Metric Embeddings with Relaxed Guarantees*

T-H. Hubert Chan[†]   Kedar Dhamdhere[‡]   Anupam Gupta[§]   Jon Kleinberg[¶]

Aleksandrs Slivkins[‖]

April, 2005
Revised: November 2005, September 2006, August 2008

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 [27], 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: there 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.

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 [23].

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. [13, 37, 42]). 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 example, 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 [13]).

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 [27] 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.) Roughly, Kleinberg et. al. [27] 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

---

[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.

own right, have also been proposed on several occasions as candidates for tractable abstractions of the set of Internet latencies (see e.g. [16, 25, 37, 39]).

There were two main open questions posed in [27].

(1) There was no evidence that the main embedding result of [27] 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* [27], 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., [21, 17, 28, 37, 38, 43] 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 [2] (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 $D(\epsilon)$.

<u>**Our Results.**</u> 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 $O(\log^2 n)$ [10], 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 [43], 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 $(\log \frac{1}{\epsilon})^{O(\log \frac{1}{\epsilon})}$.

We then study *embeddings into trees*. We extend the known results of probabilistic embedding into trees [5, 6, 14, 7] to obtain embeddings with slack. In particular, we use the technique of Fakcharoenphol et al. [14] to obtain the following two results:

Theorem 1.5. *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 1.5 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 1.6. *For any input metric (V, d), the randomized embedding of [14] into tree metrics has expected gracefully degrading distortion $D(\epsilon) = O(\log \frac{1}{\epsilon})$.* [3] *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 $D(\epsilon) = O(\log \frac{1}{\epsilon})$.*

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 1.7. *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 $D(\epsilon) = O(\beta)(\log \frac{1}{\epsilon})^{1/p}$.*

---

[3]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 [14].

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 1.7 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 1.7 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 Table 1 on page 18.

Related Work. This work is closely related to the large body of work on metric embeddings in theoretical computer science; see the surveys [23, 24] for a general overview of the area. Our results build on much of the previous work on embeddings into ℓ_p, including [10, 32, 41, 33, 19, 29, 30], and on embeddings of metrics into distributions of trees [3, 5, 6, 20, 14, 7]. Among the special classes of metrics we consider are *doubling metrics* [4, 19, 44, 36]; the book by Heinonen [22] 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 [11] and a comprehensive study was more recently developed by Bartal et. al. [8, 9]. 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 I. Abraham, Y. Bartal and O. Neiman, which lead to a merged publication [1]. The results on lower bounds (Section 5) and on embedding into distributions of trees (Theorem 1.6) were proved similarly by both groups. For the rest of the results, the techniques are quite different. The two groups of authors have agreed to write up the full versions of their results separately.

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 et al. [2], with $D(\epsilon) = O(\log \frac{1}{\epsilon})$ and dimension $O(\log n)$, using a new type of more advanced metric decompositions. They also show a tight result of $O(1/\sqrt{\epsilon})$ distortion for ϵ-slack embedding into a tree metric, and improve the distortion in Theorem 1.7 by a factor of $\beta^{1/p}$.

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 4.1 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.

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 metrics and measures. 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 $\lceil \log s \rceil$ [22, 19]. 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 [22, 36].

Padded Decompositions. Let us recall the definition of a *padded decomposition* (see e.g. [19, 29]). 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 [19]; 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 [26, 41, 15].

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 [27].

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 ϵ-*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 [27], 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 [43] 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 [43], 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

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

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,

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

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

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

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. $\qquad \square$

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 ϵn 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. $\qquad \square$

3 Embeddings into Trees

Probabilistic embedding of finite metric space into trees was introduced in [5]. Fakcharoenphol et al. [14] 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 [14] 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 [14] are *dominating*, i.e. for any edge uv we have $d(u, v) \leq d_T(u, v)$.

Theorem 3.1. *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. [14]. 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*

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

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 [14].

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 $\rho_u(\epsilon) \leq \rho_v(\epsilon)$. Then $\rho_u(\epsilon) \leq d$, so $|\mathbf{B}_u(d)| \leq \epsilon n$.

Run the randomized algorithm of [14] to build a tree T and the associated tree metric d_T. The decomposition scheme will separate u and v at some distance scale $2^i \geq d/2$. 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

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

Look at the sum for i such that $d/2 \leq 2^i < 4d$: this is at most $48d$. By the analysis of [14], the rest of the sum, i.e. the sum for $i \geq \log 4d$, is at most

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

Since the above sum telescopes, it is at most

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

which proves the second inequality in (3). The first inequality in (3) holds trivially because all metrics d_T generated by the algorithm in [14] are dominating. $\square$

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*.

Theorem 3.2. *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 $O(\frac{1}{\epsilon} \log \frac{1}{\epsilon})$.*

4 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 4.1. *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 $D(\epsilon) = O(\beta)(\log \frac{1}{\epsilon})^{1/p}$. 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 [10] and Linial et al. [32], and combines ideas given in [41, 19, 27, 43, 29] 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 the rest of this section, which has the following conceptual steps. We first define a concrete notion of "distance scales" in Section 4.1, 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 4.2, and show that it has gracefully degrading distortion in Section 4.3.

4.1 Distance Scales and Scale Bundles

Our algorithm, just like the algorithms in [10, 32, 41, 19, 27, 29, 30], 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 a 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 [41, 19], 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 [10, 32], $f_{ij}(u)$ 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, 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 [29] 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.

4.2 The Embedding Algorithm

Let us construct the embedding for Theorem 4.1. 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 $1/|\mathbf{B}_u(4f_{ij}(u))|$. Assuming U_{ij} is non-empty, define $f_{(i+1,j)}(u)$ as the minimum of $d(u, U_{ij})$ and $f_{ij}(u)/2$. If U_{ij} is empty, set $f_{(i+1,j)}(u) = f_{ij}(u)/2$. This completes the construction of the scale bundle.

To proceed, let us state a lemma that captures, for our purposes, the structure of the source metric space: this is the only place in the proof of Theorem 4.1 where we use padded decomposition.

Lemma 4.2. *Consider a metric (V, d) which admits β-padded decompositions. Then for any 1-Lipschitz coordinate map f and any $p \geq 1$ there is a randomized embedding g into ℓ_p 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 $\Omega(1)$,

$$\|g(u) - g(v)\|_p \geq \Omega(d_{uv}\, t^{1/p}/\beta). \tag{4}$$

In short, this lemma transforms a "smooth" distance scale f into a "smooth" low-dimensional embedding g which approximately preserves distances along the "relevant" edges. Here "smooth" means "1-Lipschitz", an edge (u, v) is relevant to f if $d_{uv} \approx f(u)$ or $d_{uv} \approx f(v)$, and distances are preserved in the sense of (4). Once the relevant edges are taken care of, we want the coordinates of g to be as small as possible in order to upper-bound the embedded distance on larger distance scales.

Section 4.4 and Section 4.6 contain two different proofs of this lemma; the first one uses padded decomposition techniques from [19, 29], and the other uses some Bourgain-style ideas [10, 32] which we believe are novel and possibly of independent interest.[4]

Fix a pair $i, j \in [k]$. Apply Lemma 4.2 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 4.3. *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 $\varphi_{(i,j,l)}$. $\square$

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.

4.3   Analysis

In this section, we complete the proof of Theorem 4.1 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 of the underlying metric. The proofs are deferred to Section 4.5.

---

[4]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 4.1.

Lemma 4.4. *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 4.5. *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 4.1: Fix an ϵ-long edge uv and let $d = d_{uv}$. Since $g_{(i,j,l)} \leq f_{ij}$ for each l, by Lemma 4.4a 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 4.1 and Lemma 4.2. 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 4.4b there are $\Omega(k)$ different j's such that $f_{ij}(u) \in I$ for some i. By Lemma 4.2 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

$$\Pr\left[\mathcal{E}_{ij} \mid f_{ij}(u) \in I\right] \geq \Omega(1), \quad \text{for each } (i,j)\text{-pair.} \tag{5}$$

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 4.2 $g_{ij}(uv) \geq \Omega(d/\beta)$ 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 4.5 $g_{(i,j,0)}(uv) \geq \Omega(d)$ with conditional probability $\Omega(1)$. This completes the proof of (5).

Let X_j be the indicator variable of the following random event: $\mathcal{E}_{ij}$ and $f_{ij}(u) \in I$ for some i. Since the interval I is u-broad, by Lemma 4.4b there are $\Omega(k)$ different j's such that $f_{ij}(u) \in I$ for some i. Let J be the set of all such j's. Then conditional on J, $\{X_j, j \in J\}$ are $\Omega(k)$ independent 0-1 random variables of expectation $\Omega(1)$. By Chernoff bounds (Lemma A.1a) their sum is $\Omega(1)$ with high probability, completing the proof for case (b). $\qquad \square$

4.4 Analysis: proof of Lemma 4.2

In this section we use padded decomposition techniques from [19, 29] to prove Lemma 4.2. Let us recall the definitions of a *padded decomposition* and a *decomposition bundle* [19, 29].

Definition 4.6. Given a finite metric space (V, d), a positive parameter $\Delta > 0$ and a mapping $\beta : V \to \mathbb{R}$, 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 $\mathbf{B}_x(\Delta/\beta(x))$ is partitioned by P with probability $< \frac{1}{2}$.
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 $s \in \mathbb{Z}$, let the decomposition P_s be chosen according to the distribution $\eta(s)$. We denote $P_s(x)$ to be the unique cluster in P_s containing x.

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

$$W^t := \{x \in V : \sigma_{s(x)-t}(P_{s(x)-t}(x)) = 0\}, \tag{6}$$

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 4.7. *Suppose $f(x) \in [\frac{d}{4}, 4d]$ and $t \in T$ is the integer such that $\hat{s} := s(x) - t$ satisfies $2^{\hat{s}} \in [d/8, d/4)$. Let $J := \{-1, 0, 1\}$ and $\rho := \min\{\frac{1}{32\beta(x,s)} : s \in \hat{s} + J\}$. 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 $s \in \hat{s} + J$, x and z are in the same cluster in $\eta(s)$.
- Event $\mathcal{E}_4$ occurs if for all $s \in \hat{s} + J$, $\sigma_s(P_s(x)) = 1$.

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

$$Pr\left[\mathcal{E}_1 \cap \mathcal{E}_2 \,|\, \eta(\hat{s})\right] = Pr\left[\mathcal{E}_1 \,|\, \eta(\hat{s})\right] \, Pr\left[\mathcal{E}_2 \,|\, \eta(\hat{s})\right] = \tfrac{1}{2} Pr\left[\mathcal{E}_2 \,|\, \eta(\hat{s})\right].$$

Since this fact holds for all decompositions $\eta(\hat{s})$, 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 $|s(x) - s(z)| \in J$. 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 $s(x)$ and $s(z)$ differ by at most one. Again, given the decompositions $\eta(s)$, $s \in \hat{s} + J$, the event $\mathcal{E}_4$ is independent of the event $\mathcal{E}_3 \cap \overline{\mathcal{E}_2}$. Hence,

$$Pr\left[\mathcal{E}_3 \cap \mathcal{E}_4 \cap \overline{\mathcal{E}_2}\right] = Pr\left[\mathcal{E}_4\right] \, Pr\left[\mathcal{E}_3 \cap \overline{\mathcal{E}_2}\right] = \tfrac{1}{8} Pr\left[\mathcal{E}_3 \cap \overline{\mathcal{E}_2}\right].$$

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

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

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 $s \in \hat{s} + J$ we have

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

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. $\qquad \square$

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

Here we prove Lemma 4.4 and Lemma 4.5. First we prove part (a) of Lemma 4.4, 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 4.8. *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

$$4r' \leq f_{ij}(v)/2 \leq (r + d_{uv})/2 \leq 5r/8,$$

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

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 $1/|\mathbf{B}_v(4r')| \leq 1/|B|$. $\qquad \square$

Fix a node u; for simplicity assume $k = 4k_0 + 1$ for some $k_0 \in \mathbb{N}$. 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 4.9. *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 4.8 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 $1 - (1 - 1/2|B|)^{|B|} \geq q$. $\qquad \square$

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 4.9 and Lemma A.3 (proved in Section A) we have the following:

$$\sum_{i=0}^{t} X_{ij} \succeq \sum_{i=0}^{t} Y_i, \quad \text{for any } t \in [k_0] \text{ and each } j \in [k]. \tag{7}$$

We'll use (7) to prove the following crucial claim:

Claim 4.10. *Fix node u and $\epsilon > 0$. 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 (7) for any integer $t \in [k_0]$

$$\Pr[L_j > t] = \Pr\left[\sum_{i=0}^{t} X_{ij} < \alpha\right] \leq \Pr\left[\sum_{i=0}^{t} Y_i < \alpha\right] = \Pr\left[\sum_{r=1}^{\alpha} Z_r > t\right] = \Pr[A_1 > t] \qquad (8)$$

Since $\{A_j\}$ are i.i.d., by (8) and Lemma A.2 it follows that $\sum_j L_j \succeq \sum_j A_j = Z$. Therefore by Lemma A.4

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

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

    Now we have all tools to prove Lemma 4.4a.

Proof of Lemma 4.4a: Use $T_j = T_j(u)$ from Claim 4.10. Fix some ϵ-long edge uv and let $d = d_{uv}$. Let $t_j = \max(T_j(u), T_j(v))$. Then 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

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

where the last inequality follows by Claim 4.10. $\qquad \square$

    To prove part (b) Lemma 4.4, 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 4.4b: 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 a or the smallest b, 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

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

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

14

maps f_{ij}, $i > i_0$ is equivalent to a similarly defined construction that starts with $f_{(i_0,j)} = f$. Therefore, by (7) (applied to this modified construction) we have

$$\Pr[N_l \geq t] \leq \Pr\left[\sum_{\beta=0}^{t-2} X_{(i_0+\beta, j)} = 0\right] \leq \Pr\left[\sum_{\beta=0}^{t-2} Y_\beta = 0\right] = (1-q)^{t-1};$$

$$E[N_l] = \sum_{t=1}^{\infty} \Pr[N_l \geq t] \leq \sum_{t=1}^{\infty}(1-q)^{t-1} = \tfrac{1}{q},$$

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

$$\sum_{i=0}^{k-1} \Pr[r_i \in I_l] = \sum_{\alpha=0}^{3} \sum_{i=0}^{k_0-1} \Pr[r_{4i+\alpha} \in I_l] = \sum_{\alpha=0}^{3} E\left[N_{(l,\alpha)}\right] \leq 4/q. \tag{9}$$

By Claim 4.8 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, $\Pr[\mathcal{E}_i \mid r_i \in I_l] \leq 2^{-l}/32$. By (9) it follows that

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

so some r_i lands in I with at least a constant probability. $\qquad\square$

It remains to prove Lemma 4.5 about the maps $g_{(i, j, 0)}$.

Proof of Lemma 4.5: 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

$$\Pr\left[g_{(i, j, 0)}(u) \leq r \mid r \leq X_u/8\right] \geq \Omega(\beta_r), \tag{10}$$

where $\beta_r = |\mathbf{B}_u(r)|/|\mathbf{B}_u(X_u)|$. Indeed, suppose $r \leq X_u/8$, let $B = \mathbf{B}_u(r)$ and consider any $w \in B$. Then by (13):

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

proving (10). Now let $B = \mathbf{B}_v(X_v/8)$; then by (13) any $w \in B$ is included into the set W_{ij} with probability at most $1/B$, so

$$\Pr\left[g_{(i, j, 0)}(v) \geq X_v/8\right] = \Pr[W_{ij} \text{ misses } B] \geq (1 - 1/|B|)^{|B|} \geq 1/4. \tag{11}$$

Finally, let's combine (10) and (11) to prove the claim. Let $r = d/4$ and suppose $X \geq 4d$. Since $X_v \geq X - d_{uv} \geq 3d$, by (11) event $g_{(i, j, 0)}(v) \geq 3d/8$ happens with probability at least $1/4$. This event and the one in (10) 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 $g_{(i,j,0)}(uv) \geq d/8$. $\qquad\square$

4.6 A Bourgain-style proof of Lemma 4.2 for doubling metrics.

In this section we use the ideas of [10, 32] to derive an alternative proof of Lemma 4.2 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 4.1.

Let us note that in the well-known embedding algorithms of Bourgain [10] and Linial et al. [32] 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 4.11. *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:
(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

$$f(u)/d_{uv} \in [1/4;\ 4] \text{ and } \max_{w \in \{u,v\}} \beta_\mu(w) \leq 2^t. \tag{12}$$

To prove Lemma 4.2 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 4.11, 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 4.2.

We'll need the following simple fact:

$$\text{If } d_{uv} \leq f(u)/8 \text{ for some edge } uv, \text{ then } \mathbf{B}_u(f(u)/8) \subset \mathbf{B}_v(f(v)/2) \subset \mathbf{B}_u(f(u)) \tag{13}$$

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 4.11: Define the transformation of μ with respect to f as $\mu_f(u) = \mu(u)/2\mu(B)$, where $B = \mathbf{B}_u(f_u/2)$. 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

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

Claim 4.12. *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)) with probability at least $1 - 2^{-\Omega(k)}$ we have

$$\sum_{ij} g_{ij}(uv) \geq \sum_{i=1}^t \Omega(k)(\rho_i - \rho_{i+1}) = \Omega(k)(\rho_1 - \rho_t). \tag{14}$$

Proof of Claim 4.12: 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 4.14 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. $\qquad\square$

Claim 4.13. *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 (13) 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 $\mu_f[\mathbf{B}_w(f_w/16)] \geq 1/\beta_\mu(w)$. $\qquad\square$

Suppose (12) holds; let $x = \max(f_u, f_v)$. Then by Claim 4.13 and the definitions of ρ_i and ψ_w we have:

$$\rho_1 \; \geq \; \max_{w \in \{u,v\}} \min(f_w/8, d/2) \geq \min(x/8, d/2),$$

$$\rho_t \; \leq \; \max_{w \in \{u,v\}} \alpha_w(2^{-t}) \leq \max_{w \in \{u,v\}} \alpha_w(1/\beta_\mu(w)) \leq \max_{w \in \{u,v\}} f_w/16 \leq x/16.$$

By (14) 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

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

This completes the proof of the Lemma. $\qquad\square$

In the above proof we used the following claim which is implicit in [32] and also stated in [27]; we prove it here for the sake of completeness.

Claim 4.14. *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.*

17

Type of Embedding	Our Lower Bound	Original Example
All metrics into ℓ_p, $p \geq 1$	$\Omega(\frac{1}{p})(\log \frac{1}{\epsilon})$	Constant-degree expanders [33]
$\mathcal{F}$ into ℓ_p, $p \in (1, 2]$	$\Omega(p-1)\sqrt{\log 1/\epsilon}$	Laakso fractal [31]
Growth-constrained ℓ_1-metrics into ℓ_1^d	$\Omega(\sqrt{\log_d 1/\epsilon})$	Laakso fractal [31]
$\mathcal{F}$ into distributions of dominating trees	$\Omega(\log \frac{1}{\epsilon})$	$n \times n$ grid [3]
All metrics into tree metrics	$\Omega(1/\sqrt{\epsilon})$	n-cycle [40, 18]
ℓ_2^{2m+1} into ℓ_2^{2m}	$\Omega(1/\sqrt{\epsilon})^{1/m}$	[35]

Here $\mathcal{F}$ is the family of doubling metrics that are induced by planar graphs.
Bounds for ϵ-uniform slack can be obtained by replacing $\sqrt{\epsilon}$ by ϵ.

Table 1: Embeddings with slack ϵ: lower bounds on distortion

Proof. Obviously, the probability that S hits E and misses E' can only decrease if we set $\Pr[E] = q$ and $\Pr[E'] = 2q$. Treat sampling a given point as two independent random events: first it misses E' with probability $1 - 2q$, and then (if it indeed misses) it hits E with probability $q' = \frac{q}{1-2q} \leq 2q$. Without loss of generality rearrange the order of events: first we choose whether all points miss E', and then upon success choose whether at least one point hits E. These two events happen independently with probabilities, respectively, $(1 - 2q)^{1/q} \geq 2^{-1/2}$ and

$$1 - (1 - q')^{1/q} \geq 1 - (1 - 2q)^{1/q} \geq 1 - e^{-2}.$$

So the total success probability is at least $c = (1 - e^{-2})/\sqrt{2}$, which is an absolute constant as required. □

5   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. For simplicity of exposition, we will first give a concrete example proving lower bounds for ϵ-slack embeddings into ℓ_p (which will follow from a lower bound for embedding expanders into ℓ_p [33]). Then we provide the general technique; the bounds obtained by this technique are given in Table 1. 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 [33] for embeddings into ℓ_p holds even with ϵ-slack (see Section 5.2 for more details).

Theorem 5.1. *For an arbitrarily small positive ϵ there exists a finite metric on arbitrarily many nodes that requires distortion $\Omega(\frac{1}{p} \log \frac{1}{\epsilon})$ to embed into ℓ_p, $p \geq 1$ with ϵ-slack.*

Proof. Given an ϵ such that $0 < \epsilon \leq 1/12$, let $k = 1/(3\sqrt{\epsilon})$. Fix n, the number of nodes in our counterexample.

We now construct a graph G on n vertices. Consider a constant degree expander graph H on k vertices. Let (H, d) be the shortest path metric defined by H. For each vertex $s \in H$, let L_s be a path containing n/k vertices. Attach the path L_s to s at one of its endpoints. The length of each edge of L_s is small enough so that if δ is the length of path L_s, then $\delta \cdot D \leq 1/2$. Let the new graph be G and the shortest path metric defined on it be (G, d). We now prove that if (G, d) can be embedded into ℓ_p with distortion D and ϵ-slack, then H can be embedded into ℓ_p with distortion $4D$ *without any slack*.

Let $\varphi : G \to \ell_p$ be the embedding of (G, d) into ℓ_p with distortion D and ϵ-slack. Let E denote the set of *ignored* pairs, i.e. let us assume that the complement of E incurs distortion at most D. Note that ϵ-slack

18

means that $|E| \leq \epsilon n^2 2$. We delete all the vertices that participate in more than $\sqrt{\epsilon}n$ pairs in E. By a simple counting argument, at most $\sqrt{\epsilon}n$ vertices of G can be deleted. Therefore, at least one point from each path survives. For each $s \in H$, let v_s denote a survived vertex from the path L_s. We define an embedding ψ of H into ℓ_p as $\psi(s) = \varphi(v_s)$.

We now bound the distortion of the embedding ψ by $4D$. Let x, y be two vertices in H. Then v_x and v_y are the survivors in L_x and L_y respectively. Note that v_x and v_y participate in at most $\sqrt{\epsilon}n$ pairs in E. Since $|L_y| = 3\sqrt{\epsilon}n$, it follows that there is another survivor $t \in L_y$ such that neither $\{t, v_x\}$ nor $\{t, v_y\}$ is in E. Since the distortion of the map φ is D, we can assume that for edge $(u, v) \notin E$,

$$d(u, v) \leq \|\varphi(u) - \varphi(v)\|_p \leq D \cdot d(u, v).$$

Now we can bound $\psi(xy) := \|\psi(x) - \psi(y)\|_p$ as follows:

$$
\begin{aligned}
\psi(xy) &= \|\varphi(v_x) - \varphi(v_y)\|_p \\
&\leq \|\varphi(v_x) - \varphi(t)\| + \|\varphi(t) - \varphi(v_y)\| \\
&\leq D\left(d(v_x, t) + d(t, v_y)\right) \\
&\leq D\left(1 + 3\delta\right) d(x, y) \leq 2D \, d(x, y).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\psi(xy) &\geq \|\varphi(v_x) - \varphi(t)\|_p - \|\varphi(t) - \varphi(v_y)\|_p \\
&\geq d(v_x, t) - Dd(t, v_y) \geq (1 - D\delta)d(x, y) \\
&\geq d(x, y)/2.
\end{aligned}
$$

Hence $\frac{1}{2}d(u, v) \leq \psi(uv) \leq 2D \cdot d(u, v)$, and so ψ is a map from H to ℓ_p with distortion $4D$.

To finish the proof of the theorem, we note that a constant-degree expander on k vertices requires $\Omega(\log k/p)$ distortion to embed into ℓ_p [33]. $\qquad \square$

5.1 General lower-bounding technique

The technique used in Theorem 5.1 of starting with a $O(1)$-degree expander H_k on k vertices, replacing each vertex with a path on n/k vertices to get G, and for suitable $k \approx O(1/\sqrt{\epsilon})$ arguing that ϵ-slack embeddings of G_n give us slack-less embeddings of H_k with (roughly) the same distortion is quite general. In fact, we use it to obtain lower bounds on *both the distortion and dimensions of embeddings* into ℓ_p from similar lower bounds for slack-less embeddings; similar results can be obtained for embeddings into trees, or distributions of trees. We summarize these results in Table 1.

Theorem 5.2. *Suppose for each k there exists a k-node metric H_k such that any embedding of H_k into ℓ_p with $L(k)$ dimensions has distortion at least $D(k)$. Then for an arbitrarily small positive ϵ there exist finite metrics M, M^* on arbitrarily large number of nodes such that:*

(a) any embedding of M into ℓ_p with $L(\frac{1}{3\sqrt{\epsilon}})$ dimensions has ϵ-slack distortion $\Omega(D(\frac{1}{3\sqrt{\epsilon}}))$.

(b) any embedding of M^* into ℓ_p with $L(\frac{1}{3\epsilon})$ dimensions has ϵ-uniform slack distortion $\Omega(D(\frac{1}{3\epsilon}))$.

Moreover, if metrics $\{H_k\}$ are planar (resp. K_r-minor-free, doubling, ℓ_p^d) then so are M and M^*.

Note that this result can be used to translate, e.g., the Brinkman and Charikar [12] lower bound for dimensionality reduction in ℓ_1 into the realm of ϵ-slack as well.

Similarly, we provide a lower bound theorem for (probabilistic) embeddings into trees:

19

Theorem 5.3. *Suppose for each k there exists a k-node metric H_k such that any (probabilistic) embedding of H_k into trees has distortion at least $D(k)$. Then for an arbitrarily small positive ϵ there exist finite metrics M, M^* on arbitrarily large number of nodes such that:*

(a) any (probabilistic) embedding of M into trees has ϵ-slack distortion $\Omega(D(\frac{1}{3\sqrt{\epsilon}}))$.

(b) any (probabilistic) embedding of M^* into trees has ϵ-uniform slack distortion $\Omega(D(\frac{1}{3\epsilon}))$.

Moreover, if metrics $\{H_k\}$ are planar (resp. K_r-minor-free, doubling, ℓ_p^d) then so are M and M^*.

For instance, we can now derive a lower bound of $\Omega(1/\sqrt{\epsilon})$ on the distortion incurred when embedding the n-cycle into a single tree.

The proofs of the two above theorems are based on the following lemma:

Lemma 5.4 (Master Lemma). *Suppose H is a metric on k points and $\mathcal{T}$ is a collection of metrics on k points, such that any embedding of H into $\mathcal{T}$ incurs a distortion at least D. Suppose $\mathcal{S}$ is a collection of metrics such that every subset of k points in each metric in $\mathcal{S}$ embeds into $\mathcal{T}$ with distortion at most ρ. Setting $\epsilon = 1/9k^2$, there exist arbitrarily large metrics that embed into $\mathcal{S}$ with ϵ-slack distortion $\Omega(\frac{D}{\rho})$.*

Remark. In order to obtain lower bounds for ϵ-*uniform slack* embeddings instead of ϵ-slack embeddings, we need to set $\epsilon = 1/3k$ instead of $\epsilon = 1/9k^2$; the rest of the proof remains essentially unchanged.

Before we prove Lemma 5.4, let us show how to derive the above results from it.

Proof of Theorem 5.2: Suppose $\{H_k\}$ is the given family of metrics. Let us fix a large enough k such that $\epsilon = 1/9k^2$ is small enough. Now in Lemma 5.4, let us set H to be H_k and $\mathcal{T}$ to be the collection of metrics with k points in ℓ_p with at most $L(k)$ dimensions. Hence, H embeds into $\mathcal{T}$ with distortion at least $D(k) = D(\frac{1}{3\sqrt{\epsilon}})$. We set $\mathcal{S}$ to be the family of metrics in ℓ_p with at most $L(k) = L(\frac{1}{3\sqrt{\epsilon}})$ dimensions. It follows that any subset of k points in any metric in $\mathcal{S}$ embeds into $\mathcal{T}$ with distortion 1. Hence, we conclude that there exists a family of metrics, each of which embeds into ℓ_p with at most $L(\frac{1}{3\sqrt{\epsilon}})$ dimensions with ϵ-slack distortion at least $\Omega(D(\frac{1}{3\sqrt{\epsilon}}))$. $\square$

The application of Lemma 5.4 to prove the lower bounds for embeddings into trees is very similar; we sketch it here to emphasize the general patterns, as well as the slight changes required.

Proof of Theorem 5.3: Again, we large enough k, and set $\epsilon = 1/9k^2$. As before, H is set to be H_k. We set $\mathcal{T}$ to be the family of tree metrics on k points (or distribution of tree metrics on k points). Again, H embeds into $\mathcal{T}$ with distortion at least $D(k) = D(\frac{1}{3\sqrt{\epsilon}})$. We set $\mathcal{S}$ to be the family of tree metrics (or distribution of tree metrics). Note that by a result of Gupta [18], any subset of k points in any metric in $\mathcal{S}$ embeds into $\mathcal{T}$ with distortion at most 8. Now the result of Theorem 5.3 follows from Lemma 5.4 as before. $\square$

Let us now prove the Lemma 5.4: first we show how to construct a family of metrics with the desired properties. Suppose $H = (S, d)$ is a metric such that $|S| = k$. Moreover, H embeds into $\mathcal{T}$ with distortion at least D. Without loss of generality, assume that the pairwise distance in H is at least 1. For each n that is a multiple of $3k$, we define a metric $\hat{H}$ with n points in the following way. These would be the family of metrics that exhibits the lower bound for slack embeddings.

Consider a uniform line metric with point set L of size $\frac{n}{k}$ such that the two terminal points are at distance δ away from each other, where δ is small and whose value will be specified later. For each $s \in S$, we identify s with a terminal point of a copy L_s of the line metric L. We call the augmented metric $\hat{H} = (V, d)$ with point set $V = \cup_{s \in S} L_s$. If H is already in some host space X, we just need the condition that for each $s \in S$, we can embed a copy of L of length δ isomorphically into X that identifies one end point with s. Common metric spaces like ℓ_p certainly satisfy this condition. (Note that to avoid too many symbols, we use d for the various metrics.) Hence, for $u \in L_x$ and $v \in L_y$, $|d(u, v) - d(x, y)| \leq 2\delta$.

Proposition 5.5. *Let H and $\hat{H}$ be metrics defined as above. Then,* **(a)** *if H is a metric induced by a K_r-minor free graph, then so is $\hat{H}$, and* **(b)** *if H is a doubling metric, then so is $\hat{H}$.*

The next lemma states a crucial property of the edges that are ignored by any ϵ-slack embedding.

Lemma 5.6. *Suppose an ϵ-slack embedding of some metric (V, d) ignores the set of edges E. Then, there exists a subset $T \subseteq V$ of size at least $(1 - \sqrt{\epsilon})n$ such that each vertex in T intersects with at most $\sqrt{\epsilon}n$ edges in E.*

Proof. It suffices to show that it is impossible to have a subset $S \subseteq V$ of size greater than $\sqrt{\epsilon}n$ such that each vertex in S intersects more than $\sqrt{\epsilon}n$ edges in E. Otherwise, the total number of edges ignored would be greater than $(\sqrt{\epsilon}n)^2/2 > \epsilon n^2/2 > \epsilon\binom{n}{2}$. $\quad\square$

Note that for an ϵ-*uniform slack* embedding, the number of ignored edges incident on any node is at most ϵn by definition; this is one place in the proof which changes when considering uniform slack.

The following lemma implies Lemma 5.4:

Lemma 5.7. *Let $H = (S, d)$ be a metric on k points. Suppose $\mathcal{T}$ and $\mathcal{S}$ are families of metrics such that H embeds into $\mathcal{T}$ with distortion at least D, and every subset of k points in each metric in $\mathcal{S}$ embeds into $\mathcal{T}$ with distortion at most ρ.*

Suppose δ is small enough such that $(\frac{D}{4\rho} + 2)\delta \leq \frac{1}{2}$. Let $\hat{H} = (V, d)$ be the metric be defined as above. Let $\epsilon := 1/9k^2$. Then, $\hat{H}$ embeds into $\mathcal{S}$ with ϵ-slack distortion at least $D/4\rho$.

Proof. Suppose, on the contrary, φ is an embedding of $\hat{H}$ into $\mathcal{S}$ with ϵ-slack distortion $R < D/4\rho$ that ignores the set E of edges. Then by Lemma 5.6, there exists a subset T of V such that $|T| \geq (1 - \sqrt{\epsilon})n$ and for all $v \in T$, v intersects at most $\sqrt{\epsilon}n$ edges in E.

For each $s \in S$, the set L_s contains $\frac{n}{k} = 3\sqrt{\epsilon}n$ points and hence there exists some point in $T \cap L_s$, which we call v_s. We define an embedding ψ of H into $\mathcal{S}$ given by $\psi(s) := \varphi(v_s)$. We next bound the distortion of the embedding ψ. Let $x, y \in S$. Since v_x and v_y are in T, each of them has at most $\sqrt{\epsilon}n$ neighbors. Observing that $|L_y| = 3\sqrt{\epsilon}n$, it follows that there exists a point $t \in L_y$ such that neither $\{v_x, t\}$ nor $\{v_y, t\}$ is contained in E. We can assume that for $\{u, v\} \notin E$, $d(u, v) \leq \|\varphi(u) - \varphi(v)\| \leq Rd(u, v)$.

Hence, it follows that

$$\begin{aligned}
\|\psi(x) - \psi(y)\| =& \|\varphi(v_x) - \varphi(v_y)\| \\
\leq& \|\varphi(v_x) - \varphi(t)\| + \|\varphi(t) - \varphi(v_y)\| \\
\leq& R(d(v_x, t) + d(t, v_y)) \leq R(d(x, y) + 3\delta) \\
\leq& R(1 + 3\delta)d(x, y) \leq 2Rd(x, y),
\end{aligned}$$

and similarly,

$$\begin{aligned}
\|\psi(x) - \psi(y)\| \geq& \|\varphi(v_x) - \varphi(t)\| - \|\varphi(t) - \varphi(v_y)\| \\
\geq& d(v_x, t) - Rd(t, v_y) \geq d(x, y) - 2\delta - R\delta \\
\geq& (1 - (R + 2)\delta)d(x, y) \geq d(x, y)/2,
\end{aligned}$$

where the last inequality follows from the fact that $(R + 2)\delta \leq 1/2$. It then follows that ψ embeds H into $\mathcal{S}$ with distortion at most $4R$. However, since any metric in $\mathcal{S}$ embeds into $\mathcal{T}$ with distortion at most ρ, it follows that H embeds into $\mathcal{T}$ with distortion at most $4\rho R < D$, from which we obtain the desired contradiction. $\quad\square$

5.2 Lower bounds for contracting embeddings

Let us consider contracting embeddings with slack. Formally, a contracting embedding has distortion D with ϵ-slack if no pairwise distance expands and all but ϵ-fraction of the pairs contract by no more than D. We show that such embeddings incur an $\Omega(\log n)$ distortion in order to embed constant-degree expander graphs into ℓ_p, $p \geq 1$.

Theorem 5.8. *For the shortest-paths metric of a bounded-degree expander on n vertices, distortion of any contracting embedding into ℓ_p, $p \geq 1$ is $\Omega(\frac{1}{p}\log n)$ even if we allow slack $\epsilon < \frac{1}{2}$.*

Proof. Let $G = (V, E)$ be a bounded-degree expander on n vertices, and let ρ denote its shortest path metric. Let φ be a contracting embedding of this metric to ℓ_p, $p \geq 1$ with distortion D and slack $\epsilon < \frac{1}{2}$. Let σ denote the metric on ℓ_p; to simplify the notation, we will denote $\varphi(V) \subseteq \ell_p$ by V. Define

$$
\begin{aligned}
R(\sigma) &= \sqrt{\sigma^2(V \times V)/\sigma^2(E)} \text{ , where} \\
\sigma^2(S) &= \textstyle\sum_{(x,y)\in S} \sigma(x,y)^2 \text{ for any set } S \subseteq V \times V.
\end{aligned}
$$

First we show that $R(\sigma) \leq O(\sqrt{n})$. The proof is exactly the same as that of Theorem 15.5.1 in Matousek [34] and works despite the fact that we allow $\varepsilon \cdot n^2$ pairwise distances to be as low as 0. Note that

$$
\sigma^2(E) = \textstyle\sum_{(x,y)\in E} \sigma(x,y)^2 \leq \sum_{(x,y)\in E} \rho(x,y)^2 = O(n)
$$

Now, we bound $\sigma^2(V \times V)$ from below. If all n^2 pairs were contracted by at most D, then we would get

$$
\sigma^2(V \times V) \geq \textstyle\sum_{(u,v)} \left(\frac{\rho(u,v)}{D}\right)^2 \geq \frac{n^2\log^2 n}{D^2}
$$

However, we need to take into account the fact that $\varepsilon \cdot n^2$ pairs of vertices could have distance 0 between them. Therefore, $\sigma^2(V \times V)$ is at least $(n/D)^2(\log^2 n)$ minus the loss due to the slack. To upper-bound this loss, consider a pair (x, y) of nodes for which the distortion is bigger than D. The pair will contribute 0 instead of $\rho(x,y)/D$. Thus the loss due to the pair (x, y) is at most $(\log n)/D$. Therefore, the total loss due to the slack is at most $\epsilon(n/D)^2(\log^2 n)$. Therefore, since $R(\sigma) \leq O(\sqrt{n})$, it follows that $D = \Omega(\log n)$. $\square$

References

[1] I. Abraham, Y. Bartal, H. T.-H. Chan, K. Dhamdhere, A. Gupta, J. Kleinberg, O. Neiman, and A. Slivkins. Metric Embeddings with Relaxed Guarantees. In *46th Symp. on Foundations of Computer Science (FOCS)*, pages 83–100, 2005.

[2] I. Abraham, Y. Bartal, and O. Neiman. Advances in Metric Embedding Theory. In *38th ACM Symp. on Theory of Computing (STOC)*, 2006.

[3] 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):78–100, 1995.

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

[5] Y. Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *37th Symp. on Foundations of Computer Science (FOCS)*, pages 184–193, 1996.

[6] Y. Bartal. On approximating arbitrary metrics by tree metrics. In *30th ACM Symp. on Theory of Computing (STOC)*, pages 161–168, 1998.

[7] Y. Bartal. Graph Decomposition Lemmas and their Role in Metric Embedding Methods. In *12th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 89–97, 2004.

[8] Y. Bartal, B. Bollobás, and M. Mendel. Ramsey-type Theorems for Metric Spaces with Applications to Online Problems. *J. Computer and System Science (special issue)*, 72(5):890–921, 2002. Preliminary version in *42nd IEEE FOCS, 2001*.

[9] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric ramsey-type phenomena. *Annals of Mathematics*, 162(2):643–709, 2005. Preliminary version in *35th ACM STOC, 2003*.

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

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

[12] B. Brinkman and M. S. Charikar. On the impossibility of dimension reduction in l1. In *44th Symp. on Foundations of Computer Science (FOCS)*, pages 514–523, 2003.

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

[14] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. of Computer and System Sciences*, 69(3):485–497, 2004. Preliminary version in *35th ACM STOC*, 2003.

[15] J. Fakcharoenphol and K. Talwar. An improved decomposition theorem for graphs excluding a fixed minor. *RANDOM-APPROX*, pages 36–46, 2003.

[16] 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.

[17] 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. Preliminary version in IEEE INFOCOM 1999.

[18] A. Gupta. Steiner points in tree metrics don't (really) help. In *12th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 220–227, Philadelphia, PA, 2001.

[19] A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low–distortion embeddings. In *44th Symp. on Foundations of Computer Science (FOCS)*, pages 534–543, 2003.

[20] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and ℓ_1-embeddings of graphs. *Combinatorica*, 24(2):233–269, 2004. (Preliminary version in 40th FOCS, 1999.).

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

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

[23] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *42nd Symp. on Foundations of Computer Science (FOCS)*, pages 10–33, 2001. Invited survey.

[24] 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, pages xviii+1539. Chapman & Hall/CRC, Boca Raton, FL, second edition, 2004.

[25] D. R. Karger and M. Ruhl. Finding Nearest Neighbors in Growth-restricted Metrics. In *34th ACM Symp. on Theory of Computing (STOC)*, pages 63–66, 2002.

[26] P. Klein, S. A. Plotkin, and S. B. Rao. Excluded minors, network decomposition, and multicommodity flow. In *25th ACM Symp. on Theory of Computing (STOC)*, pages 682–690, 1993.

[27] J. Kleinberg, A. Slivkins, and T. Wexler. Triangulation and Embedding Using Small Sets of Beacons. In *45th Symp. on Foundations of Computer Science (FOCS)*, pages 444–453, 2004.

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

[29] R. Krauthgamer, J. Lee, M. Mendel, and A. Naor. Measured descent: A new embedding method for finite metrics. *Geometric and Functional Analysis (GAFA)*, 15(4):839–858, 2005. Preliminary version in *45th IEEE FOCS*, 2004.

[30] J. R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *16th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 92–101, 2005.

[31] J. R. Lee, M. Mendel, and A. Naor. Metric Structures in L1: Dimension, Snowflakes, and Average Distortion. In *Latin-American Conf. on Combinatorics, Graphs and Applications (LATIN)*, Buenos Aires, Argentina, 2004.

[32] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. (Preliminary version in *35th IEEE FOCS*, 1994).

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

[34] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[35] J. Matoušek. Bi-Lipschitz embeddings into low dimensional Euclidean spaces. *Commentationes Mathematicae Universitatis Carolinae*, 31(3):589–600, 1990.

[36] M. Mendel and S. Har-Peled. Fast construction of nets in low dimensional metrics, and their applications. In *21st ACM Symp. on Computational Geometry (SoCG)*, pages 150–158, 2005.

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

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

[39] C. G. Plaxton, R. Rajaraman, and A. W. Richa. Accessing nearby copies of replicated objects in a distributed environment. *Theory Comput. Syst.*, 32(3):241–280, 1999. Preliminary version in *9th ACM SPAA*, 1997.

[40] Y. Rabinovich and R. Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *Discrete Comput. Geom.*, 19(1):79–94, 1998.

[41] S. B. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *15th ACM Symp. on Computational Geometry (SoCG)*, pages 300–306, 1999.

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

[43] A. Slivkins. Distributed Approaches to Triangulation and Embedding. In *16th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 640–649, Jan. 2005.

[44] K. Talwar. Bypassing the embedding: Algorithms for low-dimensional metrics. In *36th ACM Symp. on Theory of Computing (STOC)*, pages 281–290, 2004.

A   Tools from Probability Theory

Here we state some tools from Probability Theory that we used in Section 4.

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 $\Pr[X > \beta\mu] \leq \left[\frac{1}{e}(e/\beta)^\beta\right]^{\mu/\Delta}$.*

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 $t \in \mathbb{R}$.

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.

Proof. We first show that for all $t \in [T]$,

$$\Pr\left[\sum_{r=1}^{t} X_r + \sum_{r=t+1}^{T} Y_r \leq m\right] \leq \Pr\left[\sum_{r=1}^{t-1} X_r + \sum_{r=t}^{T} Y_r \leq m\right], \tag{15}$$

which would immediately imply the lemma. Observe that for any fixed number a (or in general any random variable that is measurable in the σ-field generated by the random variables $\{X_r : r < t\}$), we have

$$\Pr\left[X_t \leq a | X_r, r < t\right] \leq Pr[Y_t \leq a] = \Pr\left[Y_t \leq a | X_r, r < t\right].$$

Note that the interesting case is when $a \in [0, 1)$. The inequality comes from the assumption concerning the conditional probabilities of the sequence $\{X_r\}$, and the equality comes from the fact that Y_t is independent of the sequence $\{X_r\}$.

Since both X_t and Y_t are independent of $\{Y_r : r > t\}$, the above inequality would still hold if we further condition on the random variables $\{Y_r : r > t\}$. Finally, setting $a = m - \sum_{i<t} X_r - \sum_{i>t} Y_r$, which is measurable in the σ-field generated by $J := \{X_r : r < t\} \cup \{Y_r : r > t\}$, we obtain

$$\Pr\left[\sum_{r=1}^{t} X_r + \sum_{r=t+1}^{T} Y_r \leq m \,\middle|\, J\right] \leq \Pr\left[\sum_{r=1}^{t-1} X_r + \sum_{r=t}^{T} Y_r \leq m \,\middle|\, J\right].$$

Taking expectation on both sides gives (15). $\qquad\square$

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*

$$\Pr\left[\sum_{r=1}^{k} Z_r > 2k/q\right] \leq (0.782)^k. \tag{16}$$

Proof. Each Z_r has a geometric distribution with parameter q, so its moment generating function is

$$E\left[e^{tZ_r}\right] = \frac{qe^t}{q - (1-q)e^t}.$$

Let $Z = \sum_{r=1}^{k} Z_r$. Since Z_r's are i.i.d, it follows that $E\left[e^{tZ}\right] = E\left[\prod_r e^{tZ_r}\right] = \left(E\left[e^{tZ_1}\right]\right)^k$.
By Markov inequality for any $t > 0$ we have

$$Pr[Z > 2k/q] = Pr\left[e^{tZ} > e^{2tk/q}\right] \leq E\left[e^{tZ}\right] e^{-2tk/q} \leq \left(\frac{qe^t}{(1-(1-q)e^t)e^{2t/q}}\right)^k.$$

Plugging in $q = 1 - 1/\sqrt{e}$ and $t = q$ we have (16). $\qquad\square$