Triangulation and Embedding using Small Sets of Beacons*

Jon Kleinberg† Aleksandrs Slivkins‡ Tom Wexler§

April, 2004
Revised: August 2004, September 2006, August 2008

Abstract

Concurrent with recent theoretical interest in the problem of metric embedding, a growing body of research in the networking community has studied the distance matrix defined by node-to-node latencies in the Internet, resulting in a number of recent approaches that approximately embed this distance matrix into low-dimensional Euclidean space. There is a fundamental distinction, however, between the theoretical approaches to the embedding problem and this recent Internet-related work: in addition to computational limitations, Internet measurement algorithms operate under the constraint that it is only feasible to measure distances for a linear (or near-linear) number of node pairs, and typically in a highly structured way. Indeed, the most common framework for Internet measurements of this type is a *beacon-based* approach: one chooses uniformly at random a constant number of nodes ('beacons') in the network, each node measures its distance to all beacons, and one then has access to only these measurements for the remainder of the algorithm. Moreover, beacon-based algorithms are often designed not for embedding but for the more basic problem of *triangulation*, in which one uses the triangle inequality to infer the distances that have not been measured.

Here we give algorithms with provable performance guarantees for beacon-based triangulation and embedding. We show that in addition to multiplicative error in the distances, performance guarantees for beacon-based algorithms typically must include a notion of *slack* — a certain fraction of all distances may be arbitrarily distorted. For metric spaces of bounded doubling dimension (which have been proposed as a reasonable abstraction of Internet latencies), we show that triangulation-based distance reconstruction with a constant number of beacons can achieve multiplicative error $1 + \delta$ on a $1 - \epsilon$ fraction of distances, for arbitrarily small constants δ and ϵ. For this same class of metric spaces, we give a beacon-based embedding algorithm that achieves constant distortion on a $1 - \epsilon$ fraction of distances; this provides some theoretical justification for the success of the recent Global Network Positioning algorithm of Ng and Zhang, and it forms an interesting contrast with lower bounds showing that it is not possible to embed *all* distances in a doubling metric space with constant distortion. We also give results for other classes of metric spaces, as well as distributed algorithms that require only a sparse set of distances but do not place too much measurement load on any one node.

ACM categories and subject descriptors: F.2.2 [Analysis of Algorithms and Problem Complexity]: Nonnumerical Algorithms and Problems — geometrical problems and computations, computations on discrete structures; E.1 [Data Structures]: distributed data structures, graphs and networks; G.2.2 [Discrete Mathematics]: Graph Theory – graph algorithms. **General terms:** algorithms, theory. **Keywords:** metric spaces, embedding, triangulation, doubling dimension.

---

*The conference version [29] of this paper has appeared in *45th IEEE FOCS*, 2004. The present version also includes portions of Slivkins [53] from *16th ACM-SIAM SODA*, 2005. This work has been completed while A. Slivkins and T. Wexler were graduate students at Cornell University, supported, respectively, by the Packard Fellowship of J.Kleinberg and by NSF grant 0311333.

†Department of Computer Science, Cornell University, Ithaca, NY 14853. *kleinber at cs.cornell.edu*. Supported in part by a David and Lucile Packard Foundation Fellowship and NSF grants 0081334 and 0311333.

‡Microsoft Research, Mountain View, CA 94043. Email: *slivkins at microsoft.com*.

§Department of Mathematics and Computer Science, Denison University, Granville, OH 43023. *wexlert at denison.edu*.

1 Introduction

The past decade has seen many significant and elegant results in the theory of metric embeddings (for recent surveys, see [24, 37, 41, 25]). Embedding techniques have been valuable in the design and analysis of algorithms that operate on an underlying metric space; many optimization problems become more tractable when the given metric space is embedded into one that is structurally simpler.

Meanwhile, an active line of research in the networking community has studied the distance matrix defined by node-to-node latencies in the Internet [15, 18, 20, 22, 30, 57], resulting in a number of recent approaches that approximately embed this distance matrix into low-dimensional Euclidean space [11, 44, 47, 51].[1] However, there is a fundamental distinction between this Internet-related work and the large body of theoretical work on embedding, due to the following intrinsic problem: *in any analysis of the distance matrix of the Internet, most distances are not available.* The cost of measuring all node-to-node distances is simply too expensive; instead, we have a setting where it is generally feasible to measure the distances among only a linear (or near-linear) number of node pairs, and typically in a highly structured way. Indeed, the most common framework for Internet measurements of this type is a *beacon-based* approach: one chooses uniformly at random a constant number of nodes ('beacons') in the network, each node measures its distance to all beacons, and one then has access to only these $O(n)$ measurements for the remainder of the algorithm. (For example, the data can be shared among the beacons, who then perform computations on the data locally.)

This inability to measure most distances is the inherent obstacle that stands in the way of applying algorithms developed from the theory of metric embeddings, which assume (and use) access to the full distance matrix. Thus, to obtain insight at a theoretical level into recent Internet measurement studies, we need to consider problems in following two genres.

(i) What performance guarantees can be achieved by metric embedding algorithms when only a sparse (beacon-based) subset of the distances can be measured?

(ii) At an even more fundamental level, many Internet measurement algorithms are seeking not to embed but simply to reconstruct the unobserved distances with reasonable accuracy (see e.g. [15, 18, 20, 30]). Can we give provable guarantees for this type of *reconstruction* task?

Reconstruction via triangulation. Within this framework, we discuss the reconstruction problem (ii) first, as it is a more basic concern. Motivated by the research of Francis et al. on IDMaps [15], and subsequent work, we formalize the reconstruction problem here as follows. Let S be the set of beacons; and suppose for each node u, and each beacon $b \in S$, we know the distance d_{ub}. What can we infer from this data about the remaining unobserved distances d_{uv} (when neither v nor v is a beacon), assuming we know only that we have points in an arbitrary metric space? The triangle inequality implies that

$$\max_{b \in S} |d_{ub} - d_{vb}| \leq d_{uv} \leq \min_{b \in S} d_{ub} + d_{vb}, \tag{1}$$

and it is easy to see that these are the tightest bounds that can be provided on d_{uv} if we assume only that the underlying metric is arbitrary subject to the given distances. We will say that d_{uv} is reconstructed by *triangulation*[2], with distortion $\Delta \geq 1$, if the ratio between the upper and lower bounds in (1) is at most Δ.

[1] We speak of Internet latencies as defining as a "distance matrix" rather than a metric, since the triangle inequality is not always observed; however, one can view the recent networking research as indicating that severe triangle inequality violations are not widespread enough to prevent the matrix of node-to-node latencies from being usefully modeled using notions from metric spaces.

[2] Note that this is one of several standard uses of the term "triangulation" in the literature; it should not be confused with the process of dividing up a region into simplices, which goes by the same name.

Since it is much cheaper for nodes to exchange messages than to actually estimate their round-trip distance on the Internet (the latter typically requires a significant measurement period to produce a stable estimate), triangulation can be valuable as a way to assign each node a short *label* — its distances to all beacons — in such a way that the distance d_{uv} can later be estimated by a third party (or by one of v or v) just from their labels. This can be viewed as a kind of *distance labeling*, and we discuss related work on this topic (e.g. [16]) below.

To give performance guarantees for triangulation, we also need a notion of *slack*. Even in very simple metrics, there will be some distance pairs that cannot be reconstructed well using only a constant number of beacons. Consider for example a set of regularly spaced points on a line (or in a d-dimensional lattice); points u and v that are very close together will have a distance d_{uv} that is much smaller than the distance to the nearest beacon, rendering the upper bound obtainable from (1) useless. We therefore say that a set of beacons achieves a triangulation with distortion Δ and slack ϵ if all but an ϵ fraction of node pairs in the metric space are reconstructed with distortion Δ.

A fundamental question is then the following. Suppose we have an underlying metric space M, and desired levels of precision $\epsilon > 0$ and $\delta > 0$. Is there a function $f(\cdot,\cdot)$ (independent of the size of M) so that $f(\epsilon, \delta)$ beacons suffice to achieve a triangulation with distortion $1+\delta$ and slack ϵ? Clearly such a guarantee is not possible for every metric; in the n-point uniform metric, with all distances equal to 1, any distance that is not directly measured will have a lower bound from (1) equal to 0. Thus we ask: are there are natural classes of metric spaces that *are* triangulable in this way?

Beacon-based embedding. The recent work of Ng and Zhang on Global Network Positioning (*GNP*) [44] showed how a beacon-based set of measurements could embed all but a small fraction of Internet distances with constant distortion in low-dimensional Euclidean space, and this result touched off an active line of follow-up embedding studies in the networking literature (e.g. [11, 47, 51]). Note that the empirical guarantee for *GNP* naturally defines a notion of ϵ slack for embeddings: an ϵ fraction of all node pairs may have their distances arbitrarily distorted. Again, it is easy to see that this notion of slack is necessary for a beacon-based approach. The *GNP* algorithm forms an interesting contrast with the algorithms of Bourgain and Linial, London, and Rabinovich [8, 38] for embedding arbitrary metrics. These latter algorithms use access to the full distance matrix and build coordinates in the embedding by measuring the distance from a point to a *set* — in effect, sets that can be as large as a constant fraction of the space thus act as "super-beacons" in a way that would not be feasible to implement for all nodes in the context of Internet measurement.

In order to understand why beacon-based approaches in general, or the *GNP* algorithm in particular, achieve good performance for Internet embedding in practice, a basic question is the following: are there natural classes of metric spaces that are embeddable with constant distortion and slack ϵ, using a constant number of beacons?

The present work: Performance guarantees for beacon-based algorithms. We begin by showing that distances in a metric space M whose doubling dimension is bounded by k can be reconstructed by triangulation with distortion $1+\delta$ and slack ϵ, using a number of beacons that depends only on δ, ϵ, and dimension k, independent of the size of M. We define the *doubling dimension* here to be the smallest k such that every ball can be covered by at most 2^k balls of half the radius; we also call such a metric space 2^k-*doubling*. The point here is that we are not assuming a reconstruction method that explicitly knows anything about the doubling properties of M; rather, as long as the number of beacons is simply large enough relative to the doubling dimension, one obtains accurate reconstruction using upper and lower bounds obtained from the triangle inequality alone. Doubling metrics, which generalize the distance matrices of d-dimensional point sets in ℓ_p, have been the subject of recent theoretical interest in the context of embedding, nearest-neighbor search, and other problems [19, 27, 19, 33, 56]; and an increasing amount of work in the networking community has suggested that the bounded growth rate of balls may be a useful way to capture the structural properties of the Internet distance matrix (see e.g. [14, 44, 46, 59]). Thus, given that strong triangulation performance guarantees are not possible for general metrics (as noted above via the uniform metric), this positive result for doubling metrics serves as a plausible theoretical underpinning for the success of beacon-based triangulation in practice.

Certain non-trivial metric spaces exhibit a stronger version of triangulation that we term *perfect triangulation*: on all but an ϵ-fraction of node pairs, the upper and lower bounds from the triangle inequality agree exactly (i.e. with distortion 1). For example, one can show that $f(d, \epsilon)$ beacons suffice to achieve perfect triangulation with slack ϵ on the points of a d-dimensional lattice under the ℓ_1 metric. It is thus natural to ask how generally this phenomenon holds. Perfect triangulation turns out not be possible for all point sets in the ℓ_1 metric, but we show that it can be achieved for all *balanced* point sets in ℓ_1; by a balanced point set we mean an n-point subset of $\mathbb{R}^d$ in which the ratio of the largest to the smallest distance is $O(d\, n^{1/d})$.

We next move on to results for beacon-based embedding. We show that every metric space can be embedded into ℓ_p (for any $p \geq 1$) with constant distortion and slack ϵ, using a constant number of beacons, where the constants here depend only on ϵ. Moreover, for doubling metrics we show that an embedding with these properties can be achieved by a close analog of the actual *GNP* algorithm of Ng and Zhang, providing further theoretical explanation for its success in practice. It is interesting to note that arbitrary metric spaces (and even arbitrary doubling metrics) cannot be embedded into Euclidean space (or into ℓ_p for any $p \geq 2$) with constant distortion [38, 40], so this is a case where allowing slack leads to a qualitatively different result.

While beacon-based algorithms perform a manageable set of measurements, they do so by choosing a small set of nodes and placing a large computational and measurement load on them. Several recent networking papers [11, 47, 51] address the unbalanced load of beacon-based methods using *uniform probing*: each node selects a small number of virtual 'neighbors' uniformly at random and measures distances to them; all nodes then run a distributed algorithm that uses the measured distances. We show how an extension of our techniques here can be used to give performance guarantees for distributed algorithms such as these.

In particular, to analyze these uniform-probing embedding algorithms, we build on the techniques we develop for reasoning about triangulation. We consider graphs G on the set of nodes with the property that embeddings that approximately preserve all edge lengths in G must have constant distortion with slack ϵ for the full distance matrix. This is a kind of "rigidity" property (with slack) that follows naturally from the analysis of triangulation, and we can show that graphs consisting of node-to-beacon measurements, as well as graphs built in a more distributed fashion, can be usefully analyzed in terms of this property. We then simulate a beacon-based algorithm: instead of measuring distances to beacons directly, nodes cooperatively infer them from the probed distances via an appropriate distributed algorithm. The inferred distances to beacons are in fact upper and lower bounds on the true distances that are sufficiently precise to yield a good triangulation. To obtain an embedding from these bounds, one needs somewhat more elaborate technique than the one for the 'pure' beacon-based result; this is because the inferred distances do not quite obey the triangle inequality.

The present work: Extensions. We extend the above results on beacon-based approaches in two directions. First, we consider a more restrictive notion of slack (termed *ϵ-uniform slack*) when for each node u, at most an ϵ-fraction of node pairs (u, v) may be arbitrarily distorted. Our results on triangulation and embedding easily extend to this notion of slack, at the cost of making the number of beacons proportional to $\log n$. We achieve ϵ-uniform slack with a *constant* number of uniformly sampled beacons (the number of beacons dependent only on ϵ, distortion, and the doubling dimension) via a novel lemma on the structure of doubling metrics. We also discuss an extension to infinite metric spaces.

Second, we show that stronger guarantees can be obtained in the more restrictive class of *growth-constrained metrics*, in which doubling the radius of a ball increases its cardinality by at most a constant factor. Such metric spaces can be seen as generalized grids. They have been used as a reasonable abstraction of Internet latencies in the long line of work on locality-aware Distributed Hash Tables started by Plaxton et al. [49] (see the intro of [21] for a short survey); they have also been considered in the context of compact data structures [27], routing schemes [3], dimensionality in graphs [32], and gossiping protocols [28]. For growth-constrained metrics we obtain an embedding with a more "gracefully degrading" notion of slack: all but an ϵ-fraction of distances are embedded with distortion $\Delta = O(\log \frac{1}{\epsilon})$; all but an ϵ-fraction of the remainder are embedded with distortion 2Δ; and in general, all but an ϵ^j fraction are embedded with distortion $j\Delta$. We also show that the following simple *nearest-beacon* embedding is effective: select k beacons uniformly at random, embed the beacons, and then simply position each other node at the embedded location of its nearest beacon.

Related work. As discussed above, the questions we consider here differ from the bulk of algorithmic embedding research (as surveyed in [24, 37, 41, 25]) because we are able to measure only a small subset of the distances, and we allow a notion of slack in the performance guarantee. Indeed the whole problem of triangulation, which seeks simply to reconstruct the distances, would not be of interest if we already had access to all distances. Allowing slack changes the kinds of performance guarantees one can achieve; for example, as mentioned above, doubling metrics become embeddable with constant distortion in Euclidean space once a small slack is allowed. At the same time, we find that techniques from the body of previous work on embedding, combined with our results on triangulation, are useful in designing algorithms under these new constraints.

Work on distance labeling [16] seeks to assign a short label to each node in a graph so that the distance between u and v can be (approximately) determined from their labels alone. This is of course analogous to our goals in triangulation. In the most closely related work in this vein, Talwar [56] investigated distance labels for doubling metrics.[3] Both the objective and the techniques in [56] differ considerably from our work on network triangulation here, however: in [56], the concern is with labels of low bit complexity, but the encoding of distances into short labels there makes extensive use of the full distance matrix, and it is thus not adaptable to our setting in which distances to only a few beacons can be measured. The more extensive use of the distance matrix in [56] comes in pursuit of a stricter goal: distance labels in which there is no notion of slack in the performance guarantee.

Work on property testing [17] makes use of a somewhat different notion of slack in its performance guarantees: can an ϵ-fraction of the input be changed so that a given property holds? There has been some research on property testing in metric spaces (see e.g. [34, 45], and related work on sampling for approximating metric properties in [23]), but this work has considered problems quite different from what study here, and makes use of different sampling models and objective functions. Metric Ramsey theory [6] also seeks subsets of a metric space satisfying specific properties, but it tends to operate in a qualitatively different part of the parameter space, exploring properties that hold on the sub-metric induced by relatively small subsets of the nodes, rather than properties that hold on a large fraction of the edges. Finally, *distance geometry* [10] is a large area concerned with reconstructing point sets from sparse and imprecise distance measurements; our use of triangulation here corresponds to the notion of *triangle inequality bounds smoothing* in [10], but beyond this connection we are not aware of closely related work in the distance geometry literature.

[3] After the conference version [29] of this paper has appeared, distance labels for doubling metrics have also been considered in [53, 42, 52].

Follow-up work. There is an interesting and quite natural open question raised by our work here: Can every metric space be embedded into ℓ_p with constant distortion and ϵ-slack? To this end, we demonstrate that standard examples of metric spaces that require super-constant distortion for embeddings into ℓ_p — bounded-degree expanders and hypercubes — do not serve as counterexamples for embeddings with slack, since they can actually be embedded with constant distortion and ϵ-slack into a uniform metric; see Section 7 for details. In the follow-up work [1, 9] we resolve this question in the affirmative. We achieve distortion $O(\log \frac{1}{\epsilon})$; it is shown to be optimal up to constant factors.

Another open question concerns embeddings with gracefully degrading distortion: Can we extend our result for growth-constrained metrics to more general families of metrics? Indeed, in [1, 9] we have obtained gracefully degrading embeddings into ℓ_p for *decomposable metrics* [19, 31], a wide family of metrics that includes doubling metrics and metrics induced by planar graphs, and for arbitrary metrics in the case when the target space is (high-dimensional) ℓ_1. Subsequently, Abraham et al. [2] resolved the question in full, achieving low-dimensional gracefully degrading embeddings for arbitrary metric spaces into any ℓ_p, $p \geq 1$.

Finally, can we achieve triangulation *without* ϵ-slack? In the follow-up work we considered triangulations where instead of a single global set of beacons each node has a distinct beacon set, and for each node pair (u, v) the triangle inequality is applied for every node that is a beacon for both u and v. In this framework we achieved slack-less triangulation (with distortion $1+\delta$ and only a poly-logarithmic number of beacons per node) for doubling metrics via a centralized construction [53, 52], and for growth-constrained metrics via a distributed construction [55].

The line of work started by this paper (excluding [2]) has been brought together in the Ph.D. thesis of A. Slivkins [54]. While the main open questions directly motivated by this paper have been resolved in the follow-up work, that work has lead to a number of new open questions; see Chapter 8 of [54] for a discussion.

Organization of this paper. The present paper is an extended version of Kleinberg et al. [29], which includes closely related portions from Slivkins [53]. Specifically, the results on uniform slack (Section 4) and on embedding via uniform probing (Theorem 5.10) are from [53]. We believe that such joint exposition benefits both papers.

Sections 2 and 3 are on beacon-based triangulation and embeddings, respectively. Sections 4 concerns uniform slack. In Section 5 we discuss approaches based on uniform probing. In Section 6 we strenghten our results for beacon-based embeddings for the family of growth-constrained metrics. Finally, in Section 7 we discuss possible counterexamples for embeddings with slack.

1.1 Preliminaries.

Let us start with some notation that will be used throughout the paper. Unless specified otherwise, we denote the underlying metric space by (V, d), so that $d(u, v)$ denotes the distance between nodes u and v; we also use d_{uv} whenever typographically convenient. Let $B_u(r)$ be the closed ball of radius r around node u, i.e. $B_u(r) = \{v \in V : d_{uv} \leq r\}$. Let $r_u(\epsilon)$ be the radius of the smallest closed ball around u that contains at least ϵn nodes. The *open ball* of radius r around node u is the set of all nodes within distance strictly less than r from u. The term *ball* in a metric space refers to a *closed* ball unless specified otherwise.

For $k \in \mathbb{N}$ define $[k]$ as the set $\{0, 1 \ldots k-1\}$. Throughout the paper, n denotes the number of nodes in the input graph or metric space. All metric spaces are finite except in Section 4.1.

Our results for triangulation and embedding will generally involve showing that a large enough set of beacons sampled uniformly at random from the metric space will have a certain desired property. (For brevity, we will refer to such a sampled subset of the space as "a constant number of randomly selected beacons.") Because we will be working in many cases with constant-size samples, our properties will typically hold with a constant probability that can be made arbitrarily close to 1. Hence, in this context, we will sometimes use the phrase "with probability close to 1" as an informal short-hand for: with a probability that can be made arbitrarily close to 1 by increasing the sample size by a constant factor. Some of our guarantees have a stronger form, termed *with high probability*, which means that by increasing the sample size by a constant factor c the failure probability can be reduced to $1/n^c$.

Doubling metrics Any point set in a k-dimensional ℓ_p metric, $p \geq 1$, has the following property, called the *doubling property* [5]: for some $s = O(2^k)$ every set of diameter d can be covered by s sets of diameter $d/2$. (The diameter of a set is the maximal distance between any two points in it.) This motivates the following definition: the *doubling constant* of a metric space is the smallest s such that the above property holds. A metric space is called *s-doubling* if its doubling constant is at most s. The doubling property is often characterized via the *doubling dimension* of a metric space, defined as the $\log$ of the doubling constant.

Doubling metrics are defined as metrics with bounded (and, intuitively, low) doubling constant. By definition, doubling metrics generalize constant-dimensional ℓ_p metrics. Doubling metrics is a much wider class of metrics: in particular, there exist doubling metrics on n nodes that need distortion $\Omega(\sqrt{\log n})$ to embed into any ℓ_p, $p \geq 2$ [50, 35, 36, 19]. Interestingly, the doubling property is *robust*:

Lemma 1.1. *The doubling dimension of a subset is no larger than that of the entire metric space.*

Proof. Let α be the doubling dimension of a metric on node set V, and let S be a subset. Then any subset $S' \subset S$ can be covered by 2^α subsets $S_1, S_2, S_3, \ldots \subset V$, each of diameter $d/2$. To obtain the desired covering by 2^α subsets of S, just intersect each of the S_i's with S. □

Recall that the defining property of a doubling metric is that any set of diameter d can be covered by a constant number of sets of diameter at most $d/2$. We will use this property via a more concrete corollary where we cover with a constant number of *balls*:

Lemma 1.2. *In a metric with doubling constant s, any set of diameter d can be covered by s^k balls of radius $d/2^k$, for any integer $k \geq 1$. The desired cover can be efficiently constructed.*

Proof. Consider a set S of diameter d and apply the doubling property recursively k times. It follows that S can be covered by s^k sets of diameter at most $d/2^k$. Pick any one point from each of these sets. Then S can be covered with s^k balls of radius $d/2^k$ centered in the selected points. Moreover, it follows that the desired cover can be efficiently constructed by a simple greedy algorithm select any node $u \in S$, add the ball around u to the cover, delete from S all nodes within distance $d/2^k$ from u, repeat until S is empty. □

In fact, for all our applications it suffices to redefine the doubling property in terms of covering a large ball with balls of half the radius. Moreover, it is slightly more convenient technically; in particular, the proof of Lemma 1.2 simplifies. However, under this definition we no longer have robustness (Lemma 1.1).

Chernoff Bounds. Throughout the paper we use *Chernoff Bounds*, a standard result which says that the sum of bounded independent random variables is close to its expectation with high probability (e.g. see the textbook of Motwani and Raghavan [43] for the proof).

Theorem 1.3 (Chernoff Bounds)**.** *Consider the sum X of n independent random variables $X_i \in [0, y]$.*

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

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

2 Beacon-based triangulation

We start by defining a notion of beacon-based distance estimation via triangle inequality:

Definition 2.1. If E is a set of node pairs in a metric space, we say that E is an *ϵ-dense set* if it includes all but an ϵ-fraction of all pairs. We say that E is a *uniform ϵ-dense set* if it includes all but an ϵ-fraction of all pairs of the form (u, v) for each point u.

Definition 2.2. Given a set S of beacons, we define lower and upper distance bounds for each pair (u, v) of points: $D^-_{uv} = \max_{b \in S} |d_{ub} - d_{vb}|$ and $D^+_{uv} = \min_{b \in S}(d_{ub} + d_{bv})$. We say that S achieves a triangulation with distortion $1 + \delta$ and (uniform) slack ϵ if we have $D^-_{uv} \leq (1 + \delta)D^+_{uv}$ for a (uniformly) ϵ-dense set of node pairs (u, v). For brevity, let us call it a (uniform) *(ϵ, δ)-triangulation*.

As noted in the introduction, good triangulation bounds cannot be obtained for all metric spaces since, for example, non-trivial lower bound values D^-_{uv} cannot be achieved in the uniform metric in which all distances are 1. However, it is interesting to note that in every metric space, the upper bound D^+_{uv} actually does come within a constant factor of the true distance on all but an ϵ fraction of pairs.

Theorem 2.3. *If M is an arbitrary finite metric space, then a constant number of randomly selected beacons achieves an upper bound estimate $D^+_{uv} \leq 3d_{uv}$ for all but an ϵ-fraction of pairs (u, v) with probability at least $1 - \gamma$, where the constant depends on ϵ and γ.*

Proof. Let us fix $\epsilon, \gamma \in (0, 1)$. Let B_u be the smallest ball around node u containing at least $\epsilon n/2$ nodes. For each point u in M, and with enough beacons, at least one point in B_u will be selected as a beacon with probability $1 - \epsilon\gamma/2$. Suppose this happens, and let b be a beacon in B_u. Then all but at most $\epsilon n/2$ points v lie outside B_u or on its boundary; for any such point, we have $d_{vb} \leq d_{ub} + d_{uv} \leq 2d_{uv}$ and hence $D^+_{uv} \leq d_{ub} + d_{vb} \leq d_{uv} + 2d_{uv} = 3d_{uv}$.

Let us say that a node u is *good* if for all but at most $\epsilon n/2$ node pairs (u, v) we have $D^+_{uv} \leq 3d_{uv}$, and *bad* otherwise. We have proved that each node is bad with probability at most $\epsilon\gamma/2$. Letting N be the number of bad nodes, we have $E[N] \geq \frac{1}{2}\,\epsilon\gamma n$, so by Markov inequality we have $\Pr[N > \epsilon n/2] < \gamma$. It is easy to see that if $N \leq \epsilon n/2$ then for all but at most ϵn node pairs we have $D^+_{uv} \leq 3d_{uv}$. $\square$

The upper bound of 3 in Theorem 2.3 is tight, as shown by the shortest-path metric of the complete bipartite graph $G = K_{n,n}$ with unit-distance edges. For all non-beacon pairs (u, v) on opposite sides of G, we have $D^+_{uv} = 3d_{uv}$. With a modification of this example, we can in fact show that no algorithm given access to each node's distances to all beacons can estimate d_{uv} to within a factor better than 3 for a large fraction of pairs (u, v). Specifically, we randomly generate a graph G' by deleting each edge from $G = K_{n,n}$ with probability $\frac{1}{2}$. If u and v are on opposite sides of G', then $d_{uv} = 1$ if the edge (u, v) is present, and otherwise $d_{uv} = 3$ with probability $1 - o(1)$. But if neither u nor v is a beacon, the full set of node-to-beacon distances gives no information about the presence or absence of the edge (u, v), and hence one cannot resolve whether this distance is 1 or 3.

For doubling metrics, we have a much stronger result.

Theorem 2.4. *In any s-doubling metric space M, a constant number of randomly selected beacons achieves an (ϵ, δ)-triangulation with probability $1 - \gamma$, where the constant depends on δ, ϵ, γ, and s.*

Proof. Fix any point u. Let $r = r_u(\epsilon/3)$, and consider a large ball $B = B_u(2r/\delta)$. By our definition of r, there are only a small number of points at distance strictly less than r from u, and we will ignore our estimated distances to these points. By selecting enough beacons, we can ensure that with probability close to 1 at least one beacon b lies in $B_u(r)$. Consider any point $v \notin B$. Since b is close to u and relatively very

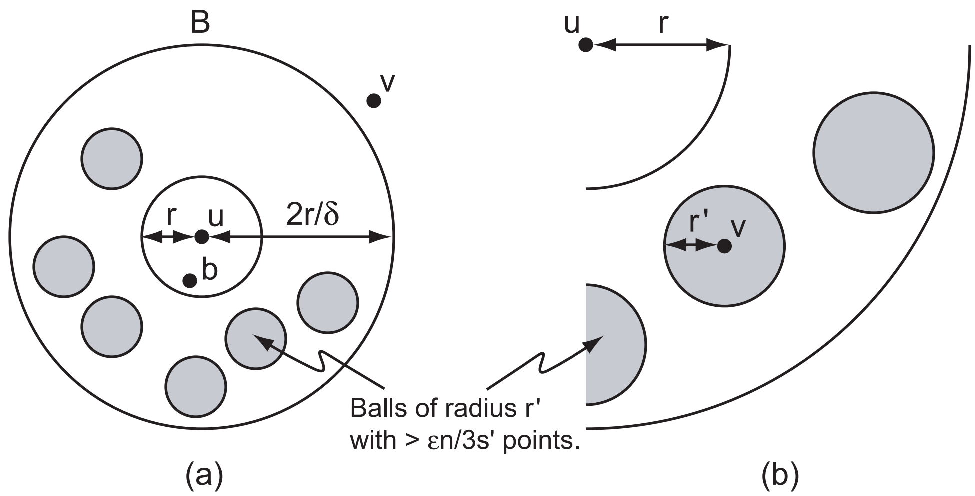

Figure 1: Triangulation in doubling metrics.

far from v, we can argue that the upper and lower bound provided by b on the distance from u to v will be good (see Figure 1a). In particular, if $d = d_{uv}$ then $d_{vb} + d_{ub} \leq d + 2d_{ub} \leq d + 2r = (1+\delta)d$, and similarly $d_{vb} - d_{ub} \geq (1-\delta)d$.

It remains to consider the possibly large set of points in the annulus $B - B_u(r)$. For these points, a beacon in $B_u(r)$ will not necessarily suffice to give the desired bound. Instead, we need to use the doubling property to show that the points in the annulus can be covered with a bounded number of very small balls, and with probability close to 1 we can ensure beacons lie in most of these. In other words, to estimate the distance d_{uv} for $v \in B - B_u(r)$, we will find a beacon close to v rather than close to u.

We would like to cover the annulus with balls of small radius $r' = \delta r/2$. By the doubling property, B (and hence $B - B_u(r)$) can be covered by $s' = (2/\delta)^{2\log s}$ balls of radius r', as shown in Figure 1(b). Disregarding balls containing fewer than $\epsilon n/3s'$ points throws out at most $\epsilon n/3$ points. Again, if we know that each of the remaining balls contains a beacon, then all points in these balls will have upper and lower bounds that are within a $1 \pm \delta$ factor of their respective distances to u.

Thus, we conclude by arguing that if we chose a sufficiently large (constant) number k of beacons, namely $k = O(s'/\epsilon)(\log \frac{1}{\epsilon})$, then each ball containing $\epsilon n/3s'$ or more points will contain a beacon with probability close to 1. So by Markov inequality with probability close to 1 a beacon will be selected in all but an $\epsilon/3$ fraction of balls containing $\epsilon n/3s'$ or more points. Combining these results shows that all but $\frac{1}{3}\epsilon n$ points have good estimated distances to all but $\frac{2}{3}\epsilon n$ points. This is the desired result. □

Remark. The above argument uses $O(\frac{1}{\epsilon}\log\frac{1}{\epsilon})\,(2/\delta)^{2\log s}$ beacons to obtain an (ϵ,δ)-triangulation with high probability. Note that a similar argument with $O(\frac{1}{\epsilon}\log n)\,(2/\delta)^{2\log s}$ beacons yields a *uniform* (ϵ,δ)-triangulation. In Section 4 we obtain a uniform (ϵ,δ)-triangulation using a number of beacons that depends only on s, ϵ and δ.

The following lemma is implicit in the proof of Theorem 2.4, and it will be very useful in our subsequent discussion of doubling metrics:

Lemma 2.5. *Consider an s-doubling metric space (V,d), fix $\epsilon, \delta \in (0,1)$, and let $\epsilon_\delta = \frac{\epsilon}{2}\left(\frac{\delta}{2}\right)^{2\log s}$. Then there exists a uniform ϵ-dense set of node pairs (u,v) such that* $\min(r_u(\epsilon), r_v(\epsilon_\delta)) \leq \delta d_{uv}$.

Perfect triangulation As mentioned in the introduction, the stronger notion of *perfect triangulation* is sometimes achievable, when $D^-_{uv} = D^+_{uv} = d_{uv}$ for all but an ϵ-fraction of node pairs, using only a constant number of beacons. A natural example where this occurs is for the points of a finite d-dimensional lattice under the ℓ_1 metric (this is a consequence of Theorem 2.6 below). It is natural to ask whether perfect triangulation is possible for all finite point sets in the ℓ_1 metric, but this is too strong; consider for example the union of the points $\{(i, n-i) : i \in [n]\}$ and $\{-i, -(n-i) : i \in [n]\}$ in the plane.

As a way to understand how general this phenomenon is, we use the following notion of a *balanced point set* as a generalization of the d-dimensional lattice: We say that a finite subset of $\mathbb{R}^d$ under the ℓ_1 metric is *balanced* if the coordinates of all points lie in the interval $[0, (kn)^{1/d}]$ for a constant k, and the minimum distance between each pair of points is 1. (We will refer to k as the *balancing parameter*.)

Theorem 2.6. *In any balanced point set M under the ℓ_1 metric, a constant number of randomly selected beacons achieves a perfect triangulation with ϵ slack and with probability $1-\gamma$, where the constant depends on ϵ, γ, the dimension, and the balancing parameter.*

Proof. We start with a proof sketch and follow up with the full proof. For ease of exposition we assume that $d = 2$, but the same techniques extend naturally to any constant dimension.

Given a balanced point set M in $[0, \sqrt{kn}]^2$, we divide M into square cells with width and height $\delta\sqrt{kn}$, for a small constant δ. We partition these cells into two types: *heavy* and *light*, where roughly speaking the heavy cells are those that contain at least $\Omega(\delta^2 n)$ points. We argue that with probability close to 1, each heavy cell will contain a beacon. Also, we can ignore errors on pairs that involve points in light cells, or that involve two points in the same heavy cell, since there are relatively few pairs like this. Thus, we only need to consider pairs of points that belong to distinct heavy cells.

We then argue that for most heavy cells C, there are heavy cells K_1, K_2, K_3, K_4 in each of the four "quadrants" of the square $[0, \sqrt{kn}]^2$ defined by treating C as the origin. This requires a geometric argument based on the balancing property; however, once the existence of K_1, K_2, K_3, K_4 is established, one beacon in each K_i is sufficient to provide a tight lower bound on any distance pair involving a point in C. Analogously, for the upper bound, we show by another application of the balancing property that for most pairs of heavy cells C and C', there is a heavy cell K in the rectangle with corners at C and C'; one beacon in K is sufficient to provide a tight upper bound on distances between points in C and C'. This completes the proof sketch.

Let us proceed with the full proof. Consider a balanced point set M in $[0, \sqrt{kn}]^2$. Divide M into cells with width and height $\delta\sqrt{kn}$, for some δ to be chosen later. There will be $\frac{1}{\delta^2}$ cells. Let x_C and y_C denote the row and column of cell C. Define $h = \min(\delta^2 n/4k, \delta^2 n\epsilon/3)$, and call a cell C *heavy* if it contains at least h points, and *light* otherwise. The idea is that we will be able to ensure that with high probability, nearly all heavy cells will contain beacons, and that a negligible number of points fall outside of the heavy cells. We will then argue that for most pairs of points that lie in heavy cells, triangulation will give matching upper and lower bounds.

Since no two points in M are within a distance of 1, no cell can have more than $4\delta^2 nk$ points. So if we let α be the fraction of cells that are heavy, then (omitting some easy arithmetic) $\alpha \geq 1/(4k+1)$.

We will begin by proving that the lower bound is correct for most pairs. Say two cells C, D are *aligned* if $x_C = x_D$ or $y_C = y_D$. Let $\mathcal{A}_C$ be the set of cells aligned with C. Note that the removal of $\mathcal{A}_C$ partitions the area into four quadrants, which we label C_1, C_2, C_3, and C_4, as shown in Figure 2(a). Say a heavy cell C is *good* if each of its four quadrants contain at least one heavy cell, and *bad* otherwise. Observe that if C is good, and all heavy cells contain beacons, then each point $x \in C$ will have correct lower bounds to each point $y \in M - \mathcal{A}_C$. (This is due to the fact that the quadrant C_i opposite to y contains a beacon.)

We now need to show that most heavy cells are good. Any heavy cell that is not good can attribute its badness to one of its quadrants. Define $\mathcal{B}_i$ for $1 \leq i \leq 4$ to be the set of heavy cells lacking a heavy cell in their i^{th} quadrant. Consider cells $C, D \in \mathcal{B}_1$ and note that $x_C + y_C \neq x_D + y_D$, since otherwise one of these cells would be to the upper-left of the other, violating our assumption. Therefore $|\mathcal{B}_1| \leq \frac{2}{\delta}$ (see Figure

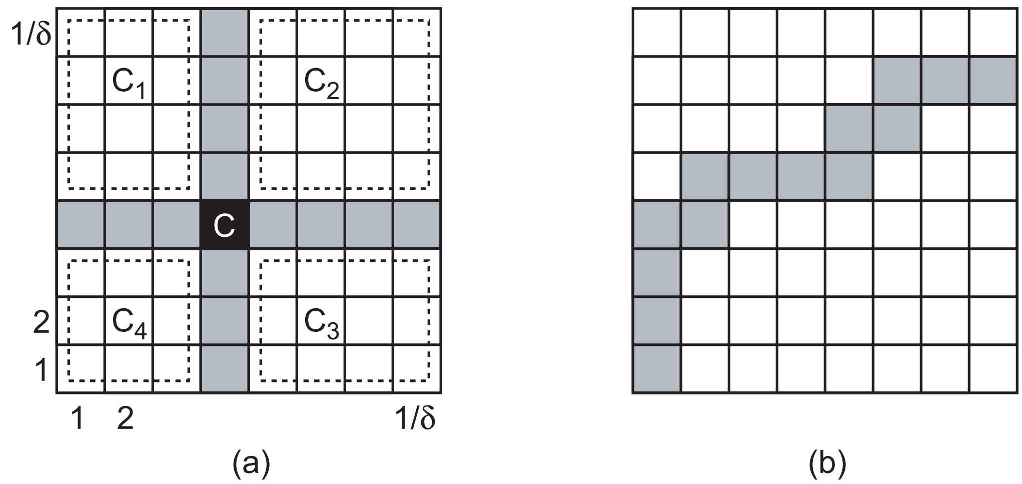

Figure 2: Balanced point sets: (a) a cell C, the set $\mathcal{A}_C$ in gray, and the corresponding quadrants; (b) a band of bad heavy cells.

2(b) for a possible $\mathcal{B}_1$ set). The argument is symmetric for all four quadrants, so in total, there can be no more than $\frac{8}{\delta}$ bad cells. Since any cell contains at most $4\delta^2 nk$ points, the total number of points in bad cells is at most $32\delta nk$. Choosing $\delta = \frac{\epsilon}{96k}$ ensures that only $\frac{\epsilon}{3}n$ points are in bad cells.

By our definition of h, the total number of points that are in light cells is also at most $\frac{\epsilon}{3}n$. Lastly, for those points in any good cell C, we have no guarantee about the lower bound to points in $\mathcal{A}_C$. But this set contains $\frac{2}{\delta} - 2$ cells, and hence fewer than $\frac{\epsilon}{3}n$ points. Hence, by selecting a large enough number of beacons, we can ensure with high probability that all but an ϵ fraction of distances have correct lower bounds.

The same general idea works for the upper bound as well. The primary difference is we need the idea of a heavy cell D being bad *relative* to some cell C, meaning there are no heavy cells in the rectangular region bounded by C and D. It is this region that needs to contain a beacon for us to have a good upper bound on distances from C to D. As before, we can show that only a small number of cells are bad relative to any other cell, and for all other cells, the calculated upper bound will be correct. The same choice of δ used above gives the desired result. □

3 Beacon-based embeddings

We now turn to the problem of beacon-based embedding. Let f map the points of M into some target metric space X with distance function d^X; we say that f is an embedding of M, and for nodes $u, v \in M$, we write d'_{uv} for $d^X_{f(u),f(v)}$. We define the *distortion* of f on a set of node pairs $E \subseteq M \times M$ to be the ratio between the maximum amount by which distances are expanded, $\max_{(u,v)\in E} d'_{uv}/d_{uv}$, and the minimum amount $\min_{(u,v)\in E} d'_{uv}/d_{uv}$. We will say that f has *non-contracting distortion* Δ on E if no distance in E is shrunk under f, and no distance is expanded by more than a factor of Δ. Following our discussion earlier, we can say that f has distortion Δ with slack ϵ if f has distortion Δ on some ϵ-dense set of pairs.

We will be able to use our triangulation analysis via the following definition.

Definition 3.1. Consider a set $S \subset V$ of beacons. Let E be the set of all node pairs (u, v) such that some beacon $b \in S$ lies within distance δd_{uv} from u or v. Let us call S a *(uniform)* (ϵ, δ)*-base* if E is a (uniform) ϵ-dense set of edges.

Note that any such set achieves a (uniform) $(\epsilon, 3\delta)$-triangulation. Let us restate the conclusions from the proof of Theorem 2.4 as follows:

Theorem 3.2. *Consider an s-doubling metric on n nodes. Let $k_0 = O(\frac{1}{\epsilon})\,(2/\delta)^{2\log s}$. Then:*

(a) $(k_0 \log \frac{1}{\epsilon})$ *randomly selected nodes form an (ϵ, δ)-base with probability close to 1.*

(b) $(k_0 \log n)$ *randomly selected nodes form a uniform (ϵ, δ)-base with high probability.*

For a set S of beacons, let E_S be the set of all node pairs (u, v) where at least one of u or v belongs to S. We show that if beacons form an (ϵ, δ)-base, for a sufficiently small δ, then in order to guarantee a low-distortion embedding with slack it suffices to achieve low distortion on E_S.

Lemma 3.3. *Consider a metric space M with an (ϵ, δ)-base S, and suppose an embedding $f : M \to X$ has non-contracting distortion Δ on E_S, where $\Delta \leq \frac{1}{4\delta}$. Then the embedding has distortion $O(\Delta)$ with slack ϵ. Furthermore, if S is a uniform (ϵ, δ)-base, then the embedding has distortion $O(\Delta)$ with ϵ-uniform slack.*

Proof. This lemma is subsumed by Lemma 5.2 in Section 5. This is because in the terminology of that section, the edge set E_S is an (ϵ, δ)-frame. □

In fact, for any beacon set S we are be able to guarantee distortion $\Delta = O(\log |S|)$ on E_S.

Lemma 3.4. *There exists an absolute constant c_0 with the following property. Consider a metric space (V, d) and a set $S \subset V$ of k beacons. Then for any $p \geq 1$ there exists an embedding into ℓ_p with $O(k \log k)$ dimensions that achieves distortion $(c_0 \log k)$ on the edge set E_S. Moreover, in this embedding the coordinates of every given node u are defined as a function of its distances to the nodes in S, and can be efficiently computed.*

Proof Sketch. We first embed S using the algorithm of Bourgain [8, 38]. Recall that this involves choosing, for each $i = 1, 2, \ldots, \lfloor \log k \rfloor$, a collection of x subsets of S of size 2^i, each uniformly at random. Let S_{ij} denote the j^{th} of these. We assign each node $b \in S$ a coordinate corresponding to each set S_{ij}, defined to be $d(b, S_{ij})$, the minimum distance between b and any point in S_{ij}.

Having embedded the beacons, we then embed every other node u using these same sets $\{S_{ij}\}$; for each S_{ij}, node u constructs a coordinate of value $d(u, S_{ij})$. In the approach of Linial et al., $x = O(\log k)$ sets of each size are chosen. Here, by way of contrast, we take $x = \Theta(k)$; we claim that with this choice of random sets $\{S_{ij}\}$ in the embedding, the set of node-beacon pairs is embedded with distortion $O(\log k)$ with probability close to 1.

To establish this claim, we give upper and lower bounds on the embedded distances; the calculations here differ from [8, 38] in that we will be taking a union bound over subsets of beacons, rather than over the much larger set of all node pairs. The upper bound is straightforward, so we focus on the lower bound. Here, we fix i and let A and A' be two disjoint subsets of S of size $k/2^i$ and $2k/2^i$ respectively. One can show there is a constant c so with probability at least c, a given S_{ij} has the property that it hits A and misses A'. Thus the expected number of S_{ij}'s with this property is ck, so applying the Chernoff bound, for large enough $x = \Theta(k)$ the probability that at most $cx/2$ of S_{ij}'s do not have this property is at most $e^{-cx/8} \leq 2^{-2k}$. Therefore with probability close to 1 for all i, for *every* pair A, A' of disjoint subsets of S of the right size, this property holds for $\Omega(k)$ sets S_{ij}. Once this is true, consider embedding any given node u, separately from all other non-beacon nodes; an analog of the telescoping-sum argument from [38] gives the desired lower bound. □

Let us define a (uniform) ***ϵ-base*** as a (uniform) $\left(\epsilon, \frac{1}{4\Delta}\right)$-base S such that $\Delta = c_0 \log |S|$. Combining the previous two lemmas, we obtain a beacon-based embedding whenever the beacons form an ϵ-base.

Theorem 3.5. *Consider a metric space (V, d) and a set $S \subset V$ of k beacons. If S is a (uniform) ϵ-base, for some $\epsilon \in (0,1)$, then there exists a $O(k \log k)$-dimensional embedding of V into ℓ_p which has distortion $O(\log k)$ with ϵ-(uniform) slack. In this embedding, the coordinates of every given node u are defined as a function of its distances to the beacons in S, and can be efficiently computed.*

We will use this theorem to obtain improved embeddings with uniform slack (Theorem 4.1); moreover, this embedding technique will be essential for our result on fully distributed embeddings in the next section.

In view of the above theorem, we need to make sure that a small set of beacons forms a (uniform) ϵ-base. Indeed, by Theorem 3.2 such beacon sets exist and can be constructed via random node selection:

Corollary 3.6. *Consider an s-doubling metric space on n nodes. Let S be the set of $k \geq 4$ randomly selected nodes. Then there exists a constant c such that:*

(a) if $k \geq (s/\epsilon)^{c \log\log(s/\epsilon)}$ then S is an ϵ-base with probability close to 1.

(b) if $k \geq x^{c \log\log x}$, $x = \frac{s}{\epsilon} \log n$, then S is a uniform ϵ-base with high probability.

Proof Sketch. Let c_0 be the constant from Lemma 3.4. We start with k and define $\delta = (4c_0 \log k)^{-1}$. Take $k_0 = O(\frac{1}{\epsilon})\,(2/\delta)^{2\log s}$ from Theorem 3.2. Then it suffices to check that $k \geq k_0 \log \frac{1}{\epsilon}$ for part (a), and that $k \geq k_0 \log n$ for part (b). □

Theorem 3.5 does not quite capture the full power of Lemma 3.3 and Lemma 3.4. We can further exploit these two lemmas to obtain a beacon-based embedding with a novel *black-box* flavor: beacons are embedded first (by inspecting only the distances between the beacons), and then the coordinates of every non-beacon node u can be computed separately by any black-box procedure that inspects the distances from u to the beacons and minimizes distortion on these distances. This closely mimics the behavior of GNP.

Definition 3.7. Consider a metric space (V, d), node set $S \subset V$, an embedding $f : S \to X$, and a node $u \notin S$. Then a ***(u, Δ)-extension*** of f is an embedding $g : S \cup \{u\} \to X$ that coincides with f on S and has distortion Δ on all node pairs (u, v) with $v \in S$.

Theorem 3.8. *Fix $p \geq 1$ and let c_0 be the constant from Lemma 3.4. For every metric space (S, d) there exists an embedding $f_{(S,d)} : S \to \ell_p^{\Theta(|S| \log |S|)}$ with distortion $c_0 \log |S|$ and the following property (*):*

Property (*). Consider a metric space (V, d) and a beacon set $S \subset V$. Let $f = f_{(S,d)}$ and let $\Delta = c_0 \log |S|$.

(a) For each node $u \notin S$ there exists a (u, Δ)-extension of f. Let g_u be any such extension.

(b) Let $g : V \to \ell_p$ be an embedding that coincides with f on S, and equals to $g_u(u)$ for every node $u \notin S$. If S is a (uniform) $\left(\epsilon, \frac{1}{4\Delta}\right)$-base, then g achieves distortion $O(\Delta)$ with ϵ-(uniform) slack.

Remark. In part (a), in order to construct a suitable g_u it suffices to inspect only the coordinates of the beacons under f and the distances from u to the beacons. A key feature of this theorem is that it requires neither any specific embedding g_u nor any specific procedure to compute it: any black-box procedure that computes a (u, Δ)-extension of f would work.

We also provide a different embedding algorithm that achieves qualitatively similar bounds: constant distortion with ϵ-slack, using a constant number of beacons. This alternate algorithm offers somewhat better quantitative guarantees but is less useful in justifying GNP. We state the result precisely as follows:

Theorem 3.9. *For any s-doubling source metric space (V, d), any target metric space ℓ_p, $p \geq 1$, and any parameter $\epsilon > 0$, we give the following two $O(\log \frac{s}{\epsilon})$-distortion embeddings:*

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

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

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

Proof. We capture the dependence on the doubling constant s via a parameter $\beta = \epsilon/2s^6$. Recall that by Lemma 2.5 there exists a uniform ϵ-dense set E of node pairs (u,v) such that $\min(r_u(\beta), r_v(\beta)) \leq \frac{1}{4}d_{uv}$. We will give a randomized algorithm that embeds any s-doubling metric d into ℓ_p with dimension $O(\log\frac{1}{\beta}\log\frac{1}{\beta\delta})$, slack ϵ and and distortion $O(\log\frac{1}{\beta})$, using only $O(\frac{1}{\beta}\log\frac{1}{\beta\delta})$ beacons, with success probability at least $1-\delta$. In particular, there exists an embedding with dimension $O(\log^2\frac{1}{\beta})$, slack ϵ and distortion $O(\log\frac{1}{\beta})$ that uses only $O(\frac{1}{\beta}\log\frac{1}{\beta})$ beacons. Note that $\log\frac{1}{\beta} = O(\log\frac{s}{\epsilon})$.

The algorithm is essentially Bourgain's algorithm [8, 38] without the smaller length scales. For each $i \in [\log\frac{1}{\beta}]$, choose $k = O(\log\frac{1}{\beta\delta})$ sets of beacons of size $1/(2^i\beta)$, call them S_{ij}. Each set is chosen independently and uniformly at random. Embed each node v into ℓ_p so that the ij-th coordinate is $k^{-1/p}\, d(v, S_{ij})$, where $d(v,S) = \min_{u\in S} d(u,v)$ is the distance between node v and set S.

Let $d_p(u,v)$ be the embedded uv-distance For simplicity we consider the case $p = 1$ first. Since for any set S we have $d_{uv} \geq |d(u,S) - d(v,S)|$, it follows that $d_1(u,v) \leq O(d_{uv}\log\frac{1}{\beta})$. The hard part is the lower bound: $d_1(u,v) = \Omega(d_{uv})$.

Fix a node pair $uv \in E$. Let $d = d_{uv}$. Let $\rho_i = \min(r_u(\beta\, 2^i), r_v(\beta\, 2^i), d/2)$. Note that the sequence ρ_i is increasing with $\rho_0 < d/4$ and $\rho_i = d/2$ for $i \geq i_0$ for some i_0.

For each i we claim that with failure probability at most $\epsilon\delta/\log\frac{1}{\beta}$ the total contribution to $d_1(u,v)$ of all sets S_{ij} is $\Omega(\rho_{i+1} - \rho_i)$. Once this claim is proved, with failure probability at most $\epsilon\delta$ the sum of these contributions telescopes:

$$d_1(u,v) = \sum \Omega(\rho_{i+1} - \rho_i) = \Omega(\rho_{i_0} - \rho_0) = \Omega(d).$$

Using Markov inequality, we have that with failure probability $O(\delta)$ this holds for an ϵ-dense set of node pairs. To make this happen for a *uniform* ϵ-dense set of node pairs (actually, for all of E) we need to replace the Markov inequality by the union bound, which is achieved by increasing the parameter k to $O(\log n)$.

It remains to prove the claim. Fix i and let $\gamma = 2^i\beta$. Without loss of generality assume that $\rho_i < d/2$ and that the ball around u reaches size γn before the ball around v does: $\rho_i = r_u(\gamma) \leq r_v(\gamma)$. A given set S_{ij} contributes at least $\frac{1}{k}(\rho_{i+1} - \rho_i)$ to $d_1(u,v)$ as long as it hits $B = B_u(\rho_i)$ and misses the open ball B' of radius ρ_{i+1} around v. By Lemma 3.10 the probability of this happenning is at least c (since the two balls are disjoint, $|B| \geq \gamma n$ and $|B'| \leq 2\gamma n$). Thus the expected number of S_{ij}'s with this property is ck, so applying the Chernoff bound, for big enough $k = O(\log\frac{1}{\beta\delta})$ the probability that less than $ck/2$ of S_{ij}'s have this property is at most $e^{-ck/8} \leq \epsilon\delta/\log\frac{1}{\beta}$. This proves the claim, and completes the proof of the theorem for the case $p = 1$.

To extend the theorem to general p, follow [38]. Let $x_{ij} = |d(u, S_{ij}) - d(v, S_{ij})|$ be the contribution of the set S_{ij}, and let $x = \log\frac{1}{\beta}$. Then $d_p(u,v) = (\frac{1}{k}\sum_{ij} x_{ij}^p)^{1/p}$, so

$$d_p(u,v) = x^{1/p}\left(\frac{1}{xk}\sum_{ij} x_{ij}^p\right)^{1/p} \geq x^{1/p}\left(\frac{1}{xk}\sum_{ij} x_{ij}\right) = x^{1/p-1}\, d_1(u,v) = x^{1/p-1}\,\Omega(d).$$

For the upper bound, recall that $x_{ij} \leq d$, so $d_p(u,v) \leq \left(\frac{1}{k}\sum_{ij} d^p\right)^{1/p} = x^{1/p}d$. Therefore the (two-sided) distortion is at most x, as required. □

We use the following lemma in the proof of Theorem 3.9. The proof is implicit in [38] but we include it for the sake of completeness.

Lemma 3.10. *There is a constant $c > 0$ with the following property. Consider a probability space with two disjoint events E and E' such that $\Pr[E] \geq \gamma$ and $\Pr[E'] \leq 2\gamma$. Let S be a set of $1/\gamma$ points sampled independently from this probability space. Then with probability at least c, S hits E and misses E'.*

Proof. Let $p = \Pr[E]$ and $p' = \Pr[E']$. Then $p' \leq \min(2p,\, 1-p) \leq \frac{2}{3}$. Treat sampling a given point as two independent random events: first it misses E' with probability $1-p'$, and then (if it indeed misses) it hits E with probability $\frac{p}{1-p'}$. Without loss of generality rearrange the order of events: first for each point we choose whether it misses E', so that

$$\Pr[\text{all points miss } E'] = (1-p')^{1/\gamma} = \left((1-p')^{1/p'}\right)^{p'/\gamma} \leq \left((1-\tfrac{2}{3})^{3/2}\right)^2 = \tfrac{1}{27}.$$

Then upon success choose whether each point hits E. Then at least one point hits E with probability at least $1-(1-p)^{1/\gamma} \geq 1-\frac{1}{e}$. So the total success probability is at least $c = (1-\frac{1}{e})/27$. □

4 Triangulation and embedding with uniform slack

Recall our results on triangulation and embeddings with uniform slack required the number of beacons which was proportional to $\log n$. It turns out that for doubling metrics we can get rid of this dependency on n and further, we extend our results to infinite metric spaces. We note in passing that the notion of ϵ-slack can be replaced by a similar notion defined with respect to an arbitrary underlying measure. We state and prove our result for finite metric spaces, and then sketch an extension to infinite metric spaces and arbitrary measures.

Theorem 4.1. *Consider an s-doubling metric space and fix $\epsilon > 0$.*

(a) For every $\delta > 0$ there exists a uniform (ϵ, δ)-base of size $k = \frac{2}{\epsilon}\,[O(\frac{1}{\delta})]^{\log s}$. Moreover, a set of $O(k \log k)$ randomly chosen nodes forms a uniform (ϵ, δ)-base with probability close to 1. Recall that using any uniform (ϵ, δ)-base as a set of beacons leads to a uniform (ϵ, δ)-triangulation.

(b) For every $p \geq 1$ there exists an embedding into $\ell_p^{O(k \log k)}$ with distortion $O(\log k)$ and ϵ-uniform slack, where $k = (\frac{s}{\epsilon})^{O(\log\log(s/\epsilon))}$; such embedding can be computed with high probability by a beacon-based algorithm with k beacons selected uniformly at random.

The key to this theorem is the following structural lemma:

Lemma 4.2. *Consider a (possibly infinite) complete metric space with doubling constant s, equipped with a probability measure μ. Let $r_u(\epsilon, \mu)$ be the radius of the smallest ball around u that has measure ϵ. Then for every $\epsilon > 0$ there exists an* **(ϵ, μ)-packing***: a family $\mathcal{F}$ of disjoint balls of measure at least ϵ/s^4 each, such that for any node u there exists a ball $B_v(r) \in \mathcal{F}$ such that $d_{uv} + r \leq 6 r_u(\epsilon, \mu)$. Moreover, if the metric is finite then such $\mathcal{F}$ can be efficiently computed.*

In this subsection we will use (ϵ, μ)-packings such that μ is the counting probability measure (a measure μ such that $\mu(u) = 1/n$ for every node u). We will need the full generality of this lemma in Section 4.1.

Proof of Lemma 4.2: Let us fix ϵ and let $r_u = r_u(\epsilon, \mu)$. For a given node u, say a ball $B_v(r)$ is *u-zooming* if it is a subset of $B_u(3r_u)$, has measure at least ϵ/s^4, and $B_v(4r)$ has measure at most ϵ.

We claim that for every node u either there exists a u-zooming ball, or there exists a node $b_u \in B_u(3r_u)$ of measure at least ϵ.

Suppose the claim does not hold for a given node u. Let $r = r_u$. By the doubling property of the metric (see Lemma 1.2), $B_u(r)$ can be covered by s^4 balls of radius $r/8$. At least one of these balls, say $B_v(r/8)$, has cardinality at least ϵ/s^4; since without loss of generality $B_v(r/8)$ overlaps with $B_u(r)$, it follows that $d_{uv} \leq \frac{9}{8}r$ and $B_v(r/2) \subset B_u(2r)$. Since there is no u-zooming ball, in particular the ball $B_v(r/8)$ is not u-zooming, so $B_v(r/2)$ has measure at least ϵ.

In fact, the above argument can be easily extended to prove the following statement: for each $i \in \mathbb{N}$ there exists a some node v_i such that $d(u, v_i) \leq \frac{9}{4}(r - \rho_i)$ and $\mu(B_{v_i}(\rho_i)) \geq \epsilon$, where $\rho_i = r/2^i$. (The proof proceeds by induction on i.)

If the metric space is finite, then for large enough i the ball $B_{v_i}(\rho_i)$ consists of only one node v_i, which therefore has measure at least ϵ, contradiction. Now if the metric space is infinite, then we have an infinite Cauchy sequence of nodes $\{v_i\}$. Since the metric space is complete, this sequence has a limit, call it v; note that $v \in B_u(3r)$. Then for each i the ball $B_v(3\rho_i)$ contains ball $B_{v_i}(\rho_i)$, hence has measure at least ϵ. Therefore node v has measure at least ϵ, contradiction. Claim proved.

In accordance with the above claim, for every given node u we define B_u to be a u-zooming ball if such a ball exists, or else we define $B_u = \{b_u\}$ where b_u is a node in $B_u(3r_u)$ that has measure at least ϵ. Note that in the finite case, a suitable B_u can be efficiently computed by simply checking whether each ball is u-zooming, and then checking each node in $B_u(2r_u)$.

Let $\mathcal{F}$ be a maximal collection of disjoint balls B_u. Note that such $\mathcal{F}$ can be computed in polynomial time by sequentially going through all balls B_u, and including a given B_u in $\mathcal{F}$ if it is disjoint from other balls that are already in $\mathcal{F}$. We will show that $\mathcal{F}$ is the desired (ϵ, μ)-packing. It suffices to prove the following claim: for each node v, some ball $B_u \in \mathcal{F}$ lies within $B_v(6r_v)$.

Suppose that for a given v the claim is false. From the definition of a v-zooming ball $B_v \subset B_v(3r_v)$, and thus $B_v \notin \mathcal{F}$. Since $\mathcal{F}$ is maximal, B_v overlaps with some ball $B_u \in \mathcal{F}$. If $B_u = \{b_u\}$ then it trivially lies in $B_v(3r_v)$, contradiction. So B_u is a u-zooming ball; say w is its center, and r is its radius. By definition of a u-zooming ball, $B_w(4r)$ has measure at most ϵ. If $4r \geq d_{vw} + r_v$, then ball $B_w(4r)$ contains ball $B_v(r_v)$; as the latter ball has measure at least ϵ, the two balls coincide, and thus B_u lies in $B_v(r_v)$, contradiction. Therefore $4r < d_{vw} + r_v$.

Recall that ball B_u overlaps with ball B_v; let x be a node that lies in both balls. Since $B_v \subset B_v(3r_v)$, applying triangle inequality to the triple (u, v, x) yields $d_{vw} \leq 3r_v + r$. Plugging this into the previous inequality, we obtain $3r < 4r_v$. It follows that $r + d_{vw} < 6r_v$. Consequently, ball $B_u = B_w(r)$ lies in the ball $B_v(6r_v)$, contradiction. Claim proved. □

Proof Sketch of Theorem 4.1: For part (a), let us fix $\epsilon, \delta \in (0, 1)$ and take $\epsilon_\delta = \frac{1}{2}\epsilon(\delta/2)^{2\log s}$ as in Lemma 2.5. Let μ be the counting probability measure, and let $\mathcal{F}_\delta$ be an (ϵ_δ, μ)-packing guaranteed by Lemma 4.2. Say $S \subset V$ is a *δ-hitting set* if it hits a ball of radius $6r_u(\epsilon_\delta)$ around every node u. Note that S is δ-hitting if it hits every ball in $\mathcal{F}_\delta$. Moreover, since the balls in $\mathcal{F}_\delta$ are disjoint and have measure at least $\epsilon^* = \epsilon_\delta/s^4$ each, it follows that $O(1/\epsilon^*)\log(1/\epsilon^*)$ randomly chosen nodes suffice to form a δ-hitting set with probability close to 1.

Let H_δ be a δ-hitting set. We claim that $H_{\delta/6}$ is a uniform (ϵ, δ)-base. Indeed, recall that by Lemma 2.5 for each node u there exists a set S_u of measure at least $1 - \epsilon$ which has the following property: for every $v \in S_u$ a ball around u or v of radius δd_{uv} has measure at least ϵ_δ. Therefore for every $v \in S_u$ some node in H_δ lies within distance $6\delta d_{uv}$ from u or v. Claim proved. It immediately follows that we can use $H_{\delta/6}$ as the beacon set to obtain the desired uniform (ϵ, δ)-triangulation.

For part (b), we claim that a set S of $k = (\frac{s}{\epsilon})^{O(\log\log(s/\epsilon))}$ randomly selected beacons is an ϵ-base with probability close to 1. Indeed, we need to use part (a) to check that S is an (ϵ, δ)-base for $\delta = (4c_0 \log k)^{-1}$, where c_0 is the constant from Lemma 3.4; we omit the details. Now part (b) follows by Theorem 3.5. □

4.1 Infinite metric spaces and arbitrary measures

Our results for beacon-based approaches are defined for finite metric spaces; ϵ-dense sets are (essentially) defined with respect to the counting measure. Let us extend them to infinite metric spaces and arbitrary probability measures. Specifically, suppose we are given a probability measure μ on V. This measure induces a product measure on node pairs. We can define an (ϵ, δ, μ)-*triangulation* and embeddings with a (ϵ, μ)-slack, where the desired properties hold for a set of edges of measure at least $1 - \epsilon$. Also, we can define a uniform (ϵ, δ, μ)-*triangulation* and embeddings with a (ϵ, μ)-uniform slack; here the desired properties hold for all node pairs (u, v), $v \in S_u$ where $\mu(S_u) \geq 1 - \epsilon$.

Our result on beacon-based embeddings (Theorem 3.9) extends to the (uniform) (ϵ, μ)-slack setting in a straightforward way. 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. Moreover, part (a) (the part about (ϵ, μ)-slack) extends to infinite metric spaces.

In order to achieve similar extensions for triangulation and for embeddings with ϵ-uniform slack, we need the machinery in this section. Specifically, for any probability measure μ on V we can prove the analog of Theorem 4.1 with (ϵ, δ, μ)-base instead of (ϵ, δ)-base, uniform (ϵ, δ, μ)-triangulation instead of uniform (ϵ, δ)-triangulation, and (ϵ, μ)-uniform slack instead of ϵ-uniform slack.

Roughly speaking, the proof of such a theorem proceeds as follows. First we note that Lemma 3.3 extends to the new setting: it suffices to guarantee low distortion on distances to beacons as long as they form an (ϵ, δ, μ)-base for a sufficiently small δ. Consequently Theorem 3.5 extends to the new setting, too. Then we just mimic the proof of Theorem 4.1 using (the full generality of) Lemma 4.2.

5 Fully distributed approaches

Recent work in the networking literature has considered so-called 'fully distributed' approaches to triangulation and embedding problems, in which no single node has to participate in a large number of measurements [11, 47, 51]. Instead, for a relatively small parameter k, each node selects k virtual 'neighbors' uniformly at random and measures distances to them; let E_k denote the set of all pairs (u, v) where v is one of the selected neighbors of u. All nodes then run a distributed algorithm that uses the measured distances on the pairs E_k to embed the full metric. The distributed algorithms in these papers are based on different heuristics: *Vivaldi* [11] simulates a network of physical springs, *Lighthouse* [47] uses global-local coordinates, and [51] claims to simulate the Big Bang explosion. They offer no proofs, but their experimental results are quite strong. In particular, *Vivaldi* [11] uses the testbed from the *GNP* algorithm [44] and claims slightly better performance. Here we consider what kinds of theoretical guarantees can be obtained for algorithms of this type; as in previous sections, we focus on doubling metrics.

First, suppose we view the distributed embedding heuristic as a black box that embeds the nodes with distortion at most Δ on the pairs E_k. Is this enough to provide a guarantee for the full metric?

Definition 5.1. Given a set E of node pairs in a metric space, we can consider the weighted graph $G(E)$ in which these pairs form the edges, and each edge (u, v) is labeled with the distance d_{uv}. We say that a uv-path P in $G(E)$ is δ-*skewed* if for some $e \in P$, the total edge weight of $P \setminus \{e\}$ is at most δd_{uv}, and e is incident to one of u or v — in other words, P consists of an initial "long hop" followed by a number of short ones. Finally, we say that the set of pairs E is a (uniform) $\boldsymbol{(\epsilon, \delta)}$**-frame** if $G(E)$ contains a δ-skewed path for all pairs in a (uniform) ϵ-dense set. We will assume throughout this section that δ is sufficiently small: $\delta < 1/4$.

Frames have a useful "rigidity" property, as the following result shows: an embedding with bounded distortion on the pairs in E must also have bounded distortion on all but an ϵ-fraction of node pairs. In this sense, frames have a similar flavor to *spanners*, but they include a slack parameter and also require the approximately distance-preserving paths to have a particular "skewed" structure.

Lemma 5.2. *Consider a metric space M with a (ϵ, δ)-frame E, and suppose an embedding $f : M \to X$ has non-contracting distortion Δ on E, where $\Delta \leq \frac{1}{4\delta}$. Then the embedding has distortion $O(\Delta)$ with slack ϵ. Furthermore, if E is a uniform (ϵ, δ)-frame, then the embedding has this distortion with ϵ-uniform slack.*

Proof. Let d^X be the distance function on X; for nodes $u, v \in M$, let us write d^*_{uv} for $d^X_{f(u),f(v)}$.

Suppose the pair (u, v) has a δ-skewed path P in $G(E)$, with long edge (u, p). By the definition of a frame combined with the triangle inequality, we have $(1-\delta)d_{uv} \leq d_{up} \leq (1+\delta)d_{uv}$. Since the embedding has non-contracting distortion Δ on E, we have $(1-\delta) \leq d^*_{up}/d_{uv} \leq \Delta(1+\delta)$ and $d_{vp} \leq \Delta\delta d_{uv}$; hence, using the assumptions that d^X is a metric and that $\delta < 1/4$, we have

$$d^*_{uv} \in [d^*_{up} - d^*_{vp},\ d^*_{up} + d^*_{vp}] \subseteq d_{uv}\,[1-\delta-\Delta\delta,\ \Delta(1+2\delta)] \subseteq d_{uv}\,[\tfrac{1}{2},\ \tfrac{3}{2}\Delta].$$

It follows that the distortion of f is $O(\Delta)$ on the set of all pairs that have a δ-skewed path. □

By Lemma 5.2, it suffices to show that the set of pairs E_k forms an (ϵ, δ)-frame for $\delta \leq \frac{1}{4\Delta}$; then we have an embedding of the full metric with distortion $O(\Delta)$ and slack ϵ.

Theorem 5.3. *Let M be a doubling metric space. There exists $k = (2\log n)^{O(1)}$ such that for any ϵ and δ that are each at least $\Omega(1/\log^{O(1)} n)$, the set E_k of probed edges is a uniform (ϵ, δ)-frame with high probability.*

The proof uses expanders; let us briefly introduce the relevant background. For an undirected graph, the (edge) *expansion* is defined as $\min \frac{|\partial(S)|}{|S|}$, where the minimum is over all nonempty sets S of at most $n/2$ vertices, and $\partial(S)$ stands for the set of edges with exactly one endpoint in S. For a pre-defined absolute constant, an *expander* is an undirected graph whose expansion is at least this constant. Expanders are well-studied and have rich applications, see [39, 4, 43, 58] for more background. We will use two standard results:

Theorem 5.4 (Folklore)**.** *An undirected graph of degree d and expansion γ has diameter at most $\frac{2d}{\gamma}\log n$.*

Theorem 5.5 (Folklore)**.** *Fix an n-node set V and a subset $Q \subset V$. Suppose for each node $u \in Q$ we choose at least $3\log n$ nodes independently and uniformly at random from Q, and create undirected links between u and these nodes. Then the induced graph on Q is an expander with high probability.*

Remark. If we choose 3 random links per node, then the resulting induced graph is an expander, but the failure probability is bounded in terms of $|Q|$. We choose $3\log n$ random links per node because for our applications we need to bound the failure probability in terms on n.

Proof of Theorem 5.3: Let s be the doubling constant of M. For some constant c to be defined later, set $\delta^* = \delta/(2\,c\log^2 n)$ and $\epsilon^* = \frac{\epsilon}{2}(\delta^*/2)^{2\log s}$. Using Chernoff bounds and taking

$$k = O(\tfrac{1}{\epsilon^*}\log n) \leq O(\tfrac{1}{\epsilon})\,(\tfrac{1}{\delta})^{2\log s}\, s^{O(1)}\,(\log n)^{1+4\log s} \tag{2}$$

suffices to make sure that with high probability each node has at least $3\log n$ neighbors, and at most $O(\log n)$ neighbors, in a ball of size $\epsilon^* n$ around every other node. By Lemma 2.5, for a uniform ϵ-dense set of node pairs uv, a ball of size $\epsilon^* n$ around one of the nodes (say v) has radius at most $\delta^* d_{uv}$. As we argued, u has a neighbor in this ball, call it w. Now, each node in this ball has at least $3\log n$ neighbors inside the same ball, chosen uniformly at random. Therefore by Theorem 5.5 the graph induced on this ball by E_k contains an $O(\log n)$-degree expander with high probability, and hence by Theorem 5.4 has diameter at most $c\log^2 n$ for some constant c. This is the c that we use in the definition of δ^* and ϵ^*. In particular, E_k contains a vw-path with at most $c\log^2 n$ hops, each of length at most $2\delta^* d_{uv}$, so the metric length of this path is at most δd_{uv}. Therefore the edge set E_k is a uniform (ϵ, δ)-frame with high probability. □

Theorem 5.3 already helps provide some underpinning for the success of distributed embedding heuristics in recent networking research. But to go beyond this black-box result to concrete distributed algorithms, we need to think about techniques for triangulation and embedding that operate in a decentralized fashion on the graph $G(E_k)$. In this section, we focus on the problem of distributed triangulation in particular.

Here's a schematic description of a distributed triangulation algorithm. First, a (small) number of nodes S declare themselves to be *beacons*. Messages are then passed over the edges of the graph $G(E_k)$, at the end of which each node u has, for each beacon b, a pair of upper and lower bounds $D^-_{ub} \leq d_{ub} \leq D^+_{ub}$. This is the crux: unlike standard beacon-based algorithms, node u never actually measures its distance to beacon b (unless they happen to be neighbors in $G(E_k)$), so it must infer bounds on the distance from the distributed algorithm. Finally, the distance between two non-beacon nodes u and v can be estimated via

$$\max_{b\in S} \left(D^-_{vb} - D^+_{ub},\ D^-_{ub} - D^+_{vb}\right) \leq d_{uv} \leq \min_{b\in S} \left(D^+_{ub} + D^+_{vb}\right).$$

We denote the left-hand and the right-hand sides by D^-_{uv} and D^+_{uv}, respectively, and say such process is a (uniform) (ϵ, δ)-*triangulation* if $D^+_{uv} \leq (1+\delta)\, D^-_{uv}$ for a (uniform) ϵ-dense set of node pairs. Note that this definition of triangulation generalizes the one for the beacon-based triangulation in Section 2: if we measure the distance between node u and beacon b, then we just set $D^+_{ub} = D^-_{ub} = d_{ub}$.

Given a set E_k of measured distances as in Theorem 5.3, our goal is to perform triangulation with only a small number of messages passed between nodes. Our algorithm completes in poly-logarithmic time, in an asynchronous message-passing model where sending or receiving a unit of data takes a unit time.

Remark. The distinction between synchronous and asynchronous models of computation is not essential to our analysis. It is perhaps more intuitive for our purposes to consider a simplistic synchronous model where in each communication round each node can send and receive a poly-logarithmic number of messages of poly-logarithmic size. Extending our result to the asynchronous model is trivial.

Theorem 5.6. *Let M be a doubling metric space, and suppose that every node has k neighbors chosen independently and uniformly at random. Then for any ϵ and δ that are each at least $\Omega(1/\log^{O(1)} n)$ there exists $k = (2\log n)^{O(1)}$ such that an (ϵ, δ)-triangulation can be achieved with high probability in time polylogarithmic in n, with only a polylogarithmic load per node, taking into account the work for distance measurements, storage, and the number of bits sent and received.*

Proof. We will use the following multi-stage algorithm:

Algorithm 5.7. *Suppose each node knows (ϵ, δ, n) and chooses (ϵ^*, k, c) as in Theorem 5.3.*

1. *Each node selects k neighbors[4] uniformly at random, measures distances to them, and decides (independently, with probability k/n) whether it is a beacon.*

2. *Beacons announce themselves to their neighbors. Specifically, each beacon b sorts its measurements from low to high and estimates $r_b(\epsilon^*)$ by the measurement ranked $2\epsilon^* k$. Call this measurement r_b. Then it sends a message $M(b, r_b, i)$ to all its neighbors, where i is the number of hops that the message has traversed, initially set to 0.*

[4]Neighbors are undirected, in the sense that if u selects v as a neighbor, then u becomes a neighbor of v, too.

3. *When node u receives $M(b, r_b, i)$ from v, node u updates its existing bounds on d_{ub} using the new bounds $d_{uv} \pm 2ir_b$. Say the message is* new *if u does not already store $M(b, r_b, i')$ with $i' \leq i$. If so and moreover $d_{uv} \leq 2r_b$ and $i < c \log n$, then u stores it and forwards $M(b, r_b, i+1)$ to all its neighbors but v.*

We now analyze this algorithm. Let $K = c \log^2 n$. Each message is forwarded at most K times, yielding the claimed running time. Each can broadcast the message from a given beacon at most K times. Therefore a given node can *receive* this message at most K times from each of its $O(k)$ neighbors. It follows that the total number of messages sent and received by a given node is poly-logarithmic. Thus, a message waits at most a poly-logarithmic time at a given node before it is forwarded, so the total completion time is poly-logarithmic.

When $M(b, r_b, i)$ is forwarded, all hops but possibly the last one have length at most $2\, r_b$, so the distance bounds in step 3 are valid.

By a straightforward application of Chernoff bounds, it holds with high probability for every beacon b that at most $2\epsilon^* k$ neighbors lie within distance $r_b(\epsilon^*)$ from b, and at least $2\epsilon^* k$ neighbors lie within distance $r_b(4\epsilon^*)$ from b, so $r_b(\epsilon^*) \leq r_b \leq r_b(4\epsilon^*)$.

Let b be a beacon, and let B_b be the smallest ball around b that has size at least $\epsilon^* n$. In the proof of Theorem 5.3 we saw that the graph induced by this ball in the edge set E_k has diameter at most K with high probability. Since $r_b \geq r_b(\epsilon^*)$, each $w \in B_b$ will receive a message from b via a path of at most K hops of length at most $2r_b$ each, so w will upper-bound d_{wb} by $D^+_{wb} \leq 2r_b K$. Moreover, since (by the proof of Theorem 5.3) every node u has a neighbor $w \in B_b$, node u will receive a message from beacon b via this node w, and consequently bound d_{ub} by $d_{uw} \pm D^+_{wb}$, which is (at worst) $d_{ub} \pm 3r_b K$. We have proved the following:

Claim 5.8. *With high probability for each node u and beacon b bounds $D^{\pm}_{ub}$ lie within $d_{ub} \pm O(r_b \log n)$.*

Now, by Lemma 2.5 there exists an ϵ-dense set of node pairs (u, v) such that the ball B around u or v of radius $r = O(\delta d_{uv} / \log n)$ has at least $4\epsilon^* n$ points. With high probability, each such ball B contains a beacon, call it b. Since $B_b(2r)$ contains B, $r_b \leq r_b(4\epsilon^*) \leq 2r$. We have proved the following:

Claim 5.9. *With high probability for each node pair (u, v) in a uniform ϵ-dense edge-set, there exists a beacon b such that $\min(d_{ub}, d_{vb}) \leq r$ and $r_b \leq r$, for some $r = O(\delta d_{uv} / \log n)$.*

It is easy to see that such beacon b yields bounds on d_{uv} that are within $d_{uv}\,(1 \pm O(\delta))$. □

Now let us extend the above algorithm for triangulation to a fully distributed algorithm that computes a low-dimensional embedding into ℓ_p, $p \geq 1$ which has low distortion with slack. In fact, for any given $\epsilon > 0$ we compute an embedding with ϵ-slack that has dimension and distortion that depend only on the doubling constant and the parameter ϵ, not on the number of nodes in the system.

Theorem 5.10. *Let M be a doubling metric space, and suppose that every node has $k = (2 \log n)^{\Omega(1)}$ neighbors chosen independently and uniformly at random. Then there exists a fully distributed algorithm that given $\epsilon \geq (\log n)^{-O(1)}$ with high probability constructs an $O(k \log k)$-dimensional embedding into ℓ_p, $p \geq 1$ which has distortion $O(\log k)$ with ϵ-uniform slack. In this algorithm the per-node load and the total completion time are at most $O(k^2 \log^3 n)$.*

In the remainder of this section we prove Theorem 5.10.

Let s be the doubling constant of M. Let us fix (ϵ, s) and assume that they are known to the participating nodes. Take $\delta = c / \log n$, where c is a constant to be specified later, and let k be defined by (2).

The high-level algorithm is simple. First the nodes compute an (ϵ, δ)-triangulation using Algorithm 5.7; note that such triangulation uses at most $O(k)$ and at least $\Omega(k)$ beacons with high probability. Then the beacons measure distances to one another and broadcast them to the entire network using a uniform gossip [48]; in this phase each beacon broadcasts one message of size $O(k)$, the total per-node load being at most $O(k^2 \log n)$. Upon receiving this information nodes update the bounds D^+ on their distances to beacons accordingly, by running a shortest-paths algorithm on the available distances. (Note that in this step D^+ can only decrease, but not below the true distance; in particular, Claim 5.8 still holds.) Finally, nodes run the embedding algorithm in Theorem 3.5 *with the same beacon set*, using the upper bounds D^+ instead of the latent true distances to the beacons.

Our proof outline follows that of Theorem 3.5, but the details are quite different and significantly more complicated. As in Theorem 3.5, first we bound the distortion on node-to-beacon distances, then use those to bound distances between other node pairs. However, we need to compensate for the fact that D^+, the distance function that we are actually embedding, is not necessarily a metric. In particular, in our proof D^+ is more than just a function that approximately obeys the triangle inequality: it is essential that D^+ is close to a specific metric, as expressed by Claim 5.8 and Claim 5.9. We will use these two claims to reason about the embedded distances to beacons, which is why we use the same set of beacons for both triangulation and embedding.

For completeness let's restate the embedding algorithm. Let S_{beac} be the beacon set from the (ϵ, δ)-triangulation; for simplicity assume there are exactly k beacons. Let $N_{\text{sc}} = \lceil \log k \rceil$. For each $i \in [N_{\text{sc}}]$ choose $\Theta(k)$ random subsets of S_{beac} of size 2^i each; let S_{ij} be the j-th of those. Here N_{sc} is the number of "size scales". These subsets S_{ij} are broadcasted to the entire network using a uniform gossip [48]: one message of size $O(k^2)$ is broadcasted, incurring a per-node load at most $O(k^2 \log n)$. Then every node u embeds itself into ℓ_p as follows. The dimensions are indexed by pairs (i, j). The coordinate in each dimension ij is defined as

$$\rho_{ij}(u) = \tfrac{1}{k}\, N_{\text{sc}}^{p-1}\, D^+(u, S_{ij}), \quad \text{where } D^+(u, S) = \min_{v \in S} D^+_{uv}.$$

Recall that we use $\Theta(k)$ beacon sets of each size scale, not $\Theta(\log k)$ as [38], in order to guarantee the following claim from the proof of Lemma 3.4:

Claim 5.11. *With high probability for any $i \in [N_{sc}]$ and any pair of disjoint subsets $S, S' \subset S_{beac}$ of size at least $k/2^i$ and at most $2k/2^i$, respectively, it is the case that at least $\Omega(k)$ sets S_{ij} hit S and miss S'.*

Then, letting d^*_{uv} be the uv-distance in the embedding, we can bound the embedded node-to-beacon distances:

Lemma 5.12. *Whp for each node u and every beacon b we have $d_{ub} \le d^*_{ub} \le O(\log k) D^+_{ub}$.*

Now by Claim 5.9 with high probability for an ϵ-dense set of node pairs (u, v) there is a beacon b within distance $O(r)$ from u or v (say, from v) such that $r_b \le O(r)$, for some $r = O(\delta d_{uv} / \log n)$. Therefore by Claim 5.8 for any such node pair (u, v) we have

$$(1 - O(\delta)) d_{uv} \le d^*_{ub} \le O(\log k) d_{uv}$$

and $d^*_{vb} \le O(\log k) \delta d_{uv}$, so it follows that

$$d_{uv}/2 \le d^*_{ub} - d^*_{vb} \le d^*_{uv} \le d^*_{ub} + d^*_{vb} \le O(\log k) d_{uv}$$

as long as the constant c that defines δ is small enough.

To complete the proof of Theorem 5.10 it remains to prove Lemma 5.12. For the upper bound thereof, we will argue that D^+ is a reasonable notion of distance. Specifically, we will prove that it satisfies the following relaxed version of triangle inequality:

Claim 5.13. $|D^+_{ub'} - D^+_{bb'}| \leq 3D^+_{ub}$ *for any node* u *and any two beacons* b, b'.

Proof. Consider the beacon b_u that is closest to u with respect to D^+; let $x = D^+(u, b_u)$ and $y = d(b', b_u)$. The beacons measure distances to each other directly, so $D^+_{bb'} = d_{bb'}$. Thus,

$$|y - D^+_{bb'}| = |d(b_u, b') - d(b, b')| \leq d(b, b_u) \leq d(u, b) + d(u, b_u) \leq 2\, D^+_{ub}. \tag{3}$$

Node u has updated $D^+_{ub'}$ according to the measurements, so $D^+_{ub'} \leq x + y$. Moreover,

$$D^+_{ub'} \geq d_{ub'} \geq d(b', b_u) - d(u, b_u) \geq y - x.$$

Therefore $|D^+_{ub'} - y| \leq x \leq D^+_{ub}$, so, using (3),

$$|D^+_{ub'} - D^+_{bb'}| \leq |y - D^+_{bb'}| + |D^+_{ub'} - y| \leq 3D^+_{ub}. \quad \square$$

Now one can use a standard metric argument to extend Claim 5.13 to sets of beacons. We write out the proof for the sake of completeness.

Corollary 5.14. $D^+(u, S) - D^+(b, S) \leq 3D^+_{ub}$ *for any node* u*, any beacon* b*, and any set of beacons* S.

Proof. Suppose $D^+(u, S) \geq D^+(b, S)$. Let b' be a beacon such that $D^+(b, S) = D^+_{bb'}$. Then

$$D^+_{ub'} \geq D^+(u, S) \geq D^+(b, S) = D^+_{bb'},$$

so by Claim 5.13 it follows that

$$|D^+(u, S) - D^+(b, S)| = D^+(u, S) - D^+(b, S) \leq D^+_{ub'} - D^+_{bb'} \leq 3D^+_{ub}.$$

The case $D^+(u, S) < D^+(b, S)$ proceeds similarly. $\square$

Proof of Lemma 5.12: For a node set S and any pair uv of nodes define $D^+_{uv}(S) = |D^+(u, S) - D^+(v, S)|$. Then the embedded uv-distance is

$$d^*_{uv} = N_{\text{sc}} \left(\frac{1}{k\, N_{\text{sc}}} \sum_{ij} D^+_{uv}(S_{ij})^p \right)^{1/p},$$

where the sum is taken over all beacon sets S_{ij}. Now if b is a beacon then by Corollary 5.14 we have $D^+_{ub}(S_{ij}) \leq D_{ub}$, which implies the claimed upper bound.

For the lower bound we will use a version of the telescoping sum argument from [8, 38]. For simplicity consider the case $p = 1$ first. Let $S_u(r)$ be the set of beacons b such that $D^+_{ub} \leq r$. For a fixed node u and beacon b, let $\rho_i = \min(\rho_u(i), \rho_v(i), d_{ub}/2)$, where $\rho_u(i)$ is the smallest r such that $S_u(r)$ contains at least $k/2^i$ beacons.

We claim that for each given i the sum $X_i = \sum_j D^+_{ub}(S_{ij})$ is at least $\Omega(k)(\rho_{i-1} - \rho_i)$. Indeed, fix i and without loss of generality assume that $\rho_u(i) \leq \rho_b(i)$. Note that the sets $S = S_u(\rho_i)$ and the interior S' of $S_b(\rho_{i-1})$ are disjoint since if a node v belongs to both S and S' then

$$d_{ub} \leq d_{uv} + d_{bv} \leq D^+_{uv} + D^+_{bv} < \rho_i + \rho_{i-1} \leq d_{ub},$$

contradiction. Therefore by Claim 5.11 with high probability for each i at least $\Omega(k)$ sets S_{ij} hit S and miss S', thus contributing at least $\rho_{i-1} - \rho_i$ each to X_i. This proves the claim.

Let $t = \lfloor \log k \rfloor$ and note that by definition $\rho_b(t) = 0$ (since $S_b(0)$ contains at least one beacon, namely b itself), so $\rho_t = 0$. Summing up the X_i's we get $d^*_{ub} \geq \Omega(k)(\rho_1 - \rho_t) = \Omega(k) d_{ub}$ as desired, as long as $\rho_1 \geq d_{ub}/4$. Now suppose $\rho_1 < d_{ub}/4$ and assume that $\rho_u(1) < \rho_b(1)$ (the case $\rho_u(1) \geq \rho_b(1)$ is treated similarly). Then the sets $S = S_u(d_{ub}/4)$ and $S' = S_{\text{beac}} \setminus S$ are disjoint and have size at least $k/2$ and at most $k/2$, respectively. Therefore by Claim 5.11 with high probability at least $\Omega(k)$ sets S_{1j} hit S and miss S', thus contributing at least $D^+_{ub}/2 = \Omega(d_{ub})$ each to X_i, so that $d^*_{ub} \geq \Omega(k) d_{ub}$ as desired. This completes the proof of the lower bound for $p = 1$.

To extend the lower bound to an arbitrary $p \geq 1$, denote the embedded uv-distance by $d^{(p)}_{uv}$ and observe that by the Generalized Mean Inequality we have $d^{(p)}_{uv} \geq d^{(1)}_{uv} \geq d_{ub}$. □

This completes the proof of Theorem 5.10.

6 Improved embeddings for growth-constrained metrics

We can obtain a number of improvements to our results when the underlying metric is *growth-constrained*, i.e. when doubling the radius of any ball increases the number of points by at most a constant factor.

We start with some background. For a k-dimensional grid and $\alpha = n + O(1)$, the following property holds: for any $x \geq 2$ the cardinality of any ball is at most x^α times smaller than the cardinality of a ball with the same center and x times the radius.[5] This motivates the following definition: the *grid dimension* of a metric space is the infimum of all α such that the above property holds. Clearly, grid dimension of any n-node metric space is at most $\log n$. *Growth-constrained metrics* are metrics such that the grid dimension is bounded by a constant.

Grid dimension is a useful notion of low-dimensionality, e.g. see [49, 21, 27, 3, 32]. It is a more restrictive notion than the doubling dimension: the latter is at most a constant times the grid dimension [19]. The converse is not true; in fact, there exist doubling metrics whose grid dimension is $\log n$. For a simple example consider the *exponential line*, which is the point set $V = \{2^i : i \in [n]\}$ equipped with a standard metric $d(x, y) = |x - y|$. Unlike the doubling dimension, grid dimension is not *robust*, in the sense that the dimension of a subset can be larger than the dimension of the entire metric space. For instance, consider the set $[n]$ with a standard metric $d(x, y) = |x - y|$. The grid dimension of such set is 1, but for a subset $[n/2] \cup \{n - 1\}$ the grid dimension is $\Omega(\log n)$.

Let us state our results. Firstly, we show that the following simple *nearest-beacon* embedding is effective in growth-constrained metrics: select k beacons uniformly at random, embed the beacons with distortion $O(\log k)$ (e.g. using the Bourgain's algorithm [8, 38]), and then simply position each non-beacon node at the embedded location of its nearest beacon. The sufficient number of beacons is then a function of grid dimension and slack ϵ.

Theorem 6.1. *Consider a metric space with grid dimension α. Then for any $\epsilon > 0$ the nearest-beacon embedding with $k = O(4^\alpha)(\frac{1}{\epsilon} \log \frac{1}{\epsilon})$ beacons has distortion $O(\alpha + \log \frac{1}{\epsilon})$ with slack ϵ.*

Combined with the fully distributed triangulation from Section 5, the nearest-beacon embedding yields a fully distributed (*Vivaldi*-style) embedding for growth-constrained metrics. Specifically, fix $\epsilon > 0$, choose k as in the above theorem, set $\delta = 1/\Theta(\log k)$ and perform a fully distributed (ϵ, δ)-triangulation from Theorem 5.6. Then for each non-beacon node u, choose the nearest beacon with respect to the triangulation (say, with respect to the upper bound D^+), and position u at the embedded location of this beacon. The proof proceeds similarly.

[5] In the literature this property is often defined for $x = 2$ only. This is essentially equivalent but slightly less convenient technically because in order to use this property one needs to round x up to the nearest power of two.

It is worth noting, on the other hand, that there are doubling metrics in which this nearest-beacon embedding does not yield good results. Specifically, consider the *exponential line* defined earlier in this section. Suppose we choose a set S of k beacons in this metric space. Then for any node $u \in V$ the nearest beacon is $b(u) = \max\{b \in S : b \leq u\}$. It is easy to argue that for most node pairs (u, v) we have $d(u, v) \gg |b(u) - b(v)|$. Indeed, if for some $u \in V$ and $\beta > 0$ there is no beacon in the interval $(u/2^{\beta}, u]$ then $b(u) \leq u/2^{\beta}$, so for any $v < u/2^{\beta}$ we have $d(u, v) \geq 2^{\beta} |b(u) - b(v)|$. Call such note pairs β-*bad*. Then for any choice of beacons all but $O(\frac{\beta k}{n})$ fraction of node pairs are β-bad (we omit the easy details).

Our second result is an embedding with gracefully degrading distortion. Recall that an embedding has *gracefully degrading distortion* $f(\epsilon)$ if for each $\epsilon > 0$ all but an ϵ-fraction of distances are embedded with distortion $f(\epsilon)$.

Theorem 6.2. *Consider a metric space with grid dimension α. Then it can be embedded into ℓ_p, $p \geq 1$ with $O(\log^2 n)$ dimensions and gracefully degrading distortion $f(\epsilon) = O(\alpha + \log \frac{1}{\epsilon})$. In particular, such embedding is achieved by Bourgain's algorithm [8, 38].*

A beacon-based version of the above theorem produces "gracefully degrading distortion with slack":

Theorem 6.3. *Consider a metric space with grid dimension α. For any $\epsilon^* > 0$ there exists a beacon-based algorithm which uses $O(\frac{1}{\epsilon^*} \log n)$ beacons and computes an embedding into ℓ_p, $p \geq 1$ with $O(\log n)(\log \frac{1}{\epsilon^*})$ dimensions and the following property: for any $\epsilon \geq \epsilon^*$, distortion on all but an ϵ-fraction of edges is $O(\alpha + \log \frac{1}{\epsilon})$.*

Remark. The embedding in Theorem 6.3 is the Bourgain's embedding [8, 38] with $O(\log \frac{1}{\epsilon^*})$ higher distance scales (i.e., scales with lower sampling density), just like in Theorem 3.9. The proof is very similar to that of Theorem 6.3 and is omitted.

Let us proceed to the proofs. We will use the grid dimension via the following simple corollary:

Lemma 6.4. *Suppose d is a metric space with grid dimension α. Fix any two nodes u, v and let $l = d(u, v)$. Then for any positive r, r^* such that $\frac{l+r}{r^*} \geq 2$ we have $|B_u(r)| \leq (\frac{l+r}{r^*})^{\alpha} |B_v(r^*)|$.*

Proof. Since $B_u(r) \subset B_v(l + r)$, we have $|B_u(r)| \leq B_v(l + r) \leq |B_v(r^*)| \, (\frac{l+r}{r^*})^{\alpha}$. ☐

We use Lemma 6.4 to derive some further structural properties which will be essential to our results. Recall that for a node u and $\epsilon \in (0, 1]$, $r_u(\epsilon)$ denotes the smallest radius of a ball around u that contains at least ϵn nodes. For brevity, let us denote $r^+_{uv}(\epsilon) = \max(r_u(\epsilon), r_v(\epsilon))$.

Lemma 6.5. *Consider a metric d with grid dimension α, and fix positive $\epsilon > 0$. Then:*

(a) for any $\delta \in (0, 1]$ there exists an ϵ-dense set of node pairs (u, v) such that $r^+_{uv}(\epsilon\, \delta^{\alpha}) \leq \delta d_{uv}$.

(b) if $d_{uv} \geq r_u(\epsilon\, 2^{\alpha})$ for some node pair (u, v) then $d_{uv} \geq r_v(\epsilon)$.

(c) for any $x \geq 1$ it is the case that $r_u(\epsilon x) \geq x^{1/\alpha}\, r_u(\epsilon)$.

Proof. For part (a), fix node u and let $r = r_u(\epsilon\, \delta^{\alpha})$. Let B be the open ball around u of radius r/δ.

To prove the lemma, it suffices to prove that there exist at least $(1 - \epsilon)n$ nodes v such that $d_{uv} \geq r/\delta$. Equivalently, we show $|B| \leq \epsilon n$. Indeed, by Lemma 6.4 for any $x > 0$ we have

$$|B_u(r/\delta - x)| \leq (\tfrac{1}{\delta})^{\alpha} |B_u(r - x/\delta)| \leq (\tfrac{1}{\delta})^{\alpha} |B_u(r)| \leq \epsilon n.$$

It follows that $|B| = \lim_{x \to +0} |B_u(r/\delta - x)| \leq \epsilon n$. This proves part (a).

For part (b), suppose not. Let $d = d_{uv}$ and $r = r_u(\epsilon\, 2^\alpha)$. Then the ball $B_v(d)$ contains less than ϵn nodes. By Lemma 6.4 it follows that

$$(\epsilon\, 2^\alpha)\, n \leq |B_u(r)| \leq |B_v(d)|\, (\tfrac{d+r}{d})^\alpha < (\epsilon\, 2^\alpha)\, n,$$

contradiction. This proves part (b).

For part (c), let $r = r_u(\epsilon)$. Note that by Lemma 6.4 for any $y > 0$ we have

$$|B_u(r\, x^{1/\alpha})| \leq |B_u(r-y)| \left(\tfrac{r\, x^{1/\alpha}}{r-y}\right)^\alpha < \epsilon x n \left(\tfrac{r}{r-y}\right)^\alpha.$$

By taking the limit $y \to +0$ we have $|B_u(r\, x^{1/\alpha})| \leq (\epsilon x) n$, which proves part (c). □

Proof of Theorem 6.1 on the nearest-beacon embedding: Let us fix $\epsilon > 0$ and set $\epsilon^* = \epsilon/4^\alpha$. Suppose we choose $k = O(\frac{1}{\epsilon^*} \log \frac{1}{\epsilon})$ beacons uniformly at random. If the constant in $O(\cdot)$ is sufficiently large, then with probability close to 1 it is the case that for at least $(1-\epsilon)n$ nodes u, there is a beacon among the $\epsilon^* n$ nodes closest to u.

Now Lemma 6.5(a) there exists an ϵ-dense set of node pairs (u, v) such that $r^+_{uv}(\epsilon^*) \leq d_{uv}/4$ and moreover there are beacons among the $\epsilon^* n$ nodes closest to u and among the $\epsilon^* n$ nodes closest to v. Let b_u and b_v be the beacons closest to u and v, respectively. It follows that $d(b_u, b_v) = \Theta(d_{uv})$. So, letting d^* be the distance in the nearest-beacon embedding, we have

$$\Omega(d_{uv}) \leq d^*(u, v) = d^*(b_u, b_v) \leq O(d_{uv}\, \log k).$$

(Without loss of generality we assume that we embed the beacons using a non-contracting embedding.) □

Proof of Theorem 6.2 on gracefully degrading distortion: Recall that Bourgain's embedding [8, 38] uses random sets S_{ij} of size 2^i, for each $i \in [\log n]$ and $j \in [k]$, where $k = O(\log n)$. The dimensions are indexed by pairs (i, j); the corresponding coordinate of node u is defined as $k^{-1/p}\; d(u, S_{ij})$. Thus, the embedded uv-distance is

$$d_p(u, v) = \left(\tfrac{1}{k} \textstyle\sum_{ij} |d(u, S_{ij}) - d(v, S_{ij})|^p\right)^{1/p}.$$

For simplicity consider the case $p = 1$ first. Then the lower bound $d_1(u, v) \geq \Omega(d_{uv})$ holds by the original Bourgain's proof.

Fix $\epsilon \in (0, 1]$. Let us say that an edge (u, v) is *long* if $d_{uv} \geq \min(r_u(\epsilon\, 2^\alpha),\, r_v(\epsilon\, 2^\alpha))$. Clearly, all but an $(\epsilon\, 2^\alpha)$-fraction of edges are long. To prove the theorem, it suffices to prove the upper bound $d_1(u, v) \leq O(d_{uv})\, (\alpha + \log \frac{1}{\epsilon})$ for all long edges (u, v). Note that the upper bound from the original Bourgain's argument is only $d_1(u, v) \leq O(d_{uv}\, \log n)$.

Let us consider a long edge (u, v). Denote the contribution of each set S_{ij} by

$$x_{ij} = |d(u, S_{ij}) - d(v, S_{ij})|.$$

Since $x_{ij} \leq d_{uv}$, it suffices to show that

$$\sum_{i > \log(1/\epsilon)} \sum_{j \in [k]} x_{ij} \leq O(d_{uv}\, k\alpha). \tag{4}$$

Let $i_0 = \log \frac{1}{\epsilon}$. Fix $i > i_0$ and let $t = (i - i_0)/2$.

Claim 6.6. $\Pr[d(u, S_{ij}) > d_{uv}\, 2^{-t/\alpha}] \leq \exp(-2^t)$.

Proof. Consider l such that $i \geq l \geq i_0$ and let $r = r_u(2^{-l})$. By Lemma 6.5(bc) we have

$$d_{uv} > r_u(\epsilon) = r_u(2^{-l}\, 2^{l-i_0}) \geq 2^{(l-i_0)/\alpha}\, r.$$

Therefore

$$\Pr\left[d(u, S_{ij}) > d_{uv}\, 2^{(i_0-l)/\alpha}\right] \leq \Pr[S_{ij} \text{ misses } B_u(r)] = \left(1 - 2^{-l}\right)^{2^i} < \exp\left(-2^{i-l}\right).$$

The claim follows if we take $l = \frac{i+i_0}{2}$. □

Let Y_j be an indicator variable for the random event $\{x_{ij} > d_{uv}\, 2^{-t/\alpha}\}$. Then $\sum_{j\in[k]} Y_j$ is a sum of independent random variables. By Claim 6.6, its expectation is at most $\mu = k\, \exp(-2^t)$. We would like to argue that this sum is not much larger than μ, with probability large enough to take a Union Bound over all long edges (u, v). A standard application of Chernoff Bounds would give

$$\Pr\left[\textstyle\sum_{j\in[k]} Y_j < O(\mu)\right] > 1 - e^{-\Omega(\mu)} \tag{5}$$

Note that for large t the failure probability in (5) is too large. Instead, let us use a non-standard version of Chernoff Bounds (Lemma 6.7 with $l = 2^t$) to show that

$$\Pr\left[\textstyle\sum_{j\in[k]} Y_j < O(k\, 2^{-t})\right] > 1 - n^{-4}. \tag{6}$$

Now we can take a Union Bound in (6) over all long edges (u, v). So we can assume that the high-probability event in (6) holds for the specific long edge that we are considering.

$$\begin{aligned}\sum_{j\in[k]} x_{ij} &\leq O(d_{uv}) \sum_{j\in[k]} (2^{-t/\alpha} + Y_j) = O(d_{uv}\, k)(2^{-t/\alpha} + 2^{-t}) \\ &= O(d_{uv}\, k)\, 2^{-(i-i_0)/2\alpha}. \end{aligned} \tag{7}$$

Summing (7) over all $i > i_0$ we obtain the desired upper bound (4), completing the proof of Theorem 6.1 for the case $p = 1$.

To extend this theorem to an arbitrary $p \leq 1$ we need a more complicated calculation than the one in [38] and Theorem 3.9. As before, fix $\epsilon > 0$, let $i_0 = \log \frac{1}{\epsilon}$, and consider a fixed long edge (u, v) and a fixed $i > i_0$. Let J be the set of all $j \in [k]$ such that $Y_j = 1$. Then, letting $d = d_{uv}$, we have

$$\begin{aligned}\sum_{j\in[k]} x_{ij}^p &= \sum_{j\in J} x_{ij}^p + \sum_{i\in[k]\setminus J} x_{ij}^p \leq |J|\, d^p + k\,(d\, 2^{-t/\alpha})^{1/p} \\ &= O(k\, d^p)(2^{-t} + 2^{-tp/\alpha}) \\ \frac{1}{k} \sum_{i>i_0} \sum_j x_{ij}^p &\leq O(d^p) \sum_{t>0} \left(2^{-t} + 2^{-tp/\alpha}\right) \\ &\leq O(d^p)(1 + \alpha/p) \\ d_p(u, v) &= \left(\frac{1}{k} \sum_{i>i_0} \sum_j x_{ij}^p + \frac{1}{k} \sum_{i\leq i_0} \sum_j x_{ij}^p\right)^{1/p} \\ &\leq O(d)(i_0 + \alpha/p)^{1/p}\end{aligned}$$

For the lower bound, we claim that $r_u(\epsilon/4^\alpha) \leq d_{uv}/4$ Indeed, by Lemma 6.4

$$\epsilon n \leq |B_u(r_u(\epsilon))| \leq 4^\alpha\, |B_u(r_u(\epsilon)/4)| \leq 4^\alpha\, |B_u(d/4)|.$$

So ball $B_u(d/4)$ contains at least $(\epsilon/4^\alpha)n$ nodes, and the claim follows. (We can also prove a similar claim for node v, although we won't use it here.) Let us denote $l = i_0 + 2s$. Given the above claim, the telescoping argument in the proof of Theorem 3.9 shows that $\sum_{i\leq l}\sum_j x_{ij} \geq \Omega(kd)$. Therefore,

$$\begin{aligned} d_p(u,v) &\geq \left(\frac{1}{k}\sum_{i\leq l}\sum_j x_{ij}^p\right)^{1/p} = l^{1/p}\left(\frac{1}{kl}\sum_{i\leq l}\sum_j x_{ij}^p\right)^{1/p} \geq l^{1/p}\left(\frac{1}{kl}\sum_{i\leq l}\sum_j x_{ij}\right) \\ &\geq \Omega(d)(i_0+\alpha)^{1/p-1} \end{aligned}$$

So the total (two-sided) distortion is at most $O(x+\alpha)$ as required. □

In the proof of the above theorem, we have used the following version of Chernoff bounds:

Lemma 6.7. *Let X_j, $j \in [k]$ be independent 0-1 random variables such that $\Pr[X_j = 1] \leq e^{-l}$ where $l > 16$. Then $\sum_{j\in[k]} X_j < k/l$ with probability at least $1 - e^{-k/2}$.*

Proof. Let $X = \sum_{j\in[k]} X_j$, $\mu = k\,e^{-l}$ and $\delta = e^l/l - 1$. Then using Chernoff Bounds we get

$$\Pr[X > \tfrac{k}{l}] = \Pr[X > (1+\delta)\mu] < e^{-\mu}\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} < \left(\frac{(el)^{1/l}}{e}\right)^k < e^{-k/2}$$

since $(el)^{1/l} < \sqrt{e}$ for any $l > 16$. □

Proof of Theorem 6.3 proceeds similarly to that of Theorem 6.2; we omit the details.

7 Embeddings with slack: beyond doubling metrics

There is an interesting and quite natural open question raised by our work here:

(*) Can every metric space be embedded into ℓ_p, $p \geq 1$ with constant distortion and ϵ slack?

To this end, we consider bounded-degree expanders[6] – a standard example of metric spaces that require super-constant distortion for embeddings into ℓ_p, $p \geq 1$. Specifically, it is known that they can be embedded into any ℓ_p, $p \geq 1$ with distortion no better than $\Omega(\log n)$ [38, 40]. Similarly, hypercubes require distortion $\Omega(\log n)$ for embedding into ℓ_2 [12]. Therefore we ask: do constant-degree expanders or hypercubes provide a counterexample for (*)? We answer this question in the negative, by showing that both can be embedded with constant distortion and slack into a uniform metric.[7]

In fact, we will use an even stronger notion of affinity to a uniform metric:

[6]The *expansion* of an undirected graph $G = (V, E)$ is defined as $\min_{S\subset V:1\leq|S|\leq n/2} \frac{|\delta(S)|}{|S|}$, where $\delta(S)$ is the set of all edges with exactly one endpoint in S. G is an *expander* if its expansion is at least some pre-defined constant.

[7]A metric is *uniform* if all distances are the same. (Another term used in the literature is *equilateral*.) Considering uniform metrics suffices since they can be embedded into ℓ_p, $p \geq 1$ with constant distortion. Isometric embedding into n dimensions is trivial: map each node i to the unit vector in the i-th dimension. Furthermore, one can obtain an embedding into $O(\log n)$ dimensions. For $p = 2$ it is a direct application of the Johnson-Lindenstrauss Lemma [26]; the case $p \in [1,2]$ follows by combining the embeddings in [26] and [13]. The general case $p \geq 1$ follows from [7].

Definition 7.1. Let $\mathcal{F}$ be a family of metric spaces that contains metrics on arbitrarily many nodes. Call $\mathcal{F}$ *asymptotically uniform* if for any given $\epsilon, \delta > 0$ and any sufficiently large n any n-node metric space in $\mathcal{F}$ can be embedded into a uniform metric with distortion $1+\delta$ and slack ϵ.

We formulate the result of this section as follows.

Theorem 7.2. *The shortest-path metrics of (a) bounded-degree expanders and (b) hypercubes are asymptotically uniform.*

Proof. **(a)** Let us consider a bounded-degree expander. Let $\beta = \alpha/d$ where α is the expansion and d is the maximal degree. Fix a node u. Any ball $B_u(r_0)$ of size $s \leq n/2$ has least αs edges coming out of it, which go to at least βs distinct nodes outside of the ball. So ball $B_u(r_0+1)$ has at least $(1+\beta)s$ nodes. Iterating this argument, ball $B_u(r_0+r)$ has at least $\min((1+\beta)^r s, n/2)$ nodes, for any $r \in \mathbb{N}$. In particular, letting $r_0 = r_u(\epsilon)$, it follows that

$$r_u(1/2) \leq r_u(\epsilon) + \log_{1+\beta}(1/\epsilon). \tag{8}$$

Similarly, any ball $B_u(r_0)$ of size $n - s \geq n/2$ has at least αs edges coming out of it, which go to at least βs distinct nodes outside of the ball. So the complement of the ball $B_u(r_0+1)$ has at most $(1-\beta)s$ nodes. Iterating this argument, we obtain $n - |B_u(r_0+r)| \leq (1-\beta)^r s$ for any $r \in \mathbb{N}$. For $r_0 = r_u(1/2) - 1$ this is at most ϵn whenever $r \geq \log_{1-\beta}(\epsilon/2)$. It follows that

$$r_u(1-\epsilon) \leq r_u(1/2) + \log_{1-\beta}(\epsilon/2). \tag{9}$$

Combining (8) and (9), we have

$$r_u(1-\epsilon) - r_u(\epsilon) \leq \log_{1+\beta}(1/\epsilon) + \log_{1-\beta}(\epsilon/2) = O_\beta(\log \tfrac{1}{\epsilon}). \tag{10}$$

We claim that for an $O(\epsilon)$-dense set E^* of node pairs, the difference between any two distances in this set is at most $O_\beta(\log \frac{1}{\epsilon})$. Indeed, consider the set of all node pairs (u,v) such that $d_{uv} < r_u(\epsilon)$ or $d_{uv} > r_u(1-\epsilon)$; these are node pairs that we ignore. Let E be the set of remaining node pairs. Then by (10) for any two node pairs that are adjacent in E the difference between their distances is at most $O(\log \frac{1}{\epsilon})$. It suffices to show that all but an $O(\epsilon)$-fraction of node pairs in E are within a constant number of E-hops from each other. This follows from the density of E: since at most $2\epsilon n^2$ node pairs are ignored, all but an $O(\epsilon)$-fraction of nodes have degree at least $\frac{2}{3}n$ in E, and any two such nodes have a common neighbor in E. Claim proved.

It remains to show that by increasing n the distances in E^* can be made arbitrarily large compared to $\log \frac{1}{\epsilon}$. Specifically, we claim that $r_u(1/2) \geq \frac{1}{d} \log \frac{n}{2}$, for any node u. Indeed, this is because for any radius r the ball $B_u(r)$ contains at most r^d nodes; claim proved. By (10) it follows that $d_{uv} \geq \frac{1}{d} \log \frac{n}{2} - O_\beta(\log \frac{1}{\epsilon})$ for any node pair (u,v) in E^*.

(b) Now let us consider hypercubes. Fix $\epsilon, \delta > 0$. Let $a = \frac{1}{1+\delta}$, $b = \frac{1}{1+\delta/2}$ and $c = \frac{1}{1-a}$. For each $j < i \leq bk/2$ we have

$$\binom{k}{j} \leq a \binom{k}{j+1} \leq \ldots \leq a^{i-j} \binom{k}{i},$$

so the number of nodes within distance $i \leq ak/2$ from a given node u in a k-dimensional hypercube is

$$\sum_{j=0}^{i} \binom{k}{j} \leq c \binom{k}{i} \leq c \binom{k}{ak/2} \leq c\, a^{(b-a)k/2} \binom{k}{bk/2} \leq O(2^k)\, a^{\Omega(k)},$$

which is less than $\epsilon 2^k$ for big enough k. Distances $i > (1+\delta)k/2$ are treated similarly. ☐

Acknowledgements. The authors would like to thank the journal referees for thorough comments.

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] I. Abraham and D. Malkhi. Name independent routing for growth bounded networks. In *17th ACM Symp. on Parallel Algorithms and Architectures (SPAA)*, pages 49–55, 2005.

[4] N. Alon and J. Spencer. *The Probabilistic Method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, New York, 2nd edition, 2000.

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

[6] 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*.

[7] Y. Bartal and M. Mendel. Dimension reduction for ultrametrics. In *15th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 664–665, 2004.

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

[9] H. T.-H. Chan, K. Dhamdhere, A. Gupta, J. Kleinberg, and A. Slivkins. Metric Embeddings with Relaxed Guarantees, 2005. Full version of [1]. Available online at `http://research.microsoft.com/users/slivkins`. Accepted for publication in *SIAM J. of Computing*, to appear. The conference version [1] combined independent efforts of two research groups, who decided to publish their full versions separately.

[10] G. Crippen and T. Havel. *Distance geometry and molecular conformation*. Wiley, New York, NY, 1988.

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

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

[13] T. Figiel, J. Lindenstrauss, and V. Milman. The dimension of almost spherical sections of convex bodies. *Acta Mathematica*, 139:53–94, 1977.

[14] 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.

[15] 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.

[16] C. Gavoille, D. Peleg, S. Perennes, and R. Raz. Distance labeling in graphs. *J. of Algorithms*, 53(1):85–112, 2004. (Preliminary version in *12th ACM-SIAM SODA*, 2001).

[17] O. Goldreich, S. Goldwasser, and D. Ron:. Property Testing and its Connection to Learning and Approximation. *J. ACM*, 45(4):653–750, 1998. Preliminary version in *37th IEEE FOCS*, 2006.

[18] K. P. Gummadi, S. Saroiu, and S. D. Gribble. King: Estimating Latency between Arbitrary Internet End Hosts. In *2nd ACM SIGCOMM Internet Measurement Workshop (IMW)*, pages 5–18, 2002.

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

[21] K. Hildrum, J. Kubiatowicz, S. Ma, and S. Rao. A note on the nearest neighbor in growth-restricted metrics. In *15th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 560–561, 2004.

[22] B. Huffaker, M. Fomenkov, D. Plummer, D. Moore, and k claffy. Distance Metrics in the Internet. In *IEEE International Telecommunications Symposium (ITS)*, 2002.

[23] P. Indyk. Sublinear time algorithms for metric space problems. In *31st ACM Symp. on Theory of Computing (STOC)*, pages 428–434, 1999.

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

[25] 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.

[26] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz maps into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

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

[28] D. Kempe, J. Kleinberg, and A. Demers. Spatial Gossip and Resource Location Protocols. *J. ACM*, 51(6):943–967, 2004. Preliminary version in *33rd ACM STOC*, 2001.

[29] 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.

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

[31] 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.

[32] R. Krauthgamer and J. R. Lee. The Intrinsic Dimensionality of Graphs. In *35th ACM Symp. on Theory of Computing (STOC)*, pages 438–447, 2003.

[33] R. Krauthgamer and J. R. Lee. Navigating nets: simple algorithms for proximity search. In *15th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 798–807, 2004.

[34] R. Krauthgamer and O. Sasson. Property testing of data dimensionality. In *14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 18–27, 2003.

[35] T. J. Laakso. Plane with A_∞-weighted metric not bi-Lipschitz embeddable to $\mathbb{R}^N$. *Bull. London Math. Soc.*, 34(6):667–676, 2002.

[36] U. Lang and C. Plaut. Bilipschitz Embeddings of Metric Spaces Into Space Forms. *Geom. Dedicata*, 87(1-3):285–307, 2001.

[37] N. Linial. Finite metric spaces – combinatorics, geometry and algorithms. In *International Congress of Mathematicians, Vol. III (Beijing, 2002)*, pages 573–586, Beijing, 2002.

[38] 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).

[39] N. Linial and A. Wigderson. Course notes: Expander graphs and their applications. http://www.math.ias.edu/ boaz/ExpanderCourse/, 2002.

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

[41] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*, chapter 15 (Embedding Finite Metric Spaces into Normed Spaces), pages 323–362. Springer-Verlag, New York, 2002.

[42] 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.

[43] R. Motwani and P. Raghavan. *Randomized algorithms*. Cambridge University Press, Cambridge, 1995.

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

[45] M. Parnas and D. Ron. Testing metric properties. *Information and Computation*, 187(2):155–195, 2003. Preliminary version in *33th ACM STOC*, 2001.

[46] R. Percacci and A. Vespignani. Scale-free behavior of the Internet global performance. *The European Physical Journal B*, 32(4):411–414, 2003. Also appeared as *arXiv e-print cond-mat/0209619*, September 2002.

[47] M. Pias, J. Crowcroft, S. Wilbur, T. Harris, and S. Bhatti. Lighthouses for scalable distributed location. In *2nd Intl. Workshop on Peer-to-Peer Systems (IPTPS)*, 2003.

[48] B. Pittel. On spreading a rumor. *SIAM J. Appl. Math.*, 47(1):213–223, 1987.

[49] 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.

[50] S. Semmes. On the nonexistence of bi-Lipschitz parameterizations and geometric problems about A1-weights. *Rev. Mat. Iberoamericana*, 12(2):337–410, 1996.

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

[52] A. Slivkins. Distance estimation and object location via rings of neighbors. In *24th Annual ACM SIGACT-SIGOPS Symp. on Principles Of Distributed Computing (PODC)*, pages 41–50, July 2005. Full version appeared in the special issue of *Distributed Computing*: **19**(4), pp. 313-333, March 2007.

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

[54] A. Slivkins. *Embedding, Distance Estimation and Object Location in Networks*. PhD thesis, Cornell University, Ithaca, NY 14850, Aug. 2006.

[55] A. Slivkins. Towards Fast Decentralized Construction of Locality-Aware Overlay Networks. In *26th Annual ACM SIGACT-SIGOPS Symp. on Principles Of Distributed Computing (PODC)*, 2007.

[56] K. Talwar. Bypassing the embedding: approximation schemes and compact representations for growth restricted metrics. In *36th ACM Symp. on Theory of Computing (STOC)*, pages 281–290, 2004.

[57] A. Vazquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of Internet. *Phys. Rev. E*, 65, 066130, 2002.

[58] D. Y. Xiao. The Evolution of Expander Graphs. BA Thesis, Harvard University, 2003. Available from *http://www.cs.princeton.edu/ dxiao*.

[59] B. Zhao, L. Huang, S. Rhea, J. Stribling, A. Joseph, and J. Kubiatowicz. Tapestry: a resilient global-scale overlay for service deployment. *IEEE Journal on Selected Areas in Communications*, 22(1):41–53, 2004.