Triangulation and Embedding using Small Sets of Beacons

Jon Kleinberg* Aleksandrs Slivkins† Tom Wexler‡
Department of Computer Science
Cornell University, Ithaca, NY 14853
{kleinber, slivkins, wexler}@cs.cornell.edu

August, 2004

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 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 metrics of bounded doubling dimension (which have been proposed as a reasonable abstraction of Internet latencies), we show that triangulation-based 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 metrics, 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 with constant distortion. We also give results for other classes of metrics, 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.

*Supported in part by a David and Lucile Packard Foundation Fellowship and NSF grants 0081334 and 0311333.
†Supported by the Packard Fellowship of the first author.
‡Supported by NSF grant 0311333.

1 Introduction

The past decade has seen many significant and elegant results in the theory of metric embeddings (for recent surveys, see e.g. [16, 25, 28]). Embedding techniques have been valuable in the design and analysis of algorithms that operate on an underlying metric; many optimization problems become more tractable when the given metric 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 [9, 12, 14, 15, 23, 37], resulting in a number of recent approaches that approximately embed this distance matrix into low-dimensional Euclidean space [5, 7, 30, 32, 36].[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. [9, 12, 14, 23]). 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 [9], 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. [10]) 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 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 metrics that *are* triangulable in this way?

Beacon-based embedding. The recent work of Ng and Zhang on Global Network Positioning (*GNP*) [30] 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. [5, 7, 32, 36]). 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 [4, 26] 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 metrics 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 (see [1, 13, 21]); we also call such a metric 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 finite d-dimensional point sets, have been the subject of recent theoretical interest in the context of embedding, nearest-neighbor search, and other problems [13, 19, 20, 21, 35]; 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. [8, 30, 31, 38]). 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 metrics exhibit a stronger phenomenon 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 *dense* point sets in L_1; by a dense point set we mean an n-point subset of $\mathbb{R}^d$ in which the ratio of the largest to the smallest distance is $\Theta(n^{1/d})$.

We next move on to results for beacon-based embedding. We show that every metric of doubling dimension k 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 on ϵ and the doubling dimension. Moreover, we show that an embedding with these properties can be achieved by a close analogue of the actual *GNP* algorithm of Ng and Zhang, providing some theoretical explanation for its success in practice. It is interesting to note that metrics of bounded doubling dimension cannot be embedded into Euclidean space (or L_p for any $p \geq 2$) with constant distortion in general [13, 34], 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 load on them. Several recent networking papers [5, 7, 32, 36] 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 beacon-based embedding algorithms, we build on the techniques we develop for reasoning about triangulation. We consider subgraphs 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 subgraphs consisting of node-to-beacon measurements, as well as subgraphs built in a more distributed fashion, can be usefully analyzed in terms of this property.

Finally, we show that stronger guarantees can be obtained in the more restrictive class of *strongly doubling metrics*. Following a definition of [19], we say that a metric is *strongly s-doubling* if doubling the radius of a ball increases its cardinality by at most a factor of s. We show that a constant number of beacons suffice to embed such metrics with constant distortion, using a more "gracefully degrading" notion of slack: all but an ϵ-fraction of distances are embedded with distortion Δ; 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 obtain improvements here for distributed algorithms that engage in uniform probing of random neighbors, in the style discussed above.

Related Work. As discussed above, the questions we consider here differ from the bulk of algorithmic embedding research (as surveyed in [16, 25, 28]) 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 [10] 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 investigated distance labels for doubling metrics [35]. Both the objective and the techniques in [35] differ considerably from our work on triangulation here, however: in [35], 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 [35] 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 [11] 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. [22, 33], and related work on sampling for approximating metric properties in [17]), 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 [2] also seeks subsets of a metric 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* [6] 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 [6], but beyond this connection we are not aware of closely related work in the distance geometry literature.

2 Triangulation

We begin with some basic definitions. For convenience let $[n]$ denote a set $\{0, 1, \ldots, n\}$. In a given metric space, let d_{uv} denote the distance between u and v, let $B_u(r)$ denote the closed ball $\{v : d_{uv} \leq r\}$, and let $r_u(\epsilon)$ be the smallest r such that $|B_u(r)| \geq \epsilon n$. Let $r^-_{uv}(\epsilon) = \min(r_u(\epsilon), r_v(\epsilon))$.

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 an (ϵ, δ)-*triangulation* if for all but an ϵ fraction of the pairs (u, v), we have $d^-_{uv} \leq (1 + \delta) d^+_{uv}$.

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.

As noted in the introduction, good triangulation bounds cannot be obtained for all metrics 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 2.1 *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 B_u be the smallest ball around 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 close to 1. 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 $d_{vb} \le d_{ub} + d_{uv} \le 2d_{uv}$ and hence $d^+_{uv} \le d_{ub} + d_{vb} \le d_{uv} + 2d_{uv} = 3d_{uv}$. □

The upper bound of 3 in Theorem 2.1 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 metrics of bounded doubling dimension, we have a much stronger result.

Theorem 2.2 *In any s-doubling metric 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.*

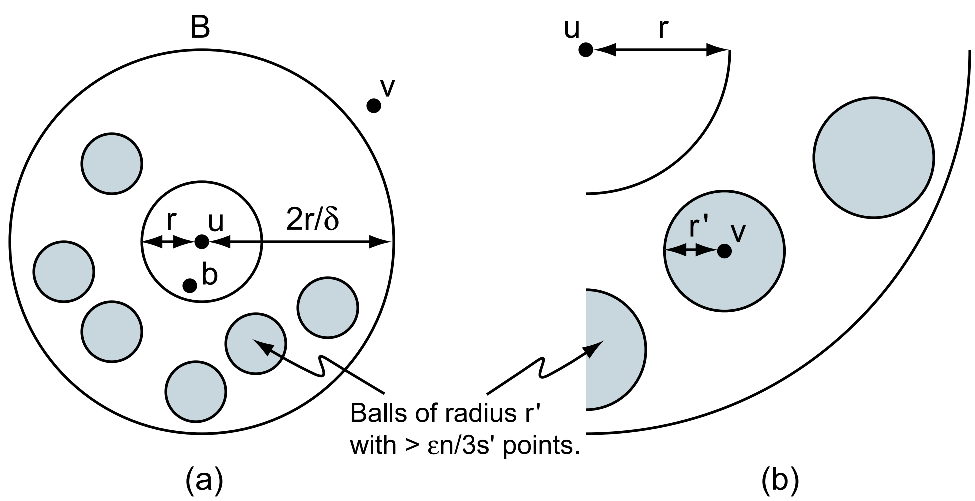

Figure 1: Triangulation in doubling metrics.

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 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 $d_{vb} + d_{ub} \le d + 2d_{ub} \le d + 2r = (1+\delta)d$, and similarly $d_{vb} - d_{ub} \ge (1-\delta)d$.

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 $s' = s^{2+2\log\frac{1}{\delta}}$ 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 of beacons, 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. □

The following lemma is implicit in the proof of Theorem 2.2, and it will be very useful in our subsequent discussion of doubling metrics. To state the lemma, we introduce the following definitions. If E is a set of pairs of points in M, we say that E is an *ϵ-set* if it includes all but an ϵ fraction of all pairs, and we say that it is a *strong ϵ-set* if it includes all but an ϵ fraction of all pairs of the form (u, v) for each point u.

Lemma 2.3 *Consider an s-doubling metric d and fixed ϵ and δ. Let $\epsilon' = \epsilon s^{2\log\delta - 1}$. Then for a strong 2ϵ-set of node pairs uv we have* $\min(r_u(\epsilon), r_v(\epsilon')) \leq \delta d_{uv}$ *and therefore* $r^-_{uv}(\epsilon') \leq \delta d_{uv}$.

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 2.4 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 *dense 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 *dense* 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 *density parameter*.)

Theorem 2.4 *In any dense 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 density parameter.*

Proof Sketch: Due to space limitations, we only provide a sketch of the proof here; the details are in Appendix A. Also, for ease of exposition we assume that $d = 2$, but the same techniques extend naturally to any constant dimension.

Given a dense 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 density 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 density 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'. □

3 Embedding with a Small Set of Beacons

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-set of pairs.

We will be able to use our triangulation analysis (particularly Lemma 2.3) via the following definition, which is phrased at a level of generality that will be useful in both this section and the next. Given a set E of node pairs in a metric, 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 (strong) (ϵ, δ)-*frame* if $G(E)$ contains a δ-skewed path for all pairs in a (strong) ϵ-set. We will assume throughout this section that δ is sufficiently small, specifically $\delta < 1/4$.

Frames E as defined here 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 3.1 *Consider a metric M with (ϵ, δ)-frame E, and suppose an embedding $f : M \to X$ has non-contracting distortion Δ on E, where $\Delta \le \frac{1}{4\delta}$. Then the embedding has distortion $O(\Delta)$ with slack ϵ.*

Proof: 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 $(1 - \delta)d_{uv} \le d_{up} \le (1 + \delta)d_{uv}$. Since the embedding has non-contracting distortion Δ on E, we have $(1 - \delta) \le d'_{up}/d_{uv} \le \Delta(1 + \delta)$ and $d_{vp} \le \Delta\delta d_{uv}$; hence, using the assumptions that X is a metric and that $\delta < 1/4$, we have

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

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

We are now in a position to discuss the performance of beacon-based embedding algorithms. We begin with a "black-box" result: in any doubling metric, an embedding will have low distortion with small slack provided it has low distortion on all measurements to a random set of beacons of constant size.

Theorem 3.2 *Let M be an s-doubling metric space, and suppose we have a black-box algorithm that for each size-k set of beacons $S \subseteq M$ will produce an embedding f_S of M into a target space X with non-contracting distortion Δ on the set of all node-beacon pairs. (That is, the set of all pairs (u, v) where at least one of u or v belongs to S.) Then provided k is large enough relative to the parameters ϵ, Δ, s, and γ, the following holds with probability at least $1-\gamma$ for a random choice of S: the embedding f_S has distortion $O(\Delta)$ with slack ϵ.*

Proof Sketch: Choose δ small enough so that $\Delta \leq \frac{1}{4\delta}$. Now Lemma 2.3 implies that with a large enough constant number of beacons relative to δ, ϵ, s, and γ, the set of all node-beacon pairs will form an (ϵ, δ)-frame with probability at least $1 - \gamma$. Hence Lemma 3.1 implies that an embedding of M into X with non-contracting distortion Δ on this set of pairs will have distortion $O(\Delta)$ with slack ϵ. □

We now turn this black-box result into an algorithm that embeds a doubling metric in $\mathbb{R}^d$ with constant distortion and ϵ-slack. For reasons of space, we only provide a sketch of the proof.

Theorem 3.3 *Let M be an s-doubling metric space, and choose a constant k so that $\frac{\log k}{\log \log k} = \Theta(\log \frac{s}{\epsilon})$. There is an algorithm that, with probability at least $1 - \gamma$, embeds M into $\mathbb{R}^d$ with distortion $O(\log k)$ and slack ϵ, using measurements to a randomly selected set of k beacons; here γ and d are constants that depend on k.*

Proof Sketch: Denote the set of node-beacon pairs by E. First we claim that with probability close to 1, E is an (ϵ, δ)-frame for $\delta = O(\log^{-1} k)$, where the constant can be adjusted by tuning the constant in the definition of k. Indeed, letting $k = \Theta(x \log x)$ and skipping some easy details, we have $\frac{\log x}{\log \log x} \geq \Omega(\log \frac{s}{\epsilon})$ and $\log_s(\epsilon x) \geq \Omega(\log \log k)$, where the constants can be adjusted similarly. The claim follows since by Lemma 2.3 E is an (ϵ, δ)-frame with probability close to 1, as long as $\log \frac{1}{\delta} = \frac{1}{2} \log_s(\epsilon x)$.

Now, by Theorem 3.2, it suffices to embed M into $\mathbb{R}^d$ so that the distortion on pairs in E is $O(\log k)$. (The embedding can be re-scaled so that it is also non-contracting on E.) We perform this embedding as follows. Let B be the set of beacons; we first embed B using the algorithm of Bourgain and Linial et al. [4, 26]. Recall that this involves choosing, for each $i = 1, 2, \ldots, \lfloor \log k \rfloor$, a collection of x subsets of B of size 2^i, each uniformly at random. Let S_{ij} denote the j^{th} of these. We assign each node $b \in B$ 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 [4, 26] 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 S and S' be two disjoint subsets of B 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 S and misses S'. 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 S, S' of disjoint subsets of B 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 analogue of the telescoping-sum argument from [26] gives the desired lower bound with probability close to 1. □

The above embedding follows the *GNP* [30] framework, in which the beacons are embedded first, and then each other node is embedded separately with respect to the beacons. This therefore provides some theoretical explanation for *GNP*'s strong empirical performance. 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*. This issue is discussed further in Appendix B.

4 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 perform a large number of measurements [5, 7, 32, 36]. 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* [5, 7] simulates a network of physical springs, *Lighthouse* [32] uses global-local coordinates, and [36] claims to simulate the Big Bang explosion. They offer no proofs, but their experimental results are quite strong. In particular, *Vivaldi* [5] uses the testbed from the *GNP* algorithm [30] 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? By Theorem 3.2, 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 4.1 *Let M be an s-doubling metric, and $k = s^{O(1)}(\log n)^{2\log s + O(1)}$. 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 strong (ϵ, δ)-frame with high probability.*

Proof: Indeed, for some constant c to be defined later, set $\delta' = \delta/(c \log n)$ and $\epsilon' = \epsilon s^{2\log \delta' - 1}/2$, so that $k = O(\frac{1}{\epsilon'} \log n)$ suffices to make sure that with high probability each node has at least three neighbors in a ball of size $\epsilon' n$ around every other node. By Lemma 2.3, for a strong ϵ-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 three neighbors in it, chosen uniformly at random. Therefore the graph induced by this ball in E_k contains a constant-degree expander, and hence has diameter at most $c \log n$. This is the c we use in the definition of δ' and ϵ' (it is enough to use an upper bound in which we assume the induced graph has n nodes). In particular, E_k contains a vw-path with at most $c \log n$ hops, each of length at most $\delta' d_{uv}$, so the metric length of this path is at most δd_{uv}. Therefore E_k is a strong (ϵ, δ)-frame. □

Theorem 4.1 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 *quasi-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 quasi-beacon b, a pair of upper and lower bounds $l_{ub} \leq d_{ub} \leq h_{ub}$. 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 nodes u and v can be estimated via

$$\max_{b \in S}(|h_{ub} - l_{vb}|, |h_{vb} - l_{ub}|) \leq d_{uv} \leq \min_{b \in S}(h_{ub} + h_{vb}).$$

We denote the left-hand and the right-hand sides by d^-_{uv} and d^+_{uv}, respectively, and say such process is a *quasi-(ϵ, δ)-triangulation* if $d^+_{uv}/d^-_{uv} \leq 1+\delta$ for an ϵ-set of node pairs. Given a set E_k of measured distances as in Theorem 4.1, our goal is to perform quasi-triangulation with only a small number of messages passed between nodes.

Theorem 4.2 *Let M be an s-doubling metric. For any ϵ and δ that are each at least $\Omega(1/\log^{O(1)} n)$, a quasi-(ϵ, δ)-triangulation can be achieved in time polylogarithmic in n with only an expected polylogarithmic load per node, taking into account the work for distance measurements, storage, and the number of bits sent and received.*

We'll use the following multi-stage algorithm. For simplicity, we refer to quasi-beacons as beacons.

Algorithm 4.3 *Suppose each node knows (ϵ, δ, n) and chooses (ϵ', k, c) as in Theorem 4.1.*

1. *Each node selects k neighbors*[3] *uniformly at random, measures distances to them, and decides (independently, with probability k/n) whether it is a quasi-beacon.*

2. *Beacons announce themselves to their neighbors. Specifically, each quasi-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.*

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 $K = c \log n$. Each message is forwarded at most K times, yielding the claimed running time. A given node can broadcast the message from a given beacon at most K times, yielding the claimed number of messages per node. When $M(b, r_b, i)$ is forwarded, all hops but possibly the last one have length at most 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_u be the smallest ball around u that has size at least $\epsilon' n$. In the proof of Theorem 4.1 we saw that the graph induced by each such ball in E_k has diameter at most K. 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 $h_{wb} \leq 2r_b K$. Moreover, since (by the proof of Theorem 4.1) every node u has a neighbor $w \in B_b$, node u will receive a message from b via this w and bound d_{ub} by $d_{uw} \pm h_{wb}$, which is (at worst) $d_{ub} \pm 3r_b K$.

Now, by Lemma 2.3 there exists an ϵ-set of node pairs uv 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, $r_b \leq r_b(4\epsilon') \leq 2r$. Therefore, omitting a few details, beacon b yields bounds on d_{uv} that are (at worst) $d(1 \pm O(\delta))$. This completes the proof of Thm. 4.2.

5 Extensions and Further Directions

Strongly doubling metrics. We can obtain a number of improvements to our results when the given metric is strongly doubling. (Recall that a metric is *strongly s-doubling* [19] if doubling the radius of a ball increases its cardinality by at most a factor of s.) We start with an improved Bourgain-style embedding; with a careful accounting argument we show that it has a 'graceful degradation' property: for this single embedding and *any* $\epsilon > 0$, the distortion is $O(\log \frac{s}{\epsilon})$ for a strong ϵ-set of node pairs. As in [26], we consider the case of L_1 first; however, extending to the general L_p is more involved than in [26]. This graceful degradation property

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

should also be contrasted with the lower bounds on embedding presented in [13]. We discuss this result in more detail in Appendix C.

We also show that the following simple *nearest-beacon* embedding is effective in strongly doubling metrics: select k beacons uniformly at random, embed the beacons, and then simply position each other node at the embedded location of its nearest beacon. It is not hard to show that in strongly-doubling metrics the nearest-beacon embedding is (essentially) as accurate as triangulation. (It is worth noting, on the other hand, that there are doubling metrics in which this nearest-beacon embedding does not yield good results, even allowing constant slack.) Combined with Algorithm 4.3, the nearest-beacon embedding yields a fully distributed (*Vivaldi*-style) embedding for strongly doubling metrics. Moreover, such an embedding will have the 'graceful degradation' property if the beacons can embed themselves using the improved Bourgain-style algorithm described above.

Embeddability with ϵ-slack for general metrics. Finally, there is an interesting and quite natural open question raised by our work here: Can every metric be embedded into L_p with constant distortion and ϵ slack? Standard examples of metrics that require super-constant distortion for embeddings into L_p — e.g., bounded-degree expanders — do not serve as counterexamples here, since they can actually be embedded with constant distortion and ϵ slack into a uniform metric. We discuss this further in Appendix D.

Acknowledgments. We thank Paul Francis, Martin Pál, Mark Sandler, and Gun Sirer for useful discussions on this topic.

References

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

[2] Y. Bartal, N. Linial, M. Mendel and A. Naor, "On Metric Ramsey-Type Phenomena," in *35th Annual ACM Symposium on the Theory of Computing*, 2003.

[3] B. Bollobas and O. Riordan, "The diameter of a scale-free random graph," preprint 2000.

[4] J. Bourgain, "On Lipschitz embeddings of finite metric spaces in Hilbert space," *Israel J. of Math*, **52**(1-2), pp. 46-52, 1985.

[5] R. Cox, F. Dabek, F. Kaahoek, J. Li and R. Morris, "Practical, Distributed Network Coordinates," in *2nd Workshop on Hot Topics in Networks* (HotNets), 2003.

[6] G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation*, Wiley, 1988.

[7] R. Cox and F. Dabek, "Learning Euclidean coordinates for Internet hosts," MIT ETR 2003.

[8] M. Fomenkov, k. claffy, B. Huffaker, D. Moore. "Macroscopic Internet topology and performance measurements from the DNS root name servers," in *USENIX Large Installation System Administration Conference* (LISA), 2001.

[9] 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*, 2001.

[10] C. Gavoille, D. Peleg, S. Perennes, R. Raz "Distance Labeling in Graphs," In *11th ACM-SIAM Symposium on Discrete Algorithms*, 2000.

[11] O. Goldreich, S. Goldwasser and D. Ron, "Property Testing and its Connection to Learning and Approximation", in *37th Annual IEEE Symposium on Foundations of Computer Science*, 1996.

[12] K.P. Gummadi, S. Saroiu and S.D. Gribble, "King: Estimating Latency between Arbitrary Internet End Hosts," in *ACM SIGCOMM Internet Measurement Workshop*, 2002.

[13] A.Gupta, R. Krauthgamer and J.R. Lee, "Bounded Geometries, Fractals, and Low-Distortion Embeddings," in *44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.

[14] J.D. Guyton and M.F. Schwartz, "Locating Nearby Copies of Replicated Internet Servers," in *ACM SIGCOMM*, 1995.

[15] B. Huffaker, M. Fomenkov, D.J. Plummer, D. Moore and k claffy, "Distance Metrics in the Internet," IEEE International Telecommunications Symposium, 2002.

[16] P. Indyk, "Algorithmic applications of low-distortion geometric embeddings," survey, in *40th Annual IEEE Symposium on Foundations of Computer Science*, 1999.

[17] P. Indyk. "Sublinear Time Algorithms for Metric Space Problems," in *40th Annual IEEE Symposium on Foundations of Computer Science*, 1999.

[18] W.B. Johnson and G. Schechtman, "Embedding l_p^m into l_1^n," *Acta Mathematica*, **149**, pp. 71-85, 1982.

[19] D. Karger and M. Ruhl, "Finding Nearest Neighbors in Growth-restricted Metrics," in *34th Annual ACM Symposium on the Theory of Computing*, 2002.

[20] R. Krauthgamer and J.R. Lee "The Intrinsic Dimensionality of Graphs," in *35th Annual ACM Symposium on the Theory of Computing*, 2003.

[21] R. Krauthgamer and J.R. Lee "Navigating nets: Simple algorithms for proximity search," In *15th ACM-SIAM Symposium on Discrete Algorithms*, 2004.

[22] R. Krauthgamer and O. Sasson, "Property testing of data dimensionality," In *14th ACM-SIAM Symposium on Discrete Algorithms*, 2003.

[23] C. Kommareddy, N. Shankar and B. Bhattacharjee, "Finding close friends on the Internet," in *9th IEEE International Conference on Network Protocols* (ICNP), 2001.

[24] T. Leighton and S. Rao, "An approximate max-flow min-cut theorem for uniform multicommodity flow problem with applications to approximation algorithms," in *29th Annual IEEE Symposium on Foundations of Computer Science*, 1988.

[25] N. Linial, "Finite metric spaces - combinatorics, geometry and algorithms," *Proc. International Congress of Mathematicians* III, pp. 573-586 Beijing, 2002

[26] N. Linial, E. London and Yu. Rabinovich, "The geometry of graphs and some of its algorithmic applications," *Combinatorica* (1995) **15**, pp. 215-245.

[27] J. Matousek, "On embedding expanders into l_p spaces," *Israel J. of Mathematics* **102** (1997).

[28] J. Matousek and P. Indyk, Chapter on embeddings in the *Handbook of Discrete and Computational Geometry*, J.E. Goodman and J. O'Rourke, eds, CRC Press LLC, FL 1997.

[29] M. Mihail, C.H. Papadimitriou and A. Saberi, “On Certain Connectivity Properties of the Internet Topology,” in *44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.

[30] E. Ng and H. Zhang, ”Predicting Internet Network Distance with Coordinates-Based Approaches”, in *IEEE INFOCOM*, 2002.

[31] R. Percacci and A. Vespignani ”Scale-free behavior of the Internet global performance,” *arXiv e-print cond-mat/0209619*, September 2002.

[32] M. Pias, J. Crowcroft, S. Wilbur, S. Bhatti and T. Harris ”Lighthouses for Scalable Distributed Location,” in *2nd International Workshop on Peer-to-Peer Systems* (IPTPS), 2003.

[33] D. Ron and M. Parnas, “Testing Metric Properties,” in *33th Annual ACM Symposium on the Theory of Computing*, 2001.

[34] S. Semmes. ”On the nonexistence of bi-Lipschitz parametrizationsand geometric problems about A_∞ weights,” *Rev. Mat. Iberoamericana* **12** (1996).

[35] K. Talwar, ”Bypassing the embedding: Approximation schemes and Compact Representations for growth restricted metrics,” in *36th Annual ACM Symposium on the Theory of Computing*, 2004.

[36] T. Tankel and Y. Shavitt, ”Big-bang simulation for embedding network distances in Euclidean space,” in *IEEE INFOCOM*, 2003.

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

[38] B. Y. Zhao, J. D. Kubiatowicz, A. D. Joseph. ”Tapestry: An Infrastructure for Fault-Tolerant Wide-Area Location and Routing,” UC Berkeley Computer Science Division, Report No. UCB/CSD 01/1141, April 2001.

Appendix A: Dense Point Sets under L_1 Metric

The proof of this theorem extends the discussion in Section 2.

Theorem A.1 *Any dense point set M under the L_1 metric can be perfectly triangulated.*
Proof: For easy of exposition we assume that $d = 2$, but the same techniques extend naturally to any constant dimension.

Consider a dense 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.

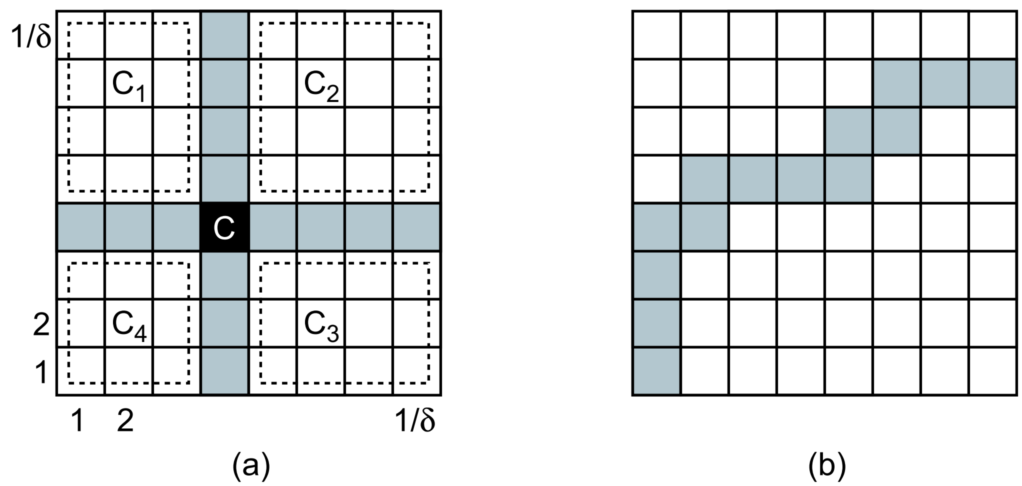

Figure 2: Dense point sets: (a) a cell C, $\mathcal{A}_C$ in gray, and corresp. quadrants; (b) a band of bad heavy cells.

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 dense 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 dense cells contain beacons, then all points in C will have correct lower bounds to all points in $M - \mathcal{A}_C$.

We now need to show that most dense cells are good. Any dense 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 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 containts 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. □

Appendix B: Bourgain-like Embedding for Doubling Metrics

The exact statement and the proof of this theorem extend the discussion in Section 3.

Theorem B.1 *Fix constants $\epsilon > 0$, $s > 1$ and $p \geq 1$. Then any s-doubling metric can be ϵ-embedded into L_p with constant dimension and constant distortion, using only a constant number of beacons. We get a strong ϵ-embedding with $O(\log n)$ dimensions and beacons. Moreover, we provide an efficient randomized algorithm.*

Proof: Using Lemma 2.3a we capture the dependence on the doubling constant s via $\beta = \frac{1}{2}\epsilon/s^5$ such that $r^-_{uv}(\beta) \leq \frac{1}{4}d_{uv}$ for a strong ϵ-set E of node pairs. We'll give a randomized algorithm that ϵ-embeds any s-doubling metric d into L_p with dimension $O(\log \frac{1}{\beta} \log \frac{1}{\beta\delta})$ 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 ϵ-embedding with dimension $O(\log^2 \frac{1}{\beta})$, distortion $O(\log \frac{1}{\beta})$ and $O(\frac{1}{\beta} \log \frac{1}{\beta})$ beacons. Note that for a *strongly* s-doubling metric $\beta = \epsilon/s^2$ would suffice by Lemma C.2. In either case $\log \frac{1}{\beta} = O(\log \frac{s}{\epsilon})$.

The algorithm is essentially the Bourgrain's algorithm without the smaller length scales. For each $i \in [\log \frac{1}{\beta}]$, choose $k = O(\log \frac{1}{\beta\delta})$ uar sets of beacons of size $1/(2^i\beta)$, call them S_{ij}. Embed each node v into L_1 so that the ij-th coordinate is $\frac{1}{k^{1/p}} d(v, S_{ij})$, where $d(v, S)$ is the distance between v and the set S.

For simplicity we'll consider the case $p = 1$ first. Since for any set S we have $d_{uv} \geq |d(u,S) - d(v,S)|$, the embedded uv-distance d'_{uv} is upper-bounded by $O(d_{uv} \log \frac{1}{\beta})$. The hard part is the lower bound: $d'_{uv} = \Omega(d_{uv})$.

Fix a node pair $uv \in E$. Let $d = d_{uv}$. Let $\rho_i = \min(r^-_{uv}(\beta\ 2^i), d/2)$. 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'_{uv} 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:

$$d'_{uv} = \sum \Omega(\rho_{i+1} - \rho_i) = \Omega(\rho_{i_0} - \rho_0) = \Omega(d).$$

Then by Markov inequality with failure probability $O(\delta)$ this holds for an ϵ-set of node pairs. To make this happen for a *strong* ϵ-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 $\gamma = 2^i\beta$. Wlog assume 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'_{uv} 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 B.2 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 at most $ck/2$ of S_{ij}'s do not have this property is at most $e^{-ck/8} \leq \beta\delta/\log \frac{1}{\beta}$. This proves the claim, and completes the proof of the theorem for the case $p = 1$.

To extend the theorem to general p, follow [26]. Let d^p_{uv} be the embedded uv-distance, 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 $d^p_{uv} = (\frac{1}{k} \sum_{ij} x^p_{ij})^{1/p}$, so

$$d^p_{uv} = x^{1/p} \left(\frac{1}{xk} \sum_{ij} x^p_{ij} \right)^{1/p} \geq x^{1/p} \left(\frac{1}{xk} \sum_{ij} x_{ij} \right) = x^{1/p-1} d^1_{uv} = x^{1/p-1} \Omega(d).$$

For a lower bound, recall that $x_{ij} \leq d$, so $d^p_{uv} \leq \left(\frac{1}{k} \sum_{ij} d^p \right)^{1/p} = x^{1/p} d$. Therefore the (two-sided) distortion is at most x, as required. $\square$

We use the following lemma in the proof of Theorem B.1. The proof is implicit in [26] but we include it for the sake of completeness.

Lemma B.2 *There is a constant $c > 0$ with the following property. Consider 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 distribution. Then with probability at least c, S hits E and misses E'.*

Proof: Let $p = \Pr[E]$ and $p' = \Pr[E']$. 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 $\frac{p}{1-p'}$. Wlog rearrange the order of events: first for each point we choose whether it misses E', so that

$$\Pr[\text{all points miss } E'] = (1 - p')^{1/\gamma} \approx e^{-p'/\gamma} \geq e^{-1/2}.$$

Then upon success choose whether each point hits E. Then at least one point hits E with probability at least $1 - (1 - p)^{1/\gamma} \geq 1 - \frac{1}{\epsilon}$. So the total success probability is at least $c = (1 - \frac{1}{\epsilon}) e^{-1/2}$. $\square$

Appendix C: Graceful Degradation Embedding for Strongly Doubling Metrics

The exact statement and the proof of this theorem extends the discussion in Section 5.

Theorem C.1 *(a) Bourgain's algorithm embeds any strongly 2^s-doubling metric into L_1 with dimesion $O(\log^2 n)$ so that each $d_{uv} \geq r^+_{uv}(2^{-i})$ is embedded with distortion $O(i + s)$, for each $i \in [\log n]$. (b) In particular, for any $\epsilon > 0$ the distortion is $O(s + \log \frac{1}{\epsilon})$ for a strong ϵ-set of node pairs.*

Proof: First note that part (b) easily follows from part (a): given $\epsilon > 0$ apply (a) for $i = \log \frac{1}{\epsilon}$ to get the desired distortion for all pairs uv such that $d_{uv} \geq r^+_{uv}(\epsilon)$, which is clearly a strong ϵ-set.

The rest of the proof is on part (a). Recall that Bourgain's algorithm uses uar sets S_{ij} of size 2^i, for each $i \in [\log n]$ and $j \in [k]$, $k = O(\log n)$. Denote the contribution of the set S_{ij} by $x_{ij} = |d(u, S_{ij}) - d(v, S_{ij})|$. For normalization purposes divide all coordinates by $k^{1/p}$, so that the embedded uv-distance is

$$d^p_{uv} = \left(\frac{1}{k} \sum_{ij} x^p_{ij} \right)^{1/p}.$$

For simplicity consider the case $p = 1$ first.

Fix x and a node pair uv such that $d_{uv} \geq r^+_{uv}(2^{-x})$. Let d and d' be the true and embedded distances, respectively. As in the Bourgain's proof, $d' = \Omega(d)$. Since $x_{ij} \leq d$, the Bourgain's upper bound is $d' = O(dk \log n)$. Here we'll improve it to $O((x+s)dk)$. Specifically, we'll show that

$$\sum_{i>x} \sum_j x_{ij} \leq O(dks).$$

Fix $i > x$. Let $\beta = 2^{1/s}$ and $t = \frac{i-x}{2}$. Let X_{ju} be a 0-1 random variable that is equal to 1 if and only if $d(u, S_{ij}) > d\beta^{-t}$. It is not hard to prove that this happens with probability at most e^{-2^t}.[4] We'd like to upper-bound $\sum_j X_{ju}$ by a constant times the expectation, but for large enough t the expectation is too small to give small enough failure probability via Chernoff bounds. However, if we give up a factor of $2^{2^t}/2^t$, then the Chernoff bound (Lemma C.3 with $l = 2^t$) gives $\sum_j X_{ju} = O(k2^{-t})$ with a sufficiently small failure probability to make sure that this happens for all u simultaneously.

Note that $x_{ij} > 2d\beta^{-t}$ only if $Y_j = 1$, where $Y_j = X_{ju} \vee X_{jv}$. So

$$\sum_j x_{ij} \leq O(d) \sum_j (\beta^{-t} + Y_j) = O(dk)(\beta^{-t} + 2^{-t}).$$

Summing this over all $i > x$ we obtain the desired upper bound since $\frac{1}{1-1/\beta} = O(s)$.

To extend this theorem to a general $p \leq 1$ we need a more complicated calculation than the one in [26]. As before, consider a fixed $i > x$. Let S be the set of all j such that $Y_j = 1$. Recall that with high probability it is the case that for all pairs uv the size of S is at most $O(k2^{-t})$. Therefore

$$\begin{aligned}
\sum_j x_{ij}^p &= \sum_{j \in S} x_{ij}^p + \sum_{i \notin S} x_{ij}^p \leq |S| d^p + k(2d\beta^{-t})^{1/p} \\
&= O(k)(2d)^p(2^{-t} + \beta^{-tp}) \\
\frac{1}{k}\sum_{i>x}\sum_j x_{ij}^p &\leq (2d)^p \sum_{i>x} O(\beta^{-ip} + 2^{-i}) \leq O\left(\frac{(2d)^p}{1-\beta^{-p}}\right) \\
&\leq (2d)^p O(s/p) \\
d_{uv}^p &= \left(\frac{1}{k}\sum_{i>x}\sum_j x_{ij}^p + \frac{1}{k}\sum_{i \leq x}\sum_j x_{ij}^p\right)^{1/p} \\
&\leq O(d)(x + s/p)^{1/p}
\end{aligned}$$

For a lower bound, note that for $l = x + 2s$ it is the case that $r^+_{uv}(2^{-l}) \leq d/4$. In the proof of Thm. B.1 we essentially show that $\sum_{i \leq l} \sum_j x_{ij} \geq \Omega(kd)$. Therefore,

$$\begin{aligned}
d_{uv}^p &\geq \left(\frac{1}{k}\sum_{i \leq l}\sum_j x_{ij}^p\right)^{1/p} = l^{1/p}\left(\frac{1}{kl}\sum_{i \leq l}\sum_j x_{ij}^p\right)^{1/p} \geq l^{1/p}\left(\frac{1}{kl}\sum_{i \leq l}\sum_j x_{ij}\right) \\
&\geq \Omega(d)(x+s)^{1/p-1}
\end{aligned}$$

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

[4] Indeed, for $i \geq l \geq x$ letting $r = r_u(2^{-l})$ we have $d > r_u(2^{-x}) = r_u(2^{-l}2^{l-x}) \geq \beta^{l-x} r$, so

$$\Pr[d(u, S_{ij}) > d\beta^{x-l}] \leq \Pr[S_{ij} \text{ misses } B_u(r)] = (1 - 2^{-l})^{2^i} < e^{-2^{i-l}}$$

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

We use these lemmas in the proof of the above theorem.

Lemma C.2 *Consider a strongly s-doubling metric d and fixed ϵ and δ. Let $\epsilon' = \epsilon s^{\log \delta}$. Then for a strong ϵ-set of node pairs uv we have $r_u(\epsilon') \leq \delta d_{uv}$. Therefore for an 2ϵ-set of pairs uv we have $r^+_{uv}(\epsilon') \leq \delta d_{uv}$.*

Lemma C.3 *Let X_j, $j \in [8 \log n]$ be independent 0-1 random variables such that $\Pr[X_j = 1] = e^{-l}$ where $l > 16$. Then $\sum X_j < \frac{8}{l} \log n$ with probability at least $1 - n^{-4}$.*
Proof: Let $X = \sum X_j$ and $\mu = E(X)$. Let $1 + \delta = e^l/l$. Then using Chernoff Bounds we get

$$\Pr[X > 8l^{-1} \log n] = \Pr[X > (1+\delta)\mu] < e^{-\mu} \left(\frac{e}{1+\delta} \right)^{(1+\delta)\mu} < \left(\frac{(el)^{1/l}}{e} \right)^{8 \log n} < \frac{1}{n^4}$$

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

Appendix D: Asymptotically-Uniform Metrics

This section extends the discussion in Section 5.

Call a metric (ϵ, δ)-uniform if it is ϵ-embeddable into a uniform metric with distortion $1 + \delta$. Call a family of metrics $\{M_n : n \in \mathbb{N}\}$ *asymptotically uniform* if for each $\epsilon, \delta > 0$ there exists N such that metric M_n is (ϵ, δ)-uniform for each $n \geq N$. We'll demonstrate several families of metrics that are non-embeddable into L_p with constant distortion, but asymptotically uniform, hence ϵ-embeddable. These are: preferential attachment graphs, constant-degree expanders, and hypercubes.

The preferential attachment graph (PA) is an expander [29]. Any *constant-degree* expander is embeddable into L_p with distortion at least $\Omega(\log n)$[26, 27]. PA (and Internet) need not contain a constant-degree expander since their high expansion might rely on the high-degree nodes. However, we can lower-bound the distortion using the average distance.

Lemma D.1 *PA is embeddable into l_p, $p \in [1, 2]$ with distortion no better than $\Omega(\log n)/\log\log n$.*
Proof: Let λ, η be the all-pairs max-concurrent flow and min-ratio cut, respectively, so that $\lambda \leq \eta$. By Linial-London-Rabinovich [26] the minimal distortion for embedding any graph into L_1 is $\gamma \geq \eta/\lambda$. By a simple argument from Leighton-Rao [24], for expanders with $O(n)$ edges $\eta/\lambda \geq \langle d \rangle$ where $\langle d \rangle$ is the average distance in the graph. Let's lower-bound the expected $\langle d \rangle$ for PA.

In Thm. 5 of Bollobas-Riordan [3], they number the vertices of PA from 1 to n, in the order of arrival, and show that for some $L = \Theta(\log n)/\log\log n$, the expected number of uv-paths of length exactly $l < L$ is at most $(n/\sqrt{uv})(\frac{2}{3})^l/(\log n)$. Therefore, for $u, v > n/2$ the expected number of uv-paths of length $< L$ is $O(1/\log n)$, so $d_{uv} \geq L$ w.h.p. for large enough n, so $E(d_{uv}) = \Omega(L)$.

This proves the lemma for $p = 1$. The general case follows since for any $p \in (1, 2]$ there exists a constant-distortion embedding from L_p to L_1 (e.g. by [18]). □

It turns out that the shortest-paths metric of PA is near-uniform with high probability.

Lemma D.2 *PA is asymptotically uniform whp. More precisely, for any fixed $\epsilon, \delta > 0$, PA is (ϵ, δ)-uniform with failure probability $o(1)$.*
Proof: Using the Bollobas-Riordan argument from Lemma D.1, for $u, v > \epsilon n$ the expected number of uv-paths of length $< L$ is $q = O(\frac{1}{\epsilon \log n})$, so $d_{uv} < L$ with probability at most q. The fraction of such pairs is at most q in expectation, hence by Markov inequality it is at most ϵ with probability at most q/ϵ. So with probability at least $1 - q/\epsilon$ all but a $O(\epsilon)$-fraction of node pairs is at distance at least L. This suffices since by [3] the diameter of PA is at most $L(1 + \delta)$ with failure probability at most $o(1)$. □

Constant-degree expanders are near-uniform for *all* n and with a small *additive* distortion.

Lemma D.3 *For any $\epsilon > 0$, any constant-degree expander is ϵ-embeddable into a uniform metric with additive distortion of $O(\log \frac{1}{\epsilon})$.*

Proof: Let $\beta = \alpha/d$ where α is the expansion and d is the maximal degree, and let $s \leq n/2$. Then any ball of radius r and size s has least αs edges coming out of it, which go to at least βs distinct nodes outside of the ball. So the ball of radius $r+1$ has at least $(1+\beta)s$ nodes. Similarly, any ball of radius r and size $n-s$ 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 of radius $r+1$ has at most $(1-\beta)s$ nodes. Therefore for any node u

$$r_u(1-\epsilon) - r_u(\epsilon) \leq \log_{1+\beta}(1/\epsilon) + \log_{1-\beta}(\epsilon) = O(\log \frac{1}{\epsilon}).$$

We obtained the required additive distortion on all node pairs adjacent to the same node. Now let's extend it to entire graph. Ignore node pairs uv such that $v \in B'_u(\epsilon)$ or $v \notin B_u(1-\epsilon)$. Let G be the graph on the remaining node pairs. Then for any pair of edges adjacent in G their distances differ by at most $O(\log \frac{1}{\epsilon})$. It remains to show that in G all but an $O(\epsilon)$-fraction of node pairs are within a constant #hops from each other. This follows from the density of G: 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 G; obviously any two such nodes have a common neighbor in G, claim proved. □

Lemma D.4 *Hypercubes are asymptotically uniform.*

Proof: Fix $\epsilon, \delta > 0$. Let $a = \frac{1}{1+\delta}$, $b = \frac{1}{1+\delta/2}$ and $c = \frac{1}{1-a}$. Note that for each $j < i \leq bk/2$ we have

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

Therefore #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. □