρ_i and ψ_w we have:

$$\begin{aligned} \rho_1 &\geq \max_{w\in\{u,v\}} \min(f_w/8, d/2) \geq \min(x/8, d/2), \\ \rho_t &\leq \max_{w\in\{u,v\}} \alpha_w(2^{-t}) \leq \max_{w\in\{u,v\}} \alpha_w\left(1/\beta_\mu(w)\right) \\ &\leq \max_{w\in\{u,v\}} f_w/16 \leq x/16. \end{aligned}$$

By (13) for $p = 1$ it remains to show that $\rho_1 - \rho_t \geq \Omega(d)$. There are two cases:

- if $f_v \leq 4d$ then $\rho_1 \geq x/8$, so $\rho_1 - \rho_t \geq x/16 \geq \Omega(d)$.
- if $f_v > 4d$ then $\rho_1 \geq d/2$ and (since f is 1-Lipschitz)
$$\rho_t \;\leq\; f_v/16 \;\leq\; (f_u + d)/16 \;\leq\; 5d/16,$$
so $\rho_1 - \rho_t \geq 3d/16$.

This completes the proof for the case $p = 1$. To extend it to $p > 1$, note that the embedded uv-distance is

$$\begin{aligned} \left(\textstyle\sum_{ij} g_{ij}(uv)^p\right)^{1/p} &= (kt)^{1/p} \left(\tfrac{1}{kt} \textstyle\sum_{ij} g_{ij}(uv)^p\right)^{1/p} \\ &\geq (kt)^{1/p} \left(\tfrac{1}{kt} \textstyle\sum_{ij} g_{ij}(uv)\right) \geq \Omega(d/t)\,(kt)^{1/p} \end{aligned}$$

This completes the proof of the Lemma. □

In the above proof we used the following claim which is implicit in [33] and also stated in [29]; for the sake of completeness, we prove this claim in the full version.

Claim C.4 *Let μ be a probability measure on a finite set V. Consider disjoint events $E, E' \subset V$ such that $\mu(E) \geq q$ and $\mu(E') \leq 2q < 1/2$ for some number $q > 0$. Let S be a set of $\lceil 1/q \rceil$ points sampled independently from V according to μ. Then S hits E and misses E' with at least a constant probability.*