Embeddings of Negative-type Metrics and An Improved Approximation to Generalized Sparsest Cut*

Shuchi Chawla Anupam Gupta Harald Räcke

Abstract

In this paper, we study the metrics of *negative type*, which are metrics (V, d) such that $\sqrt{\mathrm{d}}$ is an Euclidean metric; these metrics are thus also known as "ℓ_2-squared" metrics.

We show how to embed n-point negative-type metrics into Euclidean space ℓ_2 with distortion $D = O(\log^{3/4} n)$. This embedding result, in turn, implies an $O(\log^{3/4} k)$-approximation algorithm for the Sparsest Cut problem with non-uniform demands. Another corollary we obtain is that n-point subsets of ℓ_1 embed into ℓ_2 with distortion $O(\log^{3/4} n)$.

1 Introduction

The area of finite metric spaces and their embeddings into "simpler" spaces lies in the intersection of the areas of mathematical analysis, computer science and discrete geometry. Over the past decade, this area has seen hectic activity, partly due to the fact that it has proved invaluable in many algorithmic applications. Many details can be found in the surveys by Indyk [Ind01] and Linial [Lin02], or in the chapter by Matoušek [Mat02].

One of the first major applications of metric embeddings in Computer Science was an $O(\log k)$ approximation to the Sparsest Cut problem with non-uniform demands (henceforth called the *Generalized Sparsest Cut* problem) [LLR95, AR98]. This result was based on a fundamental theorem of Bourgain [Bou85] in the local theory of Banach spaces, which showed that any finite n-point metric could be embedded into ℓ_1 space (and indeed, into any of the ℓ_p spaces) with distortion $O(\log n)$. The connection between these results uses the fact that the Generalized Sparsest Cut problem seeks to minimize a linear function over all cuts of the graph, which is equivalent to optimizing over all n-point ℓ_1 metrics. Since this problem is NP-hard, we can optimize over all n-point metrics instead, and then use an algorithmic version of Bourgain's embedding to embed into ℓ_1 with only an $O(\log n)$ loss in performance.

A natural extension of this idea is to optimize over a smaller class of metrics that contains ℓ_1; a natural candidate for this class is NEG, the class of n-point *metrics of negative type*[1]. These are just the metrics obtained by squaring an Euclidean metric, and hence are often called "ℓ_2-squared" metrics. It is known that the following relationships hold:

$$\ell_2 \text{ metrics } \subseteq \ell_1 \text{ metrics } \subseteq \mathsf{NEG} \text{ metrics.} \tag{1.1}$$

Since it is possible to optimize over NEG via semidefinite programming, this gives us a semidefinite relaxation for the Generalized Sparsest Cut problem [Goe97]. Now if we could prove that n-point metrics in NEG embed into ℓ_1 with distortion D, we would get a D-approximation for Sparsest Cut; while this D has been conjectured to be $O(\sqrt{\log n})$ or even $O(1)$, no bounds better than the $O(\log n)$ were known prior to this work.

In a recent breakthrough, Arora, Rao, and Vazirani [ARV04] showed that every n-point metric in NEG has a contracting embedding into ℓ_1 such that the sum of the distances decreases by only $O(\sqrt{\log n})$. Formally, they showed that the SDP relaxation had an integrality gap of $O(\sqrt{\log n})$ for the case of uniform demand Sparsest Cut; however, this is equivalent to the above statement by the results of Rabinovich [Rab03].

We extend their techniques to give embeddings for n-point metrics in NEG into ℓ_2 with distortion $O(\log^{3/4} n)$. More generally, we obtain the following theorem.

Theorem 1.1. *Given (V, d), a negative-type metric, and a set of terminal-pairs $\mathbb{D} \subseteq V \times V$ with $|\mathbb{D}| = k$, there is a contracting embedding $\varphi : V \to \ell_2$ such that for all pairs $(x, y) \in \mathbb{D}$,*

$$\|\varphi(x) - \varphi(y)\|_2 \geq \frac{1}{O(\log^{3/4} k)}\, \mathrm{d}(x, y).$$

*Computer Science Department, Carnegie Mellon University, Pittsburgh PA 15213. {shuchi, anupamg, harry}@cs.cmu.edu.
The first and third authors were supported in part by NSF grant no. CCR-0122581 (The ALADDIN project).

[1] Note that NEG usually refers to all *distances* of negative-type, even those that do not obey the triangle inequality. In this paper, we will use NEG only to refer to negative-type *metrics*.

Note that the above theorem requires the embedding to be contracting for all node pairs, but the resulting contraction needs to be small only for the terminal pairs. In general, when $\mathbb{D} = V \times V$, the embedding is an $O(\log^{3/4} n)$-distortion embedding into ℓ_2. Though we also give a randomized polynomial-time algorithm to find this embedding, let us point out that optimal embeddings into ℓ_2 can be found using semidefinite programming [LLR95, Thm. 3.2(2)].

Finally, let us note some simple corollaries.

THEOREM 1.2. *Let $D(n) = O(\log^{3/4} n)$. Every n-point metric in* NEG *embeds into ℓ_1 with $D(n)$ distortion, and every n-point metric in ℓ_1 embeds into Euclidean space ℓ_2 with $D(n)$ distortion. These embeddings can be found in polynomial time.*

The existence of both embeddings follows immediately from (1.1). To find the map $\mathsf{NEG} \to \ell_1$ in polynomial time, we can use the fact that every ℓ_2 metric can be embedded into ℓ_1 isometrically; if we so prefer, we can find a distortion-$\sqrt{3}$ embedding into ℓ_1 in deterministic polynomial time using families of 4-wise independent random variables [LLR95, Lemma 3.3].

THEOREM 1.3. *There is a randomized polynomial-time $O(\log^{3/4} k)$-approximation algorithm for the Sparsest Cut problem with non-uniform demands.*

Theorem 1.3 thus extends the results of Arora et al. [ARV04] to the case of non-uniform demands, albeit it proves a weaker result than the $O(\sqrt{\log k})$ that they achieve for uniform demands.

The proof of Theorem 1.3 follows from the fact that distortion-D embeddings of negative-type metrics into ℓ_1 show the integrality gap of the semidefinite programming relaxation of the Sparsest Cut problem is at most D, and furthermore the embedding can be used to find such a cut as well. (For more details about this connection of embeddings to the Sparsest Cut problem, see the survey by Shmoys [Shm97, Sec. 5.3]; the semidefinite programming relaxation can be found in the survey by Goemans [Goe97, Sec. 6]).

1.1 Our Techniques. The proof of the Main Theorem 1.1 proceeds thus: we first classify the terminal pairs in $\mathbb{D}$ by distance scales. We define the scale-i set $\mathbb{D}_i$ to be the set of all pairs $(x, y) \in \mathbb{D}$ with $\mathrm{d}(x, y) \approx 2^i$. For each scale i, we find a partition of V into components such that a constant fraction of the terminal pairs $(x, y) \in \mathbb{D}_i$, the following two "good" events happens: (1) x and y lie in different components of the partition, and (2) the distance from x to any other component is at least $\eta\, 2^i$, and the same for y. Here $\eta = 1/O(\sqrt{\log k})$. Informally, both x and y lie deep within their *distinct* components, and this happens for a constant fraction of the pairs $(x, y) \in \mathbb{D}_i$. (The details of this process are given in Section 3; the proofs use ideas from [ARV04] and the subsequent improvements in [Lee05].)

Note that the good event happens for only a constant fraction of the pairs in $\mathbb{D}_i$, and we have little control over which of the pairs will be the lucky ones. However, to get the embedding, we want a partitioning scheme that separates a random constant fraction of the pairs in $\mathbb{D}_i$. To this end, we employ a simple reweighting scheme (reminiscent of the Weighted Majority algorithm [LW94] and many other applications). We just duplicate each unlucky pair and repeat the above process $O(\log k)$ times. Since each pair that is unlucky gets a higher weight in the subsequent runs, a simple argument given in Section 4 shows that each pair in $\mathbb{D}_i$ will be separated in at least $\log k$ of these $O(\log k)$ partitions. (Picking one of these partitions uniformly at random would now ensure that each vertex is separated with constant probability.)

We could now use these partitions naïvely to construct an embedding where the contraction for the pairs in $\mathbb{D}$ would be bounded by $\sqrt{\log k}/\eta$. However, this would be no better than the previous bounds, and hence we have to be more careful. We slightly adapt the *measured descent* embeddings of Krauthgamer et. al. [KLMN04] to combine the $O(\log k)$ partitions for the various distance scales to get a distortion-$O(\sqrt{\log k/\eta}) = O(\log^{3/4} k)$ embedding. The details of the embedding are given in Section 5.

1.2 Related Work. This work adopts and adapts techniques of Arora, Rao and Vazirani [ARV04], who gave an $O(\sqrt{\log n})$-approximation for the uniform demand case of Sparsest Cut. In fact, using their results about the behavior of projections of negative-type metrics almost as a black-box, we obtain an $O(\log^{5/6} n)$-approximation for Generalized Sparsest Cut. Our approximation factor is further improved to $O((\log n)^{3/4})$ by results of Lee [Lee05] showing that the hyperplane separator algorithm from [ARV04, Section 3] itself gives an $O(\sqrt{\log n})$ approximation for the uniform demand case.

As mentioned above, there has been a large body of work on low-distortion embeddings of finite metrics; see, e.g., [Bar98, Bou85, CGN$^+$03, FRT03, GKL03, GNRS04, KLMN04, LLR95, Mat96, Mat99, Rao99], and our work stems in spirit from many of these papers. However, it draws most directly on the technique of *measured descent* developed by Krauthgamer et. al. [KLMN04].

Independently of our work, Lee [Lee05] has used so-called "scale-based" embeddings to give low-distortion embeddings from ℓ_p $(1 < p < 2)$ into ℓ_2. The paper gives a "Gluing Lemma" of the following form: if for every distance scale i, if we are given a contracting embedding ϕ_i such that each pair x, y with $\mathrm{d}(x, y) \in [2^i, 2^{i+1})$ has $\|\phi_i(x) - \phi_i(y)\| \geq \frac{\mathrm{d}(x,y)}{K}$, one can glue them together to get an embedding $\phi : \mathrm{d} \to \ell_2$ with distortion $O(\sqrt{K \log n})$. His result is a generalization of [KLMN04], and of our Lemma 5.1; using this gluing lemma, one can derive an ℓ_2 embedding from the decomposition bundles of Theorem 4.1 without using any of the ideas in Section 5.

2 Notation and Definitions

2.1 Sparsest Cut. In the *Generalized Sparsest Cut* problem, we are given an undirected graph $G = (V, E)$ with edge capacities c_e, and k source-sink (terminal) pairs $\{s_i, t_i\}$ with each pair having an associated demand D_i. For any subset $S \subseteq V$ of the nodes of the network, let $D(S, \bar{S})$ be the net demand going from the terminals in S to those outside S, and $C(S, \bar{S})$ the total capacity of edges exiting S. Now the *generalized sparsest cut* is defined to be

$$\Phi = \min_{S \subseteq V} \frac{C(S, \bar{S})}{D(S, \bar{S})} = \min_{\text{cut metrics } \delta_S} \frac{\sum c_{uv}\ \delta_S(u, v)}{\sum_{ij} D_i\ \delta_S(s_i, t_i)}$$
$$= \min_{\mathrm{d} \in \ell_1} \frac{\sum_{(u,v) \in E} c_{uv}\ \mathrm{d}(u, v)}{\sum_{ij} D_i\ \mathrm{d}(s_i, t_i)} \qquad (2.2)$$

(If there is unit demand between all pairs of vertices, then the problem is just called the Sparsest Cut problem.) The problem is NP-hard [SM90], as is optimizing over the cone of ℓ_1-metrics [Kar85]. There is much work on this fundamental problem (see, e.g., [LR99, Shm97]), and $O(\log k)$ approximations were previously known [LLR95, AR98]. These algorithms proceeded by relaxing the problem and optimizing over *all* metrics instead of over ℓ_1-metrics in (2.2), and then rounding the solution. A potentially stronger relaxation is obtained by optimizing only over metrics $\mathrm{d} \in \mathsf{NEG}$ instead of over all metrics:

$$\Phi_{\mathsf{NEG}} = \min_{\mathrm{d} \in \mathsf{NEG}} \frac{\sum_{(u,v) \in E} c_{uv}\ \mathrm{d}(u, v)}{\sum_i D_i\ \mathrm{d}(s_i, t_i)} \qquad (2.3)$$

This quantity is the semidefinite relaxation of the problem, and can be approximated well in polynomial time (see, e.g., [Goe97]). Since $\ell_1 \subseteq \mathsf{NEG}$, it follows that $\Phi_{\mathsf{NEG}} \leq \Phi$. Now if we can embed n-point metrics in NEG into ℓ_1 with distortion at most D, one can show that $\Phi \leq D \times \Phi_{\mathsf{NEG}}$, and also get a D approximation for Generalized Sparsest Cut.

2.2 Metrics. The input to our embedding procedure will be a *negative-type metric* (V, d) with $|V| = n$. We can, and indeed will use the following standard correspondence between finite metrics and graphs: we set V to the node set of the graph $G = (V, E = V \times V)$, where the length of an edge (x, y) is set to $\mathrm{d}(x, y)$. This correspondence allows us to perform operations like deleting edges to partition the graph. By scaling, we can assume that the smallest distance in (V, d) is 1, and the maximum distance is some value $\Phi(\mathrm{d})$, the *diameter* of the graph.

It is well-known that any *negative-type* distance space admits a geometric representation as the *square of a Euclidean metric*; i.e., there is a map $\psi : V \to \mathbb{R}^n$ such that $\|\psi(x) - \psi(y)\|_2^2 = \mathrm{d}(x, y)$ for every $x, y \in V$ [DL97, Thm. 6.2.2]. Furthermore, the fact that d is a metric implies that the angle subtended by any two points at a third point is non-obtuse. Since this map can be found in polynomial time using semidefinite programming, we will assume that we are also given such a map ψ. For any node $x \in V$, we use $\vec{x}$ to denote the point $\psi(x) \in \mathbb{R}^n$.

2.3 Terminal Pairs. We are also given a set of *terminal pairs* $\mathbb{D} \subseteq V \times V$; these are the pairs of nodes for which we need to ensure a small contraction. In the sequel, we will assume that each node in V takes part in *at most one* terminal-pair in $\mathbb{D}$. This is without loss of generality; if a node x belongs to several terminal pairs, we add new vertices x_i to the graph at distance 0 from x, and replace x in the i-th terminal pair with x_i. (Since this transformation adds at most $O(|\mathbb{D}|)$ nodes, it does not asymptotically affect our results.) Note that a result of this is that $\mathbb{D}$ may have two terminal pairs (x, y) and (x', y') such that $\mathrm{d}(x, x') = \mathrm{d}(y, y') = 0$.

A node $x \in V$ is a *terminal* if there is a (unique) y such that $(x, y) \in \mathbb{D}$; call this node y the *partner* of x. Define $\mathbb{D}_i$ to be the set of node-pairs whose distance according to d is approximately 2^i.

$$\mathbb{D}_i = \{(x, y) \in \mathbb{D} \mid 2^i \leq \mathrm{d}(x, y) < 2^{i+1}\} \qquad (2.4)$$

We use the phrase *scale-i* to denote the distances in the interval $[2^i, 2^{i+1})$, and hence $\mathbb{D}_i$ is merely the set of terminal pairs that are at distance scale i. If $(x, y) \in \mathbb{D}_i$, then x and y are called *scale-i terminals*. Let $\mathcal{D}$ be the set of all terminal nodes, and $\mathcal{D}_i$ be the set of scale-i terminals.

The radius r *ball* around $x \in V$ is naturally defined to be $\mathbf{B}(x, r) = \{z \in V \mid \mathrm{d}(x, z) \leq r\}$. Given a set $S \subseteq V$, the ball $\mathbf{B}(S, r) = \cup_{x \in S} \mathbf{B}(x, r)$.

2.4 Metric Decompositions: Suites and Bundles. Much of the paper will deal with finding decompositions of metrics (and of the underlying graph) with specific properties; let us define these here. Given a distance scale i and a partition P_i of the graph, let $C_i(v)$ denote the component containing a vertex $v \in V$. We say that a pair $(x, y) \in \mathbb{D}_i$ is *δ-separated* by the partition P_i if

- the vertices x and y lie in different components; i.e., $C_i(x) \neq C_i(y)$, and
- both x and y are "far from the boundary of their components", i.e., $\mathrm{d}(x, V \setminus C_i(x)) \geq \delta\, \mathrm{d}(x, y)$ and $\mathrm{d}(y, V \setminus C_i(y)) \geq \delta\, \mathrm{d}(x, y)$.

A *decomposition suite* Π is a collection $\{P_i\}$ of partitions, one for each distance scale i between 1 and $\lfloor \log \Phi(\mathrm{d}) \rfloor$. Given a *separation function* $\delta(x, y) : V \times V \to [0, 1]$, the decomposition suite Π is said to *$\delta(x, y)$-separate* $(x, y) \in \mathbb{D}$ if for the distance scale i such that $(x, y) \in \mathbb{D}_i$, (x, y) is $\delta(x, y)$-separated by the corresponding partition $P_i \in \Pi$.

Finally, a *$\delta(x, y)$-decomposition bundle* is a collection $\{\Pi_j\}$ of decomposition suites such that for each $(x, y) \in \mathbb{D}$, at least a constant fraction of the Π_j $\delta(x, y)$-separate the pair (x, y).

In Section 2.4, we show how to create a decomposition suite that $\Omega(1/\sqrt{\log k})$-separates a constant fraction of the pairs $(x, y) \in \mathbb{D}_i$, for all distance scales i. Using this procedure and a simple reweighting argument, we construct a $\Omega(1/\sqrt{\log k})$-decomposition bundle with $O(\log k)$ suites. Finally, in Section 5, we show how decomposition bundles give us embeddings of the metric d into ℓ_2.

3 Creating Decomposition Suites

3.1 The Procedure Project-&-Prune**.** In this section, we will give the procedure Project-&-Prune that takes a distance scale i, and constructs a partition of V that η-separates at least a constant fraction of the pairs in $\mathbb{D}_i$. Here the parameter $\eta = \frac{1}{4c\sqrt{\log k}}$, where c is a constant to be defined later; let us also define $f = \frac{1}{4\eta} = c\sqrt{\log k}$.

Procedure Project-&-Prune**:**

Input: The metric (V, d), and its geometric representation where $x \in V$ is mapped to $\vec{x}$. Also, a distance scale i.

1. **Project.** In this step, we pick a random direction and project the points in V on the line in this direction. Formally, we pick a random unit vector u. Let $p_x = \sqrt{n}\, \langle \vec{x}, u \rangle$ be the *normalized projection* of the point $\vec{x}$ on u.

2. **Bucket.** Let $\ell = 2^{i/2}$, and set $\beta = \ell/6$. Informally, we will form *buckets* by dividing the line into intervals of length β. We then group the terminals in $\mathcal{D}_i$ according to which interval (mod 4) they lie in. (See Figure 3.1.) Formally, for each $a = 0, 1, 2, 3$, define

$$A_a = \{\ x \in \mathcal{D}_i \mid p_x \in \cup_{m \in \mathbb{Z}} \big((4m+a)\beta, (4m+1+a)\beta\big)\ \}$$

A terminal pair $(x, y) \in \mathbb{D}_i$ is *split* by A_a if $x \in A_a$ and $y \in A_{(a+2) \bmod 4}$. If the pair (x, y) is not split by any A_a, we remove both x and y from the sets A_a. For $a \in \{0, 1\}$, let $\mathbb{B}_a \subseteq \mathbb{D}_i$ be the set of terminal pairs split by A_a or A_{a+2}.

3. **Prune.** If there exist terminals $x \in A_a$ and $y \in A_{(a+2) \bmod 4}$ for some $a \in \{0, 1\}$ (not necessarily belonging to the same terminal pair) with $\mathrm{d}(x, y) < \ell^2/f$, we remove x and y and their partners from the sets $\{A_a\}$.

4. **Cleanup.** For each a, if $(x, y) \in \mathbb{B}_a$ and the above pruning step has removed either of x or y, then we remove the other one as well, and remove (x, y) from $\mathbb{B}_a$. Once this is done, $A_a \times A_{(a+2) \bmod 4} \subseteq \mathbb{D}_i$, and $\mathbb{B}_a$ is once again the set of terminal pairs split by A_a or A_{a+2}.

5. If $\max\{|\mathbb{B}_0|, |\mathbb{B}_1|\} \leq \frac{1}{64}|\mathbb{D}_i|$, go back to Step 1, else go to Step 6.

6. Say the set $\mathbb{B}_a$ has more pairs than $\mathbb{B}_{(1-a) \bmod 2}$. Define the partition P_i by deleting all the edges at distance $\ell^2/2f$ from the set A_a. (This step can be thought of as taking $C = \mathbf{B}(A_a, \ell^2/2f)$, and defining the partition P_i to be $G[C]$ and $G[V \setminus C]$, the components induced by C and $V \setminus C$.)

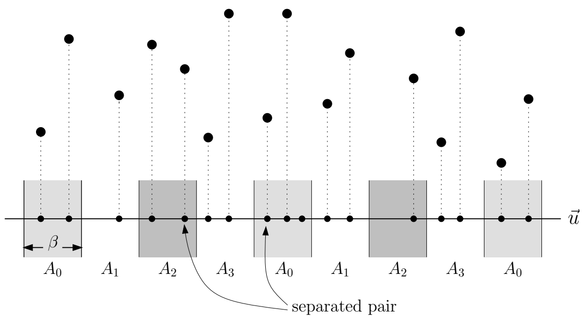

Figure 3.1: Projection and Bucketing

Note the procedure above ensures that for *any pair of terminals* $(x, y) \in A_a \times A_{(a+2) \bmod 4}$, the distance $\mathrm{d}(x, y)$ is at least $\ell^2/f = 2^i/f$, even if $(x, y) \notin \mathbb{D}_i$. Why do we care about these pairs? It is because the separation of ℓ^2/f between the sets A_a and $A_{(a+2) \bmod 4}$ ensures that the balls of radius $\frac{\ell^2}{2f}$ around these sets are disjoint.

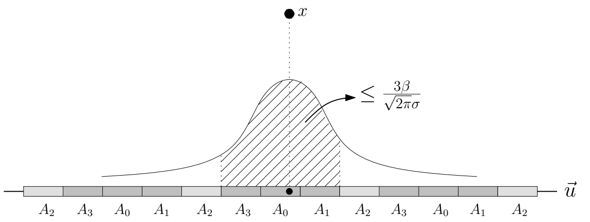

Figure 3.2: The distribution of projected edge lengths in the proof of Lemma 3.1. If y falls into a light-shaded interval, the pair (x, y) is split.

This in turn implies that terminal pairs $(x, y) \in \mathbb{D}_i \cap (A_a \times A_{(a+2) \bmod 4})$ are η-separated upon deleting the edges in Step 6. Indeed, for such a pair (x, y), the components $C_i(x)$ and $C_i(y)$, obtained upon deleting the edges at distance $\frac{\ell^2}{2f}$ from the set A_a, are distinct, and both $\mathrm{d}(x, V \setminus C_i(x))$ and $\mathrm{d}(y, V \setminus C_i(y))$ are at least $\frac{\ell^2}{2f} \geq \frac{d(x,y)}{4f}$.

The following theorem now shows that the procedure Project-&-Prune indeed terminates quickly.

THEOREM 3.1. *For any distance scale i, the procedure* Project-&-Prune *terminates in a constant number of iterations. This gives us a random polynomial-time algorithm that outputs a partition P_i which η-separates at least $\frac{1}{64}|\mathbb{D}_i|$ pairs of $\mathbb{D}_i$.*

The proof of this theorem has two parts, which we will prove in the next two subsections. We first show that the sets $\mathbb{B}_0$ and $\mathbb{B}_1$ contain most of $\mathbb{D}_i$ before the pruning step (with a high probability over the random direction u). We then show that the pruning procedure removes only a constant fraction of the pairs from these sets $\mathbb{B}_0$ and $\mathbb{B}_1$ with a constant probability, and that the size of $\mathbb{B}_0 \cup \mathbb{B}_1$ is at least $|D_i|/32$. It follows that one of these sets must have half of the terminal pairs, proving the theorem.

3.1.1 The Projection Step

LEMMA 3.1. *Fix a distance scale i. At the end of the bucketing stage, the set $\mathbb{B}_0 \cup \mathbb{B}_1$ contains at least $\frac{1}{16}|\mathbb{D}_i|$ terminal pairs w.p. $\frac{1}{15}$.*

Proof. Recall that a terminal pair $(x, y) \in \mathbb{D}_i$ is split if x lies in the set A_a and y lies in $A_{(a+2) \bmod 4}$ for some $a \in \{0, 1, 2, 3\}$. Also, we defined $\ell^2 = 2^i$, and hence $(x, y) \in \mathbb{D}_i$ implies that $\|\vec{x} - \vec{y}\| = \mathrm{d}(x, y) \in [\ell^2, 2\,\ell^2)$. Consider the normalized projections p_x and p_y of the vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ on the random direction u, and note that $p_y - p_x$ is distributed as a Gaussian random variable $Z_u \sim N(0, \sigma^2)$ with a standard deviation $\sigma \in [\ell, \sqrt{2}\ell)$ (see Figure 3.2.)

Now consider the bucket of width β in which p_x lies. The pair (x, y) will not be separated if p_y lies in either the same bucket, or in either of the adjoining buckets. (The probability of each of these three events is at most $\frac{1}{\sqrt{2\pi}\,\sigma} \times \beta$.) Also, at least $\frac{1}{4}$ of the remainder of the distribution causes (x, y) to be split, since each good interval is followed by three bad intervals with less measure.

Putting this together gives us that the probability of (x, y) being split is at least

$$\frac{1}{4}\big(1 - 3\beta\,\frac{1}{\sqrt{2\pi}\,\sigma}\big) \geq \frac{1}{4}\big(1 - \frac{3(\ell/6)}{\sqrt{2\pi}\,\ell}\big) \geq \frac{1}{8}$$

Since each pair $(x, y) \in \mathbb{D}_i$ is separated with probability $1/8$, the linearity of expectations and Markov's inequality implies that at least one-sixteenth of $\mathbb{D}_i$ must be split at the end of the bucketing stage with probability $\frac{1}{15}$. $\square$

3.1.2 The Pruning Step.

We now show that a constant fraction of the terminal pairs in $\mathbb{D}_i$ also survive the pruning phase. This is proved by contradiction, and follows the lines of the argument of Arora et al. [ARV04].

Assume that, with a large probability (over the choice of the random direction u), a large fraction of the terminal pairs in $\mathbb{D}_i$ (say $\frac{1}{64}|\mathbb{D}_i|$) get removed in the pruning phase. By the definition of the pruning step, the projection of $\vec{x} - \vec{y}$ on u must have been large for such removed pair (x, y). In our algorithm, this happens when $\mathrm{d}(x, y) < \ell^2/f$, or equivalently when $\beta > \sqrt{\mathrm{d}(x, y)} \times \sqrt{f}/6$. Since p_x and p_y are separated by at least one bucket of width β, and the expected value of $|p_x - p_y|$ is exactly $\sqrt{\mathrm{d}(x, y)}$, this implies that the expectation is exceeded by a factor of at least $\sqrt{f}/6 = \Omega(\log^{1/4} k)$. Setting $t = \sqrt{f}/6$, we can say that such a pair (x, y) is "stretched by a factor t in the direction u". For any given direction u, the stretched pairs removed in the pruning step are disjoint, and hence form a matching M_u.

Arora et al. showed the following geometric property—for a given set W and some constant C, the number of disjoint t-stretched pairs in $W \times W$ cannot be more than $C|W|$ with constant probability (over the choice of u); however, their proof only proved this for stretch $t = \Omega(\log^{1/3}|W|)$. The dependence on t was improved subsequently by Lee [Lee05] to $t = \Omega(\log^{1/4}|W|)$.

In order to make the above discussion more precise, let us first recall the definition of a stretched set of points.

DEFINITION 3.1. ([ARV04], DEFN. 4) *A set of n points $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ in $\mathbb{R}^n$ is said to be (t, γ, β)-stretched at scale l, if for at least a γ fraction of the directions u, there is a partial matching $M_u = \{(x_i, y_i)\}_i$,* *with $|M_u| \geq \beta n$, such that for all $(x,y) \in M_u$, $\mathrm{d}(x,y) \leq l^2$ and $\langle u, \vec{x}-\vec{y}\rangle \geq tl/\sqrt{n}$. That is, the pair (x,y) is* stretched by a factor of t in direction u.

Theorem 3.2. ([ARV04], Thm. 5) *For any $\gamma, \beta > 0$, there is a $C = C(\gamma,\beta)$ such that if $t > C \log^{1/3} n$, then no set of n points in $\mathbb{R}^n$ can be (t,γ,β)-stretched for any scale l.*

The above theorem has been subsequently improved by Lee, and follows from [Lee05, Thm. 4.1].

Theorem 3.3. *For any $\gamma, \beta > 0$, there is a $C = C(\gamma,\beta)$ such that if $t > C\log^{1/4} n$, then no set of n points in $\mathbb{R}^n$ can be (t,γ,β)-stretched for any scale l.*

Summarizing the implication of Theorem 3.3 in our setting, we get the following corollary.

Corollary 3.1. *Let W be a set of vectors corresponding to some subset of terminals satisfying the following property: with probability $\Theta(1)$ over the choice of a random unit vector u, there exist subsets $S_u, T_u \subseteq W$ and a constant ρ such that $|S_u| \geq \rho|W|$ and $|T_u| \geq \rho|W|$, and the length of the projection $|\langle u, \vec{x}-\vec{y}\rangle| \geq \ell/(6\sqrt{n})$ for all $\vec{x} \in S_u$ and $\vec{y} \in T_u$. Then with probability $\Theta(1)$ over the choice of u, the pruning procedure applied to sets S_u and T_u returns sets S'_u and T'_u with $|S'_u| \geq \frac{3}{4}|S_u|$ and $|T'_u| \geq \frac{3}{4}|T_u|$, such that for all $\vec{x} \in S'_u$ and $\vec{y} \in T'_u$, $\mathrm{d}(x,y) \geq \ell^2/f$.*

Proof. For a unit vector u, let $M(u)$ denote the matching obtained by taking the pairs (x,y) of terminals that are deleted by the pruning procedure when given the vector u. Note that pairs $(x,y) \in M(u)$ have the property that $\mathrm{d}(x,y) < \ell^2/f$ and $|p_x - p_y| > \ell/6$. For the sake of contradiction, suppose there is a constant γ such that the matchings $M(u)$ are larger than $\rho/4|W|$ with probability at least $1-\gamma$ over the choice of u.

Using Definition 3.1 above, we get that the vectors in W form an $(6\sqrt{f}, \gamma, \rho/4)$-stretched set at scale $\ell/\sqrt{f}$. Theorem 3.3 now implies that $6\sqrt{f} = 6\sqrt{c}(\log k)^{1/4}$ must be at most $C\log^{1/4}|W|$. However, since $|W| \leq 2k$, setting the parameter c suitably large compared to C would give us the contradiction. □

Finally, we are in a position to prove Theorem 3.1 using Lemma 3.1 and Corollary 3.1.

Proof. [Proof of Theorem 3.1] Define W to be $\mathcal{D}_i$, the set of all terminals that belong to some terminal pair in $\mathbb{D}_i$. Let a be the index corresponding to the larger of $\mathbb{B}_0$ and $\mathbb{B}_1$ before the pruning step, and set $S_u = A_a$ and $T_u = A_{(a+2) \bmod 4}$ for this value of a. Lemma 3.1 assures us that $|S_u| = |T_u| \geq \frac{1}{32}|\mathbb{D}_i| = \frac{1}{16}|W|$ with probability $\frac{1}{15}$ (over the random choice of the vector $u \in \mathbb{R}^n$). Furthermore, for each $\vec{x} \in S_u$ and $\vec{y} \in T_u$, the fact that $|p_x - p_y| \geq \beta$ translates to the statement that $\langle \vec{x}-\vec{y}, u\rangle \geq \ell/(6\sqrt{n})$.

These vectors satisfy the conditions of Corollary 3.1, and hence we can infer that with a constant probability, the pruning procedure removes at most $\frac{1}{4}|S_u|$ and $\frac{1}{4}|T_u|$ vertices from S_u and T_u respectively. Their partners may be pruned in the cleanup step as well, and hence the total number of terminal pairs pruned is at most $\frac{1}{2}|S_u|$. Thus the number of terminal pairs remaining in $\mathbb{D}_i \cap (S'_u \times T'_u)$ is at least $\frac{1}{2}|S_u| \geq \frac{1}{64}|\mathbb{D}_i|$ pairs.

Since this happens with a constant probability, we will need to repeat Steps 1-3 of the procedure (each time with a new unit vector u) only a constant number of times until we find a partition that η-separates at least $\frac{1}{64}|\mathbb{D}_i|$ of the terminal pairs; this proves the result. □

Running the procedure Project-&-Prune for each distance scale i between 1 and $\lfloor \log \Phi(\mathrm{d}) \rfloor$, we can get the following result with $\gamma = \frac{1}{64}$.

Theorem 3.4. *Given a negative-type metric d, we can find in randomized polynomial time a decomposition suite $\Pi = \{P_i\}$ that η-separates a constant fraction γ of the terminal pairs at each distance scale i.*

In the next section, we will extend this result to get a set of $O(\log k)$ decomposition suites $\{\Pi_j\}$ so that each terminal pair $(x,y) \in \mathbb{D}$ is separated in a constant fraction of the Π_j's.

4 Obtaining Decomposition Bundles: Weighting and Watching

To start off, let us observe that the result in Theorem 3.4 can be generalized to the case where terminal pairs have an associated weight $w_{xy} \in \{0,1,2,\ldots,k\}$.

Lemma 4.1. *Given terminal pairs $(x,y) \in \mathbb{D}$ with weights w_{xy}, there is a randomized polynomial time algorithm that outputs a decomposition suite Π which, for each distance scale i, $\Omega(1/\sqrt{\log k})$-separates terminals with total weight at least $\gamma \sum_{(x,y)\in\mathbb{D}_i} w_{xy}$.*

Proof. The proof is almost immediate: we replace each terminal pair $(x,y) \in \mathbb{D}_i$ having weight $w_{xy} > 0$ with w_{xy} new terminal pairs (x^j, y^j), where the points $\{x^j\}$ and $\{y^j\}$ are placed at distance 0 to x and y respectively. Doing this reduction for all weighted pairs gives us an unweighted instance with a set $\mathbb{D}'_i$ of terminal pairs. Now Theorem 3.4 gives us a decomposition suite η-separating at least $\frac{1}{64}|\mathbb{D}'_i|$ of the new terminal pairs at distance scale i, where

$\eta = 1/O(\sqrt{\log \mathbb{D}_i'}) = 1/O(\sqrt{\log k})$. Finally, observing that the separated terminal pairs at scale i contribute at least $\frac{1}{64} \sum_{(x,y)\in\mathbb{D}_i} w_{xy}$ completes the claim. □

In the sequel, we will associate weights with the terminal pairs in $\mathbb{D}$ and run the procedure from Lemma 4.1 repeatedly. The weights start off at k, and the weight of a pair that is separated in some iteration is halved in the subsequent iteration; this reweighting ensures that all pairs are separated in significantly many rounds. (Note: this weighting argument is fairly standard and has been used, e.g., in geometric algorithms [Cla95], machine learning [LW94], and many other areas; see Welzl [Wel96] for a survey.)

The Algorithm:

1. Initialize $w^{(0)}(x,y) = 2^{\lceil \log k \rceil}$ for all terminal pairs $(x,y) \in \mathbb{D}$. Set $j = 0$.
2. Use the algorithm from Lemma 4.1 to obtain a decomposition suite Π_j. Let T_j be the set of terminal pairs η-separated by this decomposition.
3. For all $(x,y) \in T_j$, set $w^{(j+1)}(x,y) \leftarrow w^{(j)}(x,y)/2$. If $w^{(j+1)}(x,y) < 1$ then $w^{(j+1)}(x,y) \leftarrow 0$.
4. Increment $j \leftarrow j+1$. If $\sum_{(x,y)\in\mathbb{D}_i} w^{(j)}(x,y) \geq 1$ for some i, go to step 2, else halt.

Note that the distance function d in each iteration of the algorithm remains the same.

LEMMA 4.2. *In each iteration j of the above algorithm* $\sum_{(x,y)\in\mathbb{D}_i} w^{(j+1)}(x,y) \leq (1-\frac{\gamma}{2}) \sum_{(x,y)\in\mathbb{D}_i} w^{(j)}(x,y)$.

Proof. In each iteration, the algorithm of Lemma 4.1 separates at least a γ fraction of the weight $\sum_{(x,y)\in\mathbb{D}_i} w^{j}(x,y)$, and hence the total weight in the next round drops by at least half this amount. □

Noting that initially we have $\sum_{(x,y)\in\mathbb{D}_i} w^{(0)}(x,y) \leq k^2$, one derives the following simple corollary:

COROLLARY 4.1. *The above algorithm has at most* $\frac{4}{\gamma}\log k$ *iterations.*

LEMMA 4.3. *Every pair $(x,y) \in \mathbb{D}_i$ is η-separated in at least* $\log k$ *iterations.*

Proof. Since we start off with $w^{(0)}(x,y) = k$ and end with $w^{(j)}(x,y) < 1$, the weight $w^{(j)}(x,y)$ must have been decremented at least $\log k$ times. Each such reduction corresponds to a round j in which (x,y) was η-separated by Π_j. □

THEOREM 4.1. *The above procedure outputs an η-decomposition bundle with at most* $\frac{4}{\gamma}\log k$ *ccdecomposition suites, such that each terminal pair (x,y) is η-separated in at least* $\log k$ *of these suites.*

5 Embedding via decomposition bundles

In the previous sections we have constructed a decomposition bundle with a large separation between terminal pairs. Now, we show how to obtain a small distortion ℓ_2-embedding from this. The proof mainly follows the lines of [KLMN04].

THEOREM 5.1. *Given an $\alpha(x,y)$-decomposition bundle for the metric* d *and a set $\mathbb{D}$, there exists a randomized contracting embedding $\varphi : V \longrightarrow \ell_2$, such that for each pair $(x,y) \in \mathbb{D}$,*

$$||\varphi(x) - \varphi(y)||_2 \geq \Omega\left(\sqrt{\frac{\alpha(x,y)}{\log k}}\right) \cdot \mathrm{d}(x,y)$$

Note that for $\alpha(x,y) = \Omega(1/\sqrt{\log k})$ this theorem implies Theorem 1.1.

Along the lines of [KLMN04], we define a measure of "local expansion". Let

$$V(x,y) = \max\left\{\log\frac{|B(x, 2\mathrm{d}(x,y))|}{|B(x, \mathrm{d}(x,y)/8)|}, \log\frac{|B(y, 2\mathrm{d}(x,y))|}{|B(y, \mathrm{d}(x,y)/8)|}\right\}$$

where $B(x,r)$ denotes the set of *terminal nodes* within the ball of radius r around x. We derive Theorem 5.1 from the following lemma.

LEMMA 5.1. *Given an $\alpha(x,y)$-decomposition bundle, there is a randomized contracting embedding $\varphi : V \longrightarrow \ell_2$ such that for every pair (x,y) with constant probability*

$$||\varphi(x) - \varphi(y)||_2 \geq \Omega\left(\sqrt{\frac{V(x,y)}{\log k}} \cdot \alpha(x,y)\right) \cdot \mathrm{d}(x,y) \ .$$

COROLLARY 5.1. *Given an $\alpha(x,y)$-decomposition bundle, there is a randomized contracting embedding $\varphi : V \longrightarrow \ell_2$ such that for every pair (x,y),*

$$||\varphi(x) - \varphi(y)||_2 \geq \Omega\sqrt{\frac{V(x,y)}{\log k}} \cdot \alpha(x,y) \cdot \mathrm{d}(x,y) \ .$$

Proof. The corollary follows by applying Lemma 5.1 repeatedly and independently for each decomposition suite several times[2]. Then concatenating and rescaling the resulting maps gives with high probability an embedding that fulfills the corollary. □

To see that the above corollary implies Theorem 5.1, we use a decomposition due to [FRT03] (and its extension to general measures, as observed in [LN04] and

[2]Note that this algorithm may give an embedding with a large number of dimensions. However, we can use semidefinite programming to obtain a nearly-optimal embedding of the metric into ℓ_2 in polynomial time.

[KLMN04]) that has the property that with probability at least 1/2, a pair (x, y) is $\Omega(1/V(x, y))$-separated in this decomposition. Applying the corollary to this decomposition bundle, we get an embedding φ_1, such that

$$||\varphi_1(x) - \varphi_1(y)||_2 \geq \Omega\left(\frac{1}{\sqrt{V(x, y) \cdot \log k}}\right) \cdot \mathrm{d}(x, y) \ .$$

Applying the corollary to the decomposition bundle assumed by the theorem gives an embedding φ_2 with

$$||\varphi_2(x) - \varphi_2(y)||_2 \geq \Omega\left(\sqrt{\frac{V(x, y)}{\log k}} \cdot \alpha(x, y)\right) \cdot \mathrm{d}(x, y) \ .$$

Concatenating the two mappings and rescaling, we get a contracting embedding $\varphi = \frac{1}{2}(\varphi_1 \otimes \varphi_2)$, with

$$\begin{aligned} &||\varphi(x) - \varphi(y)||_2 \\ &\geq \Omega\left(\frac{1}{\sqrt{\log k}} \cdot \left(\frac{1}{V(x, y)^{\frac{1}{2}}} + V(x, y)^{\frac{1}{2}} \ \alpha(x, y)\right)\right) \cdot \mathrm{d}(x, y) \\ &\geq \Omega\left(\sqrt{\frac{\alpha(x, y)}{\log k}}\right) \cdot \mathrm{d}(x, y) \end{aligned}$$

as desired.

The embedding. Let $T = \{1, \ldots, \log k\}$ and $Q = \{0, \ldots, m - 1\}$, for some suitably chosen constant m. In the following we define an embedding into $|T| \cdot |Q|$ dimensions. For $t \in T$, let $\mathrm{r}_t(x)$ denote the minimum radius r such that the ball $B(x, r)$ contains at least 2^t terminal nodes. We call $\mathrm{r}_t(x)$ the *t-radius of x*. Further, let $\ell_t(x) \in \mathbb{N}$ denote the distance class this radius belongs to (i.e., $2^{\ell_t(x)-1} \leq \mathrm{r}_t(x) \leq 2^{\ell_t(x)}$).

Fix a decomposition suite $\Pi = \{P_s\}$ from the decomposition bundle. For every distance scale s, we pick a partitioning P_s from the distribution $\mathcal{P}_s$. In the following $\delta(x, y)$ denotes the separation-factor between x and y in this suite, i.e., $\delta(x, y) = \min\{\mathrm{d}(x, V \setminus C_s(x)), \mathrm{d}(y, V \setminus C_s(y))\}$ if $C_s(y) \neq C_s(x)$ and 0, otherwise. Observe that with constant probability we have $\delta(x, y) \geq \alpha(x, y)$.

The standard way to obtain an embedding from a decomposition suite is to create a coordinate for every distance scale and embed points in this coordinate with respect to the partitioning for this scale. For example, one could assign a random color, 0 or 1, to each cluster $C \in P_i$. Let W_i denote the set of nodes contained in clusters with color 0 in partitioning P_i. By setting the i-th coordinate of the image $\varphi(x)$ of a point x to $\mathrm{d}(x, W_0^i)$, a pair (x, y) gets a distance $\Omega(\delta(x, y)\mathrm{d}(x, y))$ with probability 1/2, because this is the probability that the clusters $C_i(x)$ and $C_i(y)$ get different colors (in this case the distance is $\Omega(\delta(x, y)\mathrm{d}(x, y))$ since both nodes are at least that far away from the boundary of their cluster). Overall this approach gives an embedding into ℓ_2 with distortion $O(\sqrt{\log k}/\delta(x, y))$, and has e.g. been used in [Rao99] for getting a $\sqrt{\log n}$ embedding of planar metrics into ℓ_2.

In order to improve this, along the lines of [KLMN04], the goal is to construct an embedding in which the distance between (x, y) increases as the local expansion $V(x, y)$ increases. This can be achieved by constructing a coordinate for every $t \in T$ and then embed points in this coordinate according to the partitioning for the corresponding distance scale $\ell_t(x)$ (i.e., different points use different distance scales depending on their local expansion). Thereby, for a pair with a high $V(x, y)$-value the nodes will often ($\approx V(x, y)$ times) be embedded according to the partitioning for distance scale $i = \lfloor \log \mathrm{d}(x, y) \rfloor$ that corresponds to $\mathrm{d}(x, y)$. Therefore, the pair (x, y) gets a higher distance in the embedding than in the standard approach.

However, transferring the rest of the standard analysis to this new idea has some difficulties. If we define the set W_t as the nodes x that are colored 0 in the partitioning $\ell_t(x)$ we cannot argue that for a pair (x, y) either $\mathrm{d}(x, W_t)$ or $\mathrm{d}(y, W_t)$ is large, because nodes very close to x or y may have distance scales different from $\ell_t(x)$ or $\ell_t(y)$. In order to ensure local consistency such that all nodes close to x obtain their color from the same partitioning, we construct several coordinates for every t, such that for each distance scale $\ell_t(x)$ there is a coordinate in which all nodes close to x derive their color from the partitioning for scale $\ell_t(x)$. The details are as follows.

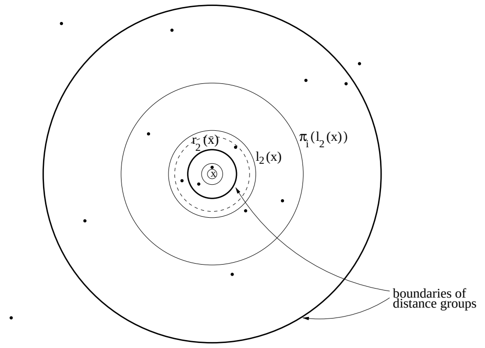

Figure 5.3: A partition of distance scales into distance groups.

For a parameter $q \in Q$ we partition the distance classes into *distance groups* by assigning the ℓ-th class to group $g_q(\ell) := \lceil \frac{\ell - q}{m} \rceil$ (see Figure 5.3). Note that each distance group contains (at most) m consecutive distance classes which means that distances within a group differ at most by a constant factor. All distances in group g are in $\Theta(2^{m\cdot g})$. We define a mapping π_q between distance classes that maps all classes of a group to the medium distance class in this group.

$$\pi_q(\ell) := \begin{cases} 0 & \text{if } q + m\cdot g_q(\ell) - \lfloor \frac{m}{2} \rfloor < 0 \\ p & \text{if } q + m\cdot g_q(\ell) - \lfloor \frac{m}{2} \rfloor > p \\ i + m\cdot g_q(\ell) - \lfloor \frac{m}{2} \rfloor & \text{otherwise} \end{cases}$$

Observe that for each distance class i there is a q such that $\pi_q(i) = i$.

Based on this mapping we define a set W_t^q for each choice of $t \in T$ and $q \in Q$ by $W_t^q = \{x \in V : \mathrm{color}_{\pi_q(\ell_t(x))}(x) = 0\}$, where $\mathrm{color}_i(x)$ denotes the color of the cluster that contains x in partitioning P_i. Note that all nodes whose t-radii fall into the same distance group (w.r.t. parameter q) derive their color (and hence whether they belong to W_t^q) from the same partitioning.

Based on the sets W_t^q we define an embedding $f_{t,q} : V \longrightarrow \mathbb{R}$ for each coordinate (t,q) — $f_{t,q}(x) = \mathrm{d}(x, W_t^q)$. The embedding $f : V \longrightarrow \mathbb{R}^{|T||Q|}$ is defined by $f(x) := \otimes_{t,q} f_{t,q}(x)$. Each coordinate of this embedding is contracting. Therefore, we have for all $x, y \in V$

$$\begin{aligned} ||f(x) - f(y)||_2 &\le \sqrt{|T|\cdot|Q|\cdot \mathrm{d}(x,y)^2} \\ &\le O(\sqrt{\log k})\cdot \mathrm{d}(x,y) \end{aligned}$$

Now, we show that for a pair x, y that is $\delta(x,y)$-separated in the partitioning corresponding to its distance scale, with a constant probability, we get

$$||f(x) - f(y)||_2 \ge \Omega(\delta(x,y)\cdot \mathrm{d}(x,y)) \cdot \sqrt{V(x,y)}$$

This gives Lemma 5.1 since $\delta(x,y) > \alpha(x,y)$ with constant probability.

Fix a pair (x,y) that is $\delta(x,y)$-separated in the partitioning for distance scale $\lfloor \log(\mathrm{d}(x,y)) \rfloor$. Without loss of generality assume that the maximum in the definition of $V(x,y)$ is attained by the first term, i.e., $\frac{|B(x,2\mathrm{d}(x,y))|}{|B(x,\mathrm{d}(x,y)/8)|} \ge \frac{|B(y,2\mathrm{d}(x,y))|}{|B(y,\mathrm{d}(x,y)/8)|}$. We show that for each t with $|B(x,\mathrm{d}(x,y)/8)| \le 2^t \le |B(x,2\mathrm{d}(x,y))|$ there is a $q \in Q$ such that the coordinate (t,q) gives a large contribution, i.e., $|f_{t,q}(x) - f_{t,q}(y)| \ge \Omega(\delta(x,y)\cdot \mathrm{d}(x,y))$.

Fix an integer t with $\log(|B(x,\mathrm{d}(x,y)/8)|) \le 2^t \le \log(|B(x,2\mathrm{d}(x,y))|)$, and let $i = \lfloor \log \mathrm{d}(x,y) \rfloor$ denote the distance class of $\mathrm{d}(x,y)$. Clearly, the distance class $\ell_t(x)$ of the t-radius of x is in $\{i-4,\dots,i+2\}$, because $\mathrm{d}(x,y)/8 \le \mathrm{r}_t(x) \le 2\mathrm{d}(x,y)$. The following claim gives a similar bound on the t-radius for nodes that are close to x.

Claim 5.1. *Let $z \in B(x, \frac{1}{16}\mathrm{d}(x,y))$. Then $\ell_t(z) \in \{i-5, i+3\}$.*

Proof. For the t-radius $\mathrm{r}_t(z)$ around z we have $\mathrm{r}_t(x) - \mathrm{d}(x,y)/16 \le \mathrm{r}_t(z) \le \mathrm{r}_t(x) + \mathrm{d}(x,y)/16$. Since $\mathrm{d}(x,y)/8 \le \mathrm{r}_t(x) \le 2\mathrm{d}(x,y)$ we get $\frac{1}{16}\mathrm{d}(x,y) \le \mathrm{r}_t(z) \le \frac{33}{16}\mathrm{d}(x,y)$, which yields the claim. □

In the following we choose m (the number of distances classes within a group) as 10, and q such that $\pi_q(i) = i$, i.e., i is the median of its distance group. Then the above claim ensures that for all nodes $z \in B(x, \frac{1}{16}\mathrm{d}(x,y))$, the distance class $\ell_t(z)$ is in the same distance group as i. Furthermore, these nodes choose their color (that decides whether they belong to W_t^q) according to the partitioning for distance scale i. Recall that x is $\delta(x,y)$-separated in this partitioning. Therefore, we can make the following claim.

Claim 5.2. *If x does not belong to the set W_t^q then $\mathrm{d}(x, W_t^q) \ge \min\{\frac{1}{16}, \delta(x,y)\}\mathrm{d}(x,y) \ge \frac{1}{16}\delta(x,y)\mathrm{d}(x,y)$.*

Now, we consider the following events concerning the distances of x and y from W_t^q, respectively.

- $X_0 = \{\mathrm{d}(x, W_t^q) = 0\}$, i.e., $x \in W_t^q$
- $X_{\mathrm{far}} = \{\mathrm{d}(x, W_t^q) > \frac{1}{16}\delta(x,y)\mathrm{d}(x,y)\}$
- $Y_{\mathrm{close}} = \{\mathrm{d}(y, W_t^q) \le \frac{1}{32}\delta(x,y)\mathrm{d}(x,y)\}$
- $Y_{\mathrm{far}} = \{\mathrm{d}(y, W_t^q) > \frac{1}{32}\delta(x,y)\mathrm{d}(x,y)\}$

These events only depend on the random colorings chosen for the partitionings in different distance classes. The events X_0 and X_{far} are independent of events Y_{close} and Y_{far}. To see this, note that X_0 and X_{far} only depend on colors chosen for nodes in $B(x, \frac{1}{16}\delta(x,y)\mathrm{d}(x,y))$. Claim 5.1 implies that these colors are derived from the partitioning for distance class i, and all nodes in $B(x, \frac{1}{16}\delta(x,y)\mathrm{d}(x,y))$ get the color assigned to the cluster that contains x.

The events Y_{close} and Y_{far}, however, depend on colors chosen for nodes in $B(y, \delta(x,y)\frac{1}{32}\mathrm{d}(x,y))$. Such a color is either derived from a partitioning for a distance class different from i (in this case independence is immediate), or it is equal to the color assigned to the cluster that contains y in the partitioning for class i. In the latter case the independence follows, since x and y lie in different clusters in this partitioning as they are separated by it.

If $X_0 \cap Y_{\mathrm{far}}$ or $X_{\mathrm{far}} \cap Y_{\mathrm{close}}$ happens, then the dimension (t,q) gives a contribution of $\Omega(\delta(x,y)\mathrm{d}(x,y))$. This happens with probability

$$\begin{aligned} \Pr[X_0 \cap & Y_{\mathrm{far}} \uplus X_{\mathrm{far}} \cap Y_{\mathrm{close}}] \\ &= \Pr[X_0 \cap Y_{\mathrm{far}}] + \Pr[X_{\mathrm{far}} \cap Y_{\mathrm{close}}] \\ &= \Pr[X_0]\cdot\Pr[Y_{\mathrm{far}}] + \Pr[X_{\mathrm{far}}]\cdot\Pr[Y_{\mathrm{close}}] \\ &= \Pr[X_0]\cdot\Pr[Y_{\mathrm{far}}] + \Pr[X_{\mathrm{far}}]\cdot(1 - \Pr[Y_{\mathrm{far}}]) \\ &= 1/2\ . \end{aligned}$$

Here we used the fact that $\Pr[X_0] = \Pr[X_{\text{far}}] = 1/2$ which holds due to Claim 5.2. This completes the proof of Lemma 5.1.

Acknowledgments. We would like to thank Sanjeev Arora, Avrim Blum, Hubert Chan, Vineet Goyal, Satish Rao, R. Ravi, and Mohit Singh for useful discussions, and Jiri Matoušek for pointing us to [Wel96]. Many thanks to James R. Lee for useful conversations, and for sending us a manuscript of [KLMN04].

References

[AR98] Yonatan Aumann and Yuval Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301, 1998.

[ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *Proc. 36th STOC*, pages 222–231, 2004.

[Bar98] Yair Bartal. On approximating arbitrary metrics by tree metrics. In *Proc. 30th STOC*, pages 161–168, 1998.

[Bou85] Jean Bourgain. On Lipschitz embeddings of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52(1-2):46–52, 1985.

[CGN+03] Chandra Chekuri, Anupam Gupta, Ilan Newman, Yuri Rabinovich, and Alistair Sinclair. Embedding k-outerplanar graphs into ℓ_1. In *Proc. 14th SODA*, pages 527–536, 2003.

[Cla95] Kenneth L. Clarkson. Las Vegas algorithms for linear and integer programming when the dimension is small. *J. Assoc. Comput. Mach.*, 42(2):488–499, 1995.

[DL97] Michel Marie Deza and Monique Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1997.

[FRT03] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the thirty-fifth ACM symposium on Theory of computing*, pages 448–455. ACM Press, 2003.

[GKL03] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low–distortion embeddings. In *Proc. 44th FOCS*, pages 534–543, 2003.

[GNRS04] Anupam Gupta, Ilan Newman, Yuri Rabinovich, and Alistair Sinclair. Cuts, trees and ℓ_1-embeddings of graphs. *Combinatorica*, 24(2):233–269, 2004. (Preliminary version in 40th FOCS, 1999.).

[Goe97] Michel X. Goemans. Semidefinite programming and combinatorial optimization. *Mathematical Programming*, 49:143–161, 1997.

[Ind01] Piotr Indyk. Algorithmic aspects of geometric embeddings. In *Proc. 42nd FOCS*, pages 10–33, 2001.

[Kar85] Alexander V. Karzanov. Metrics and undirected cuts. *Mathematical Programming*, 32(2):183–198, 1985.

[KLMN04] Robert Krauthgamer, James Lee, Manor Mendel, and Assaf Naor. Measured descent: A new embedding method for finite metrics. In *Proc. 45th FOCS*, 2004. To appear.

[Lee05] James R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *Proc. 16th SODA*, 2005. To appear.

[Lin02] Nathan Linial. Finite metric-spaces—combinatorics, geometry and algorithms. In *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*, pages 573–586, Beijing, 2002. Higher Ed. Press.

[LLR95] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. (Preliminary version in *35th FOCS*, 1994).

[LN04] James Lee and Assaf Naor. Extending lipschitz functions via random metric partitions. *Inventiones Mathematicae*, 2004. To appear.

[LR99] F. Thomson Leighton and Satish B. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM*, 46(6):787–832, 1999. (Preliminary version in *29th FOCS*, pages 422–431, 1988).

[LW94] Nick Littlestone and Manfred K. Warmuth. The weighted majority algorithm. *Inform. and Comput.*, 108(2):212–261, 1994.

[Mat96] Jiří Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel Journal of Mathematics*, 93:333–344, 1996.

[Mat99] Jiří Matoušek. On embedding trees into uniformly convex Banach spaces. *Israel Journal of Mathematics*, 114:221–237, 1999. (Czech version in : *Lipschitz distance of metric spaces*, C.Sc. degree thesis, Charles University, 1990).

[Mat02] Jiří Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[Rab03] Yuri Rabinovich. On average distortion of embedding metrics into the line and into ℓ_1. In *Proceedings of the thirty-fifth ACM symposium on Theory of computing*, pages 456–462. ACM Press, 2003.

[Rao99] Satish B. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *15th Annual ACM Symposium on Computational Geometry*, pages 300–306, 1999.

[Shm97] David B. Shmoys. Cut problems and their application to divide-and-conquer. In Dorit S. Hochbaum, editor, *Approximation Algorithms for NP-hard Problems*, pages 192–235. PWS Publishing, 1997.

[SM90] Farhad Shahrokhi and D. W. Matula. The maximum concurrent flow problem. *J. Assoc. Comput. Mach.*, 37(2):318–334, 1990.

[Wel96] Emo Welzl. Suchen und Konstruieren durch Verdoppeln. In Ingo Wegener, editor, *Highlights der Informatik*, pages 221–228, Berlin, 1996. Springer.