Approximating the bandwidth via volume respecting embeddings

Uriel Feige*
Department of Computer Science and Applied Mathematics
The Weizmann Institute
Rehovot 76100, Israel
feige@wisdom.weizmann.ac.il.

September 13, 1999

Abstract

A linear arrangement of an n-vertex graph is a one-to-one mapping of its vertices to the integers $\{1, \ldots, n\}$. The bandwidth of a linear arrangement is the maximum difference between mapped values of adjacent vertices. The problem of finding a linear arrangement with smallest possible bandwidth in NP-hard. We present a randomized algorithm that runs in nearly linear time and outputs a linear arrangement whose bandwidth is within a polylogarithmic multiplicative factor of optimal. Our algorithm is based on a new notion, called *volume respecting embeddings*, which is a natural extension of *small distortion embeddings* of Bourgain and of Linial, London and Rabinovich.

1 Introduction

We consider the problem of minimizing the *bandwidth* of an undirected connected graph $G(V, E)$, where $n = |V|$ and $m = |E|$. One needs to find a linear arrangement of the vertices, namely, a one-to-one mapping $f : V \longrightarrow \{1, 2, \ldots n\}$, for which the bandwidth, i.e. $\max_{(i,j)\in E} |f(i) - f(j)|$, is minimized. A prime motivation for the bandwidth problem is its equivalence to the problem of minimizing the bandwidth of a sparse symmetric square matrix M. That is, finding a permutation matrix P such that the nonzero entries of PMP^T all lie on a narrow band along the diagonal. Minimizing the bandwidth of matrices helps in their storage and manipulations, including Gaussian elimination. See [6, 12, 29] for more information.

*Incumbent of the Joseph and Celia Reskin Career Development Chair.

On special families of graphs, the bandwidth can be computed in polynomial time [1, 17]. However, computing the bandwidth on general graphs is NP-hard [25], even on some subfamilies of trees [11, 24]. Deciding whether the bandwidth of a graph is at most 2 can be done in linear time [11]. For any fixed value k, deciding whether the bandwidth is at most k can be done in time $O(n^k)$ [28, 14], but apparently not much faster, as the problem is hard for any fixed level of the fixed parameter tractability hierarchy, even for trees [4].

In practice, heuristics for minimizing the bandwidth appear to work rather well [9]. Some theoretical justification for the success of heuristics is given in [30], where their performance on random graphs is investigated.

Algorithms for approximating the bandwidth in special classes of graphs are presented in [15, 18, 16]. However, for general graphs, there were no known algorithms giving nontrivial approximation ratios (such as n^δ for some $\delta < 1$) for the bandwidth, not even for trees. We note that there are known graphs on which the heuristics studied in [30] fail to give approximation ratios better than $\Omega(n/\log n)$.

A known lower bound on the bandwidth is obtained via the *local density* bound. Let $N(v,\rho)$ be the set of vertices at distance at most ρ from v. Then the local density of a graph is $D = \max_{v,\rho}[|N(v,\rho)|/2\rho]$, and the optimal bandwidth B^* satisfies $B^* \geq D$. Previous research suggests the conjecture that $B^* = O(D\log n)$. There are families of graphs with local density bounded above by a universal constant, whereas their bandwidth can be arbitrarily large [8, 7]. A gap of $\Omega(\log n)$ between local density and bandwidth can be demonstrated on trees [8] and on expander graphs.

We present a randomized algorithm that produces a linear arrangement with bandwidth $B = O(D(\log n)^3\sqrt{\log n \log\log n})$. This implies a polylogarithmic approximation of the optimum bandwidth. Our algorithm runs in time $O((n+m)(\log n)^c)$, for some small integer $c > 0$.

Independently of our work, Blum, Konjevod, Ravi and Vempala [3] obtained an algorithm based on semi-definite relaxation that approximates the bandwidth within a ratio of $\sqrt{n/B^*}\log n$.

Recently, Blache, Karpinski and Wirtgen [2] announced that it is NP-hard to approximate the bandwidth within a ratio better than 3/2. This was later improved by Unger to every constant factor [31].

A preliminary version of our work appeared in the *Proceedings of the 30th Annual ACM Symposium on the Theory of Computing, 1988, 90–99.* The current version gives tighter analysis of our algorithm, improving the bound on the approximation ratio by a factor of nearly $\log n$.

1.1 Overview

Our algorithm was inspired by the work of Linial, London and Rabinovich [22], who studied the geometry of graphs and its algorithmic applications. They considered embeddings of the vertices of a graph in a normed space, with small distortion of distances. Namely, if the distance between two vertices in a graph is d, then the distance of their images is between d and d/c, where $c \geq 1$ is the distortion factor. In [5, 22] it was shown how to obtain embeddings of small distortion (typically, $c \simeq \log n$), and how these embeddings can be used in the design of algorithms.

For the purpose of approximating the bandwidth we generalize small distortion embeddings to *volume respecting* embeddings. Rather than require small distortion only for the embeddings of pairs of vertices, we require small distortion for the embeddings of sets of k vertices, for values of k up to $\log n$. For sets of k vertices, we replace the notion of distance by a notion of volume. Roughly speaking, we want to embed the graph in Eulidean L-dimensional space, such that distances between pairs of vertices do not increase, and for any set of k vertices, the $(k-1)$-dimensional volume of the convex hull of their images is as large as possible, up to some distortion factor c. Observe that for $k = 2$, volume coincides with distance. (For exact definitions, see Section 3.) We note that for an embedding to be volume respecting for sets of size k, the host space must have dimension at least $(k-1)$. In Theorem 5 we show that known small distortion embeddings of [5, 22] can be modified to have the additional property of being volume respecting with only polylogarithmic distortion c.

Our algorithm PLOGBAND for the bandwidth can be viewed as having two major steps. The first step is a volume respecting embedding ϕ of the input graph G in Euclidean L-dimensional space. The second step is a projection on a random line ℓ. The algorithm outputs a linear arrangement that agrees with the ordering of the projections of the vertices on ℓ.

The analysis of our algorithm is based on the following principles. Consider any vertex u and edge (u, v). As both the volume respecting embedding and the projection are contractions, the distance between the images of u and v on ℓ is at most one. Now consider any other vertex w, whose distance (in G) from u is d. Then by the fact that ϕ is volume respecting, the distance between $\phi(u)$ and $\phi(w)$ is at least d/c, for some distortion factor c. Thereafter, if the line ℓ is chosen at random, the probability that w is projected to a point of distance at most one from the image of u on ℓ is inversely proportional to d/c. Using this observation it is easy to show that the expected number of vertices that fall between the projections of u and v (the *stretch* of edge (u, v)) is $O(cD \log n)$, where D is the local density of G. If this bound on the stretch of an edge happens to hold simultaneously for all edges, then the approximation ratio of the algorithm is $O(c \log n)$.

However, in addition to having bounds on the expected stretch of an edge, we need to give bounds on the deviation from expected stretch. Small distortion embeddings of [22] give conditions on pairs of vertices, and from this we can derive bounds on the variance of the stretch. This leads to a provable approximation ratio of roughly $O(\sqrt{n}(\log n)^{5/4})$. However, analysis based on the variance alone cannot give approximation ratios that are $o(\sqrt{n})$. (See Proposition 13.)

This is where the property of being volume respecting comes into play. By our requirements that sets of k vertices have large volumes under the embedding, we can get hold of the kth moment of the stretch. The key lemma in this respect is that the probability that all k vertices are projected to points at distance at most one from the image of u on ℓ is inversely proportional to their volume. Using kth moment analysis, we can obtain approximation ratios of roughly $n^{1/k}$. When $k = \log n$, the approximation ratio becomes polylogarithmic in n.

1.2 Roadmap

In Section 2 we describe algorithm PLOGBAND in a combinatorial fashion. In Section 3 we develop the theory of volume respecting embeddings. In Section 4 we use the theory of volume respecting embeddings to show that algorithm PLOGBAND approximates the bandwidth within polylogarithmic ratios. Section 5 contains the proofs missing from Section 3. Some open questions are presented in Section 6.

1.3 Notation

Let $G(V, E)$ be an input graph, with $|V| = n$ and $|E| = m$. G is always assumed to be connected. For two vertices $v_i, v_j \in V$, their distance $d(v_i, v_j)$ is the length (number of edges) of the shortest path connecting them (and 0 if $v_i = v_j$). For a set of vertices $S \subset V$ and a vertex $v \in V$, we let $d(S, v) = \min_{u \in S}[d(u, v)]$. For a graph, D denotes its local density ($D = \max_{v,\rho}[|N(v, \rho)|/2\rho]$, where $N(v, \rho)$ is the set of vertices at distance at most ρ from v), B^* denotes its optimal bandwidth, and B denotes the bandwidth returned by an algorithm under consideration.

R denotes the set of real numbers. Natural logarithms are denoted by ln and logarithms in base 2 are denoted by log.

The term *with high probability* is used for probabilities that are above 1/2 in cases that we do not care about exact bounds on the probability. The normal distribution with mean zero and standard deviation one, denoted by $N[0, 1]$, is given by the density function $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, for $x \in R$.

2 The algorithm

Our algorithm for approximating the bandwidth is described below. The parameter L influences both the running time of the algorithm and the approximation ratio. Our analysis (Theorem 1) assumes $L = \Theta((\log n)^4 \log\log n)$, but lower values of L may give better results in practice.

Algorithm PLOGBAND:

1. For $1 \leq j \leq L$, let p_j be spread evenly between $(1 - 1/\log\log n)$ and $1/n$. Namely, all ratios p_j/p_{j+1} are (approximately) the same.
2. Select sets $S_j \subset V$ for $1 \leq j \leq L$ as follows. For each vertex v_i and each set S_j, $v_i \in S_j$ with probability p_j, independently of all other events.
3. For each set S_j, compute for all $v_i \in V$ the distances $\phi_j(v_i) = d(S_j, v_i)$. (This can be done efficiently by adding a new vertex connected to all members of S_j, and performing breadth first search from this vertex.) If a set S_j is empty, then use the default $\phi_j(v_i) = 0$.
4. Choose independently at random $r_j \in N[0,1]$, for $1 \leq j \leq L$.
5. For each vertex v compute $h(v) = \sum_{j=1}^{L} r_j \phi_j(v)$.
6. Sort the vertices by order of increasing value of $h(v_i)$, breaking ties arbitrarily.
7. Output the sorted list of vertices as the linear arrangement.

We note that algorithm PLOGBAND is relatively simple to implement and runs in time roughly mL. Step 4 requires sampling from the normal distribution. For methods of doing so, see [19], for example. The standard deviation of the normal distribution is irrelevant for our purpose, as long as it is the same standard deviation for all r_j. Having a mean of zero is important. For the analysis of the algorithm to work, $\log(nL)$ bits of precision suffice.

Theorem 1 *For any $k \geq 1$, if $L = \Theta(k^2(\log n)^2 \log k)$, then with high probability, the linear arrangement output by algorithm PLOGBAND has bandwidth that is within a ratio of $O(k(\log n)^2 n^{1/k}\sqrt{\log n + k\log k})$ from the local density bound. In particular, if $k = \log n$, then the approximation ratio is $O((\log n)^3\sqrt{\log n \log\log n})$.*

The proof of Theorem 1 appears in Section 4.

3 Volume respecting embeddings

3.1 Volumes of finite metric spaces

A finite metric space (S,d) contains a finite set S of points, and a distance function $d : S \times S \to R^+$ that satisfies $d(i,i) = 0$, $d(i,j) = d(j,i)$, $d(i,j) + d(j,k) \geq d(i,k)$, for all points $i,j,k \in S$. Below we present several examples of finite metric spaces.

- *Euclidean finite metric space.* S is a set of k points in the plane, where the distance $d(i,j)$ between two points $i,j \in S$ is the Euclidean distance between them. More generally, S may be a set of k points in R^n equipped with Euclidean norm (or with some other norm such as l_1 or l_∞), and the distance between two points is the distance induced by the norm.

- *Graphical metric space.* S is the set of n vertices of an undirected graph $G(V,E)$ (i.e., $S = V$), and the distance between two points is the length of the shortest path connecting them. In fact, every finite metric space can be represented as an undirected graph, if we allow for variable edge lengths. This is done by identifying the points of S with the vertices of a complete graph on $|S|$ vertices, and setting the length of edge (i,j) to $d(i,j)$ (where $d(i,j)$ is as given by the finite metric space).

- *Subgraph metric space.* S is a subset of k vertices in an undirected graph $G(V,E)$, and the distance $d(i,j)$ between two vertices $i,j \in S$ is the length of the shortest path connecting i and j in G. Here the whole graph G is used in order to deduce the distance function for the finite metric space, even though the finite metric space contains only some of the vertices of G.

Throughout we shall assume that all finite metric spaces are connected (all distances are finite), and that all graphs are connected.

We shall now define a notion of *volume* of arbitrary finite metric spaces. As a motivating example, consider the Euclidean finite metric space (S,d) consisting of k points in R^{k-1}, equipped with l_2 (Euclidean) norm. In Euclidean spaces in general, there is a well known notion of volume. For R^1, this notion coincides with length, for R^2 it coincides with area, and for R^3 it coincides with three dimensional volume. This can be generalized to any dimension via integration. In general, for an object to have nonzero volume in R^{k-1}, it must have dimension $k-1$. The finite set S of points has dimension zero. Hence it by itself has volume zero. Our notion of volume also takes into account the arrangement of the points of S in space (which gives rise to the distances d). To obtain a "natural" $k-1$ dimensional object from S, we consider the *convex hull* of S. Namely, we treat the points of S as vertices of a $k-1$ dimensional simplex, and consider the volume of this simplex.

Though the above gives a natural definition of volume of Euclidean finite metric spaces, it still does not capture our notion of volume, not even for the special case of Euclidean finite metric spaces. Intuitively, we would like our notion of volume to have the property that if points in the finite metric space get further apart, the volume grows. However, this is not captured by just considering the convex hull of S, because we can simultaneously push the points of S further apart and bring them all to a $k-2$ dimensional subspace of R^{k-1}. This causes the volume to shrink to zero, rather than grow.

To handle such matters, we first relocate the points of S, subject to the constraint that for any pair of points in S, their distance does not grow (but it may shrink). Such a relocation is called a contraction. By applying a contraction, one may get out of degeneracies, and ensure that the volume is always positive. The volume of (S,d) is the maximum of the volumes of all simplexes that can be obtained by contractions.

We are now ready to define volumes of arbitrary finite metric spaces, and not just of Euclidean finite metric spaces. This is done by mapping the arbitrary finite metric space to an Euclidean finite metric space, and then taking volumes. In general, not all finite metric spaces can be mapped onto Euclidean finite metric spaces in a way that preserves distances. However, this will not cause a problem, as we allow for mappings that are contractions.

Definition 1 *An embedding of a finite metric space (S,d) in an L-dimensional Euclidean space is a function $\phi : S \rightarrow R^L$. Choosing some arbitrary orthonormal coordinate system, we view ϕ as a collection of L functions, $\phi_i : S \rightarrow R$, for $1 \leq i \leq L$, each specifying a different coordinate in R^L.*

For a point $p \in R^L$, we let $|p|$ denote its Euclidean norm (the l_2 norm of the vector p). For two points $p, q \in R^L$, their l_2 distance is given by $|p-q|$. The function ϕ is a *contraction* if for every $u, v \in S$, $|\phi(u) - \phi(v)| \leq d(u,v)$.

Definition 2 *For a set S of k points in R^L, their l_2 volume Evol(S) is the $(k-1)$-dimensional volume of the simplex that has the points of S as its vertices.*

Note that Evol(S) may be 0, if the points lie in a subspace of $k-2$ dimensions. Note also that Evol(S) is in general not the full dimensional volume of the convex hull of S with respect to the host space R^L. When $L > k-1$, the L-dimensional volume of the convex hull of S is 0, but Evol(S) may be positive.

Definition 3 *The volume of a finite metric space (S,d), Vol(S,d) (or just Vol(S), when d is clear from the context), is $\max[Evol(\phi(S))]$, where the maximum is taken over all contractions $\phi : S \rightarrow R^{|S|-1}$.*

We give some examples of volume for finite metric spaces. In all examples, the finite metric space is an undirected graph, and we give an embedding that maximizes the volume.

- (A path with two vertices.) $S = \{v_1, v_2\}$. $d(v_1, v_2) = \delta$. Here ϕ can embed S in R^1 as two points of distance δ apart, giving a volume of δ.
- (A path with three vertices.) $S = \{v_1, v_2, v_3\}$. $d(v_1, v_2) = d(v_2, v_3) = 1$. $d(v_1, v_3) = 2$. If one wants an embedding $\phi : S \to R^2$ that preserves distances, one can take $\phi(v_1) = (0, 0)$, $\phi(v_2) = (1, 0)$, $\phi(v_3) = (2, 0)$. However, then $\mathrm{Evol}(\phi(S)) = 0$, as the embedding is one-dimensional rather than two-dimensional. To maximize the volume it is necessary to shrink the distance between $\phi(v_1)$ and $\phi(v_3)$. One can use the embedding $\phi(v_1) = (0, 0)$, $\phi(v_2) = (1, 0)$, $\phi(v_3) = (1, 1)$. Here $|\phi(v_1) - \phi(v_3)| < d(v_1, v_3)$, and the embedding gives a right angle triangle with area 1/2.
- (A triangle.) $S = \{v_1, v_2, v_3\}$. $d(v_1, v_2) = d(v_2, v_3) = d(v_1, v_3) = 1$. Here the best one can do is to embed S as an equilateral triangle, giving volume $\sqrt{3}/4$.

For a finite metric spaces (S, d) with $|S| = k$, it is easy to verify the following properties of $\mathrm{Vol}(S)$:

- $\mathrm{Vol}(S) > 0$.
- When $k = 2$, volume coincides with distance.
- Volume is monotonically increasing in the distances d.
- When all distances scale by α, volume scales by α^{k-1}.
- Volume is continuous in distance.

3.2 Tree volume

Recall that a finite metric space can be represented as an undirected graph. In this section we characterize the volume of a finite metric space in terms of more familiar graph properties.

We consider first the special case of *tree metric spaces*. Let the graph $G(V, E)$ be a tree with arbitrary positive edge lengths. Let $V = \{v_1, \ldots, v_n\}$, let v_1 be the root of the tree, and for $i \geq 2$, let δ_i denote the length of the edge connecting vertex v_i to its parent vertex. Consider (S, d), the tree metric space with $S = V$ and d being the shortest path distance along the tree edges. We wish to compute $\mathrm{Vol}(S)$.

For the tree G (and hence for the tree metric space (S, d)), we introduce a notion of *tree volume* that is defined as $\mathrm{Tvol}(S) = \Pi_{i=2}^{n} \delta_i$. Namely, the tree volume is the product of the lengths of the tree edges.

Theorem 2 *For every tree metric space (S, d) with n points,*

$$Vol(S) = \frac{Tvol(S)}{(n-1)!}$$

Proof: To show that $\mathrm{Vol}(S) \geq \frac{\mathrm{Tvol}(S)}{(n-1)!}$, use a contraction that maps S to R^{n-1} by placing the root of the tree at the origin, and then placing each new vertex v_i at distance δ_i from its parent vertex, in a direction orthogonal to all previously used directions. This mapping is a contraction because l_2 distances are never longer than l_1 distances. The volume obtained is $(\Pi_{i=2}^{n}\delta_i)/(n-1)!$, as desired.

We now prove by induction on n that $\mathrm{Tvol}(S) \geq (n-1)!\mathrm{Vol}(S)$. When $n = 2$, then $\mathrm{Tvol}(S) = \mathrm{Vol}(S)$. For the inductive step, prove the theorem for an arbitrary $(n+1)$-set S. Remove from S a vertex u that is a leaf in G, to obtain an n-set S' and an associated tree G'. By the inductive hypothesis, $\mathrm{Tvol}(S') \geq (n-1)!\mathrm{Vol}(S')$. Let δ_u be the length of the tree edge connecting u to $v \in G'$. Then $\mathrm{Tvol}(S) = \delta_u \mathrm{Tvol}(S')$. Observe that adding a vertex u to an embedding of S' in R^n gives a simplex whose volume increases by a factor of δ'_u/n compared to the volume of the embedding of S', where δ'_u is the distance of u from the $(n-1)$-dimensional subspace containing the embedding of S'. As $\delta'_u \leq \delta_u$, we obtain that $\mathrm{Vol}(S) \leq \delta_u \mathrm{Vol}(S')/n$. □

When the graph representing a finite metric space is not a tree, we do not get an exact characterization of $\mathrm{Vol}(S)$, but we can get a very good estimate. A *spanning tree* of S is a set of $|S| - 1$ edges between points of S that makes the set S connected. The length of each edge in the spanning tree is taken to be the distance between the respective vertices according to the metric on S. A *minimum spanning tree* on S is one in which the sum of the edge lengths in minimized. It is well known that the greedy algorithm (sorting edges by edge length and greedily adding low weight edges to construct a spanning tree) produces a minimum spanning tree. It follows that the minimum spanning tree also minimizes the product of the edge lengths (e.g., by applying the greedy algorithm on the logarithms of the lengths).

Definition 4 *The* tree volume *of a finite metric space (S, d), $Tvol(S, d)$ (or just $Tvol(S)$, when d is clear from the context), is the product of the edge lengths in a minimum spanning tree for S.*

Theorem 2 does not necessarily hold when the finite metric space is not induced by a tree. For example, for the triangle with unit edge lengths, its tree volume is 1, whereas its volume is $\sqrt{3}/4$ (Theorem 2 would imply volume 1/2). Nevertheless, for every metric space on k points, tree volume characterizes volume up to a factor of $2^{(k-2)/2}$.

Theorem 3 *Let (S, d) be a finite metric space with k points. Then*

$$Vol(S) \leq \frac{Tvol(S)}{(k-1)!} \leq Vol(S)2^{(k-2)/2}$$

The proof of Theorem 3 is given in Section 5.1.

3.3 Computing volumes

Theorem 3 shows that Vol(S) can be approximated within a factor of $2^{(|S|-2)/2}$ in polynomial time, by computing Tvol(S). We do not know whether Vol(S) can be computed exactly in polynomial time (note in particular that Vol(S) may be irrational even if all distances are rational – the triangle example). However, there is a polynomial time algorithm that approximates Vol(S) within arbitrary precision.

Let (S, d) be a finite metric space with k points. To compute Vol(S), one needs to find a contraction $\phi(S) \rightarrow R^{k-1}$ that maximizes Evol($\phi(S)$). This contraction maps the points of S onto the vertices of a simplex. We shall denote this simplex by K. Each vertex of K is a vector in R^{k-1}. Consider the k by $(k-1)$ matrix M_K, whose rows are the vertices of K. We now translate K such that one of its vertices coincides with the origin, by subtracting the last row of M_K from all other rows, and removing it from M_K to obtain an order $k-1$ square matrix M. The row vectors $b_1, \ldots, b_{k-1}$ of M can be thought of as defining a $k-1$ dimensional lattice, whose fundamental parallelohedron is the set of all points $\sum_{i=1}^{k-1} \alpha_i b_i$, where $0 \leq \alpha_i \leq 1$. As is well known (see for example [21], Section 2), $|Det(M)|$ is the volume of the above parallelohedron, and the volume of the simplex K is $|Det(M)|/(k-1)!$.

More generally, the k vertices of the simplex K may be expressed as points in L-dimensional space, where $L \geq k-1$. We can associate with K a k by L matrix M_K and derive from it a $(k-1)$ by L matrix M, as above. To compute the volume of K, one should now consider the $k-1$ by $k-1$ matrix $Q = MM^T$, and then Evol(K) $= \sqrt{Det(Q)}/(k-1)!$. (If M is a square matrix, then of course $\sqrt{Det(Q)} = |Det(M)|$.)

Algebraically, computing Vol(S) can be done by finding an order $(k-1)$ matrix M, where each row b_i has l_2 norm at most $d(v_i, v_k)$, the l_2 norm of $b_i - b_j$ is at most $d(v_i, v_j)$, and the determinant of M is maximized. Such a matrix M can be found indirectly, by first using the ellipsoid algorithm to find the matrix $Q = MM^T$, and then decomposing Q to obtain M.

Theorem 4 *For every $\epsilon > 0$, there is a polynomial time algorithm that for every finite metric space (S, d) computes Vol(S) within a multiplicative error of $1 + \epsilon$.*

For the proof of Theorem 4 see Section 5.2.

3.4 Volume respecting embeddings

Let $G(V, E)$ be an undirected graph on n vertices. The reader may assume for simplicity that all edges have unit length, though the results presented apply also to the case of arbitrary positive length edges (and hence, to every finite metric space). By selecting a set S of $k < n$ vertices in G, we obtain a subgraph metric space on k points, where the distance between two points is the distance between the respective vertices in G. There are $\binom{n}{k}$ subgraph metric spaces induced by the graph G. Each such metric space has a volume, which is determined by an optimal contraction that maps the points of S to Euclidean space. One may ask whether there is a collection of $\binom{n}{k}$ optimal contractions, one for each k-set $S \subset V$, such that all contractions are mutually consistent. Namely, if some vertex v belongs both to S and to S', then v is mapped to the same point in Euclidean space under the two optimal contractions ϕ and ϕ' associated with S and S'. Equivalently, one may ask whether there is a single contraction $\phi : V \to R^{n-1}$ such that for every k-set $S \subset V$, $\mathrm{Vol}(S) = \mathrm{Evol}(\phi(S))$. The answer to this question is negative.

Consider for example a tree composed of a root and three leaves (where edges have unit length). This is a graph on four vertices. Let $k = 2$. Hence we are interested in a contraction that maps the vertices of the tree to R^3, such that the volumes of all 2-sets (i.e., distances between pairs of points), are preserved. W.l.o.g., assume that the root is mapped to the origin. Then the three leaves must be mapped to points at distance one from the origin. But then, there is no way of arranging the leaves in Euclidean space such that the distance between any two of them is two. This demonstrates the known fact that some graphs do not have an isometric embedding in Euclidean space.

An additional type of problem that may arise is the following. Consider the graph above and the set S of three leaves. The volume of S is $\sqrt{3}$, obtained by embedding the leaves as vertices of an equilateral triangle with side length 2. However, there is no contraction of the whole graph in which the leaves form such a triangle, because this would imply that the root of the tree is at distance greater than one from at least one of the leaves.

The above discussion leads us to the consideration of contractions that do not produce maximal volume.

Definition 5 *For a finite metric space (S, d) and contraction ϕ, the* distortion $\eta(\phi, S)$ *is defined as $(Vol(S)/Evol(\phi(S)))^{1/(|S|-1)}$. If $Evol(\phi(S)) = 0$, then $\eta(\phi, S) = \infty$.*

Now we come to the central definition of this section.

Definition 6 *For an undirected graph $G(V, E)$, a contraction $\phi : V \to R^L$ is (k, η)-*volume respecting *if for every set $S \subset V$ of k vertices, the distortion $\eta(\phi, S)$ is at most η.*

The notion of volume respecting embedding is a generalization of the small distortion embeddings of Bourgain [5] and of Linial, London and Rabinovich [22]. They consider only the distortion of distances (the case $k = 2$), whereas we consider volumes. It is shown in [5, 22] that every n-vertex connected graph has a $(2, O(\log n))$-volume respecting embedding $\phi : V \to R^{O(\log n)}$, and that for some graphs (expanders) a distortion of $\Omega(\log n)$ is unavoidable. The geometry of low distortion embeddings was used in [22] to obtain algorithms for various graph partitioning problems. The geometric properties of the embeddings are easier to use when the host space is of low dimension. Hence one of the goals of [22] is to obtain small distortion embeddings in R^L where L is as small as possible, preferably a constant independent of n. We note that our generalization to $k > 2$ (we shall take $k = \log n$) conflicts with keeping the dimension L small, because we must have $L \geq k - 1$ to keep the distortion of volumes finite.

We now present a randomized algorithm for constructing volume respecting embeddings $\phi : V \to R^L$. For $k = 2$, Bourgain [5] shows that a similar algorithm achieves a distortion of $O(\log n)$. For a different choice of parameters in this algorithm (a larger value of L), we show that the k-volume distortion of this algorithm is bounded by a polynomial in k and $\log n$. The dimension of the embedding is $L = \Theta(k^2 \log^2 n \log k)$.

The random subsets embedding:

1. Select $\log k \log n$ different *defining probabilities* p_l, spread out at roughly equal ratios between $1 - 1/2k$ and $1/n$. More specifically, for $1 \leq l \leq \log k \log n$, let
   - $t_l = (1 + 1/\log k)^l$, and
   - $p_l = 1 - (1/2k)^{1/t_l}$, implying
   - $(1 - p_l)^{t_l} = 1/2k$.

   Divide the L coordinates into $\log k \log n$ blocks of coordinates. Each block contains $L/\log k \log n = \Theta(k^2 \log n)$ coordinates. For a coordinate $1 \leq j \leq L$, let p_j be equal to the defining probability p_l, where l is the index of the block to which the coordinate belongs.

2. Select sets $S_j \subset V$ for $1 \leq j \leq L$ as follows. For each vertex v_i and each set S_j, $v_i \in S_j$ with probability p_j, independently of all other events.

3. For each vertex v_i and component $1 \leq j \leq L$, let $\phi_j(v_i) = d(S_j, v_i)$. If a set S_j is empty, then use the default $\phi_j(v_i) = 0$.

4. Take $\phi(v)$ as $\frac{1}{\sqrt{L}}(\phi_1(v), \ldots, \phi_L(v))$.

Observe that the first three steps of the random subsets embedding almost coincide with the first three steps of algorithm PLOGBAND (using $\log k \simeq \log \log n$). The differences in step 1 are due to the fact that we wanted to keep the description of algorithm

PLOGBAND simple, whereas we wanted to describe the random subsets embedding in a way that facilitates its analysis.

Theorem 5 *For every connected undirected graph $G(V, E)$ and $L = \Theta(k^2 \log^2 n \log k)$, the random subsets embedding $\phi : V \to R^L$ has high probability of being (k, η)-volume respecting, with $\eta = O(\sqrt{\log n}\sqrt{\log n + k \log k})$.*

The proof of Theorem 5 appears in Section 5.3.

The random subsets embedding provides a randomized algorithm and an existence proof for embeddings with low distortion. However, it does not necessarily give an embedding with smallest possible distortion. The following theorem can be useful when one seeks an embedding with optimal distortion.

Theorem 6 *For every fixed k and every $\epsilon > 0$, there is a polynomial time algorithm that for every input graph computes a $(k, (1 + \epsilon)\eta)$-volume respecting embedding, where η is the least distortion for which the input graph has a (k, η)-volume respecting embedding.*

For the proof of Theorem 6 see Section 5.2, which also addresses the case of nonconstant k.

3.5 Large volumes have large projections

Volume respecting embeddings are introduced in this work as an algorithmic tool. For certain graph problems, such as estimating the bandwidth of a graph, it is proposed to first embed the graph in Euclidean space using a volume respecting embedding, and then to use the geometry of the embedding so as to solve the underlying graph problem. In volume respecting embeddings, k-subsets of vertices of the graph form simplexes with large volumes. In this section we show implications that can be derived from the fact that an object has large volume. They concern the length of the projection of the object on a random line.

In our analysis, we shall use the following notation and facts. For a unit ball in R^i, let V_i denote its volume, and let A_i denote its surface area. Then $V_i = 2\pi^{i/2}/i\Gamma(i/2)$ and $A_i = 2\pi^{i/2}/\Gamma(i/2)$, where $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$, and $\Gamma(n + 1) = n\Gamma(n)$.

Let us first consider a one dimensional object, namely, a unit vector, lying in R^L. Taking an arbitrary orthonormal basis for R^L, the unit vector decomposes into vectors $v_1, \ldots, v_L$ such that $\sum |v_i|^2 = 1$. It follows that the expectation of the square of the length of the projection of a unit vector on a random subspace R^k is k/L. Intuitively, we then expect the length of the projection to be roughly $\sqrt{k/L}$. The following proposition bounds the probability of large deviations from this estimate.

Proposition 7 *Let r be a random unit vector in R^L chosen with spherical symmetry, and let R^k be a subspace of R^L. Let ℓ denote the length of the projection of r on R^k. Then:*

- Small projection: *for every* $0 < \epsilon < 1$, $Pr_r[\ell < \epsilon\sqrt{k}/\sqrt{L}] \leq (\beta\epsilon)^k$, *for some universal* $\beta > 0$.
- Large projections: *for* $c > 1$ *and* $k = 1$, $Pr_r[\ell > \sqrt{c/L}] \leq e^{-c/4}$. *When* L *is large, the exponent tends to* $-c/2$ *(rather than* $-c/4$*).*

The proof of Proposition 7 appears in Section 5.4.

Consider a full dimensional convex body K in R^k. Its projection on a line is an interval. We show that if the line is chosen at random, then the probability that this interval is short is inversely proportional to the volume of K.

Lemma 8 *Let* K *be a convex body in* R^k, *and let* r *be a unit vector in* R^k. *Consider the projection of* K *on the line passing through the origin in the direction of* r, *given by the inner products* $\langle v, r\rangle$, *for points* $v \in K$, *and let* $P(K,r) = \max_{v\in K}\langle v,r\rangle - \min_{v\in K}\langle v,r\rangle$. *(*$P(K,r)$ *is the length of the segment covered by the projection of* K *on a line in the direction of* r*.) If the direction of* r *is chosen uniformly at random, then* $Pr[P(K,r) \leq 1] < (O(1)/k)^{k/2}/Evol(K)$.

Proof: To verify steps of this proof, see for example [20] (pages 6,32,107,111) and [27] (pages 9,10). Consider the difference body $K - K$, where $p \in K - K$ iff there are $p_1, p_2 \in K$ such that $p = p_1 - p_2$. The body $K - K$ is convex and o-symmetric (if $p \in K - K$, then so is $-p$). As for volumes, $\mathrm{Evol}(K - K) \geq \mathrm{Evol}(K)$. (When K is a simplex, as will be in cases that interest us, then in fact $\mathrm{Evol}(K - K) = \binom{2k}{k}\mathrm{Evol}(K)$.) Consider now the *polar* of $K - K$, denoted by $(K - K)^*$, defined as the set of points y having $\langle x, y\rangle \leq 1$ for each point $x \in K - K$. This polar body is convex and o-symmetric. Let V_k denote the volume of the k-dimensional unit ball. Then $\mathrm{Evol}((K - K)^*) \leq (V_k)^2/\mathrm{Evol}(K - K)$.

Observe that $P(K,r) \leq 1$ iff the radius of $(K - K)^*$ in the direction r is at least one (i.e., if $r \in (K - K)^*$). This can be verified as follows. Assume that $P(K,r) \leq 1$. Then for any two points $u, v \in K$, $\langle u - v, r\rangle \leq 1$. Hence for every point $x \in K - K$ we have $\langle x, r\rangle \leq 1$, implying $r \in (K - K)^*$. Conversely, assume that $P(K,r) > 1$. Then there are two points $u, v \in S$ with $\langle u - v, r\rangle > 1$. Hence for the point $(u - v) \in K - K$ we have $\langle (u - v), r\rangle > 1$, implying $r \notin (K - K)^*$.

The measure of directions in which the radius of $(K - K)^*$ is of length at least one is at most $\mathrm{Evol}((K - K)^*)/V_k$. Hence $Pr[P(K,r) \leq 1] \leq V_k/\mathrm{Evol}(K)$. Using $V_k = 2\pi^{k/2}/k\Gamma(k/2)$ (where $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$, and $\Gamma(n+1) = n\Gamma(n)$), the proof of the lemma is completed. □

We now generalize Lemma 8 to the situation in which K is a k-dimensional convex body embedded in R^L, where $L > k$, and the unit vector r is chosen at random in R^L. Lemma 8 does not apply in this case, because the part of r that K sees (the projection of r to R^k) is in general shorter than a unit vector. Hence it is now more likely that the projection of K on r is short. Theorem 9 handles this case.

Theorem 9 *Let K be a k-dimensional convex body in R^L, where $L > k$, and let r be a unit vector in R^L. Consider the projection of K on the line passing through the origin in the direction of r, given by the inner products $\langle v, r\rangle$, for $v \in K$, and let $P(K,r) = \max_{v\in K}\langle v,r\rangle - \min_{v\in K}\langle v,r\rangle$. If the direction of r is chosen uniformly at random, then*

$$Pr[P(K,r) \leq 1] < \max[1, \log(Evol(K))]\frac{(O(1)L)^{k/2}}{k^k Evol(K)}$$

The proof of Theorem 9 appears in Section 5.4.

Scaling K in Theorem 9 by $1/c$, we obtain that

$$Pr[P(K,r) \leq c] < \max[1, \log(\mathrm{Evol}(K))]\frac{(O(1)L)^{k/2}c^k}{k^k \mathrm{Evol}(K)}$$

3.6 A local density bound on volumes

Consider a graph $G(V,E)$ with unit edge lengths and in it select a random subset S of k vertices. This defines a finite metric space, with the distance between two points in S being their respective distance in G. How large can we expect $\mathrm{Vol}(S)$ to be? In graphs, it is easier to deal with tree volume, so the above question can be rephrased in terms of $\mathrm{Tvol}(S)$.

Recall that the *local density* D of a graph is $D = \max_{v,\rho}[|N(v,\rho)|/2\rho]$, where $N(v,\rho)$ is the set of vertices at distance at most ρ from v. Hence there are at most n/k vertices at distance at most $n/2kD$ from v. Intuitively (and nonrigorously), when we select a set of k vertices at random, the distance from $v \in S$ to the closest other vertex in S is expected to be $\Omega(n/kD)$, and then the tree volume of S is expected to be $\Omega((n/kD)^{k-1})$.

For our purposes (applications such as those in section 3.5), it is more convenient to consider the expectation of $1/\mathrm{Tvol}(S)$ rather than of $\mathrm{Tvol}(S)$. If it were true that the expectation of $1/T(S)$ is the inverse of the expectation of $T(S)$, we would then get

$$\sum_S \frac{1}{\mathrm{Tvol}(S)} \leq \binom{n}{k}(kD/n)^{k-1} < \frac{n}{k}(eD)^{k-1}$$

The above estimate which was derived nonrigorously is not very far from the truth, as the following theorem shows.

Theorem 10 *For any connected graph $G(V,E)$ with local density D and for any k,*

$$\sum_S \frac{1}{Tvol(S)} \leq n(D(1+\ln(n/D)))^{k-1}$$

where summation is taken over all subsets $S \subset V$ of cardinality k.

The proof of Theorem 10 appears in Section 5.5.

4 Analysis of algorithm PLOGBAND

Let us describe algorithm PLOGBAND in geometric terms.

Algorithm PLOGBAND – Geometric version.

1. Embed G in Euclidean space using a $(\log n, \eta)$-volume respecting embedding with polylogarithmic distortion η. More specifically, use the random subsets embedding to embed the input graph G in R^L, where $L = \Theta(\log^4 n \log\log n) = \Theta(k^2 \log^2 n \log k)$, and $k = \log n$.

2. Project the embedding on a line in the direction of a random unit vector r, obtaining for each vertex v a point $h'(v)$ on the line.

3. Sort the vertices in the order of $h'(v)$, and output the sorted list of vertices as the linear arrangement.

The geometric version of algorithmic PLOGBAND is essentially the same as the combinatorial version. The ϕ_j defined in the combinatorial version and the ϕ_j defined in the random subsets embedding are the same, except for a scaling factor of $\sqrt{L}$. (There is also a technical difference in our explanation of how to choose the probabilities p_j in each case. This is discussed after the description of the random subsets embedding and is ignored here.) The r_js chosen in the combinatorial version can be scaled by $(\sum r_j^2)^{-1/2}$, giving a unit vector $r = (\sum r_j^2)^{-1/2}(r_1, \ldots, r_L)$ in a direction chosen at random with spherical symmetry (see [19], page 130). Hence we obtain that for every vertex v, $h(v) = \sqrt{L(\sum r_j^2)}h'(v)$, and the sorted orders obtained by the two versions of the algorithm are the same.

We now prove Theorem 1. The proof refers to the geometric version of algorithm PLOGBAND.

Proof: (**Theorem 1**.) The bandwidth B is determined by the maximum number of vertices v_i for which $h'(u) \leq h'(v_i) \leq h'(v)$, for some edge $(u, v) \in E$. Note that h' is a random function (depending on the random subsets embedding ϕ and on the random vector r).

Proposition 11 *With high probability over the choice of h', for every $(u, v) \in E$, $|h'(u) - h'(v)| \leq 2\sqrt{\log n / L}$.*

Proof: For the volume respecting embedding ϕ and for edge (u, v) we have $|\phi(u) - \phi(v)| \leq d(u, v) = 1$. Now $h'(u) - h'(v)$ is just the value of the product of the vector $(\phi(u) - \phi(v))$ by a random unit vector r. From Proposition 7 it follows that the probability that $|h'(u) - h'(v)| > 2\sqrt{\log n / L}$ is at most $1/n^2$. Hence with high probability, Proposition 11 holds for all edges of G simultaneously. □

Hence we can assume that indeed, for every $(u, v) \in E$, $|h'(u) - h'(v)| \leq 2\sqrt{\log n/L}$. Call a set S of k vertices *bad* if for every two vertices $u, v \in S$, $|h'(u) - h'(v)| \leq 2\sqrt{\log n/L}$. Let N_k denote the number of bad k-sets. The following lemma connects between N_k and the local density D, and is the key to the proof of Theorem 1.

Lemma 12 *If $L = \Theta(k^2(\log n)^2 \log k)$, then with high probability over the choice of h' in algorithm PLOGBAND, $N_k \leq n \log n(\alpha(\log n)^2 D\sqrt{\log n + k \log k})^{k-1}$, for some universal constant $\alpha > 0$.*

Assume Lemma 12, and let B be the bandwidth of the solution returned by algorithm PLOGBAND. Then there is some value z such that $B + 1$ vertices w satisfy $z \leq h'(w) \leq z + 2\sqrt{\log n/L}$. Hence $N_k \geq \binom{B+1}{k}$. But from Lemma 12 we may assume $N_k \leq n \log n(\alpha(\log n)^2 D\sqrt{\log n + k \log k})^{k-1}$. By comparing the upper bound and lower bound on N_k, we have that $\binom{B+1}{k} \leq n \log n(\alpha(\log n)^2 D\sqrt{\log n + k \log k})^{k-1}$, implying $B/D = O(k(\log n)^2 n^{1/k}\sqrt{\log n + k \log k})$. As $D \leq B^*$, the proof of Theorem 1 follows. $\square$

We now prove Lemma 12.

Proof:(**Lemma 12**.) For simplicity, we shall ignore multiplicative terms of the form $(O(1))^{k-1}$ throughout the proof. These terms affect only the constant α in Lemma 12.

The choice of h' can be decomposed to a choice of ϕ which is volume respecting, and a projection on a random unit vector r. We combine the following **facts**:

1. Thm. 3: for every k-subset S,

$$\text{Vol}(S) \simeq \text{Tvol}(S)/(k-1)!$$

2. Thm. 5: with high probability over the choice of ϕ, for every k-subset S, $\text{Evol}(\phi(S)) \geq \text{Vol}(S)/(\log n(\log n + k \log k))^{(k-1)/2}$. Hence

$$\text{Evol}(\phi(S)) \geq \text{Tvol}(S)/(k^2 \log n(\log n + k \log k))^{(k-1)/2}$$

3. Thm. 9: for k vertex simplex $\phi(S)$ and random r, $Pr[P(\phi(S), r) \leq 2\sqrt{\log n/L}] \leq \log(\text{Evol}(\phi(S))(\sqrt{\log n})^{k-1}/k^{k-1}\text{Evol}(\phi(S))$. As S is a set of vertices in an undirected graph in which edges have unit length, we obtain that $\log(\text{Evol}(\phi(S))) < \log(n^{k-1}) < k \log n$. Ignoring the term $k \leq (O(1))^{k-1}$ we conclude that

$$Pr[P(\phi(S), r) \leq 2\sqrt{\log n/L}] \leq \log n(\log n\sqrt{\log n + k \log k})^{k-1}/\text{Tvol}(S)$$

4. Thm. 10: for k-sets S,

$$\sum_S 1/\text{Tvol}(S) \leq n(D \log n)^{k-1}$$

Recall that $P(\phi(S), r) \leq 2\sqrt{\log n / L}$ means that S is bad in the sense of Theorem 1. From fact 3, the expected number of bad sets is at most $\log n(\log n\sqrt{\log n + k\log k})^{k-1}\sum_S 1/\mathrm{Tvol}(S)$. Using fact 4 this expectation is at most $n\log n((\log n)^2 D\sqrt{\log n + k\log k})^{k-1}$. To obtain a high probability upper bound on N_k, use Markov's inequality. $\square$

4.1 Volumes versus distances

Our algorithm for approximating the bandwidth has two conceptual steps: a volume-respecting embedding, and a random projection. As discussed earlier, volume respecting embeddings are generalizations of small distortion embeddings of [5, 22], which only respect distances. Could we have used small distortion embeddings rather than volume respecting embeddings in algorithm PLOGBAND? Consider the following algorithm.

1. Embed G in Euclidean space using a $(2, O(\log n))$-volume respecting embedding (i.e., a small distortion embedding).
2. Project the embedding on a line in the direction of a random unit vector r, obtaining for each vertex v a point $h'(v)$ on the line.
3. Sort the vertices in the order of $h'(v)$, and output the sorted list of vertices as the linear arrangement.

With high probability, the linear arrangement output by the above algorithm has bandwidth $O(\sqrt{nD}(\log n)^{5/4})$. (The proof of this is a simplified version of the proof of Theorem 1, and is left as an exercise.) The following proposition shows that the above algorithm does not guarantee a polylogarithmic approximation ratio for the bandwidth.

Proposition 13 *There are graphs $G(V, E)$ with bandwidth $O(1)$, and embeddings $\phi : V \rightarrow R^2$ that are $(2, O(1))$-volume respecting, such that for any line $\ell \in R^2$, projecting $\phi(G)$ on ℓ gives a linear arrangement with bandwidth $B = \Omega(\sqrt{n})$.*

Proof: Consider a 2 by $n/2$ grid, in which each vertex is connected to its grid neighbors. This graph has bandwidth 2 (by interleaving the top path with the bottom path). It can be embedded in R^2 with only constant distortion by bending the grid so that it becomes half of a circle centered at the origin. The graph now looks like a railroad track making a half-circle, where the outer rail has radius roughly $n/2\pi$ and the inner rail has radius roughly $n/2\pi - 1$.

We now show that for any choice of line ℓ passing through the origin, the linear arrangement that results from orthogonal projection on ℓ has bandwidth $\Omega(\sqrt{n})$. Let p_1 be the vertex of G closest to where ℓ crosses the inner circle, and let p_2 be the vertex of G closest to where ℓ crosses the outer circle. These two vertices are adjacent in G. From p_2, one must rotate by an angle of $\Omega(1/\sqrt{n})$ to get a point on the outer circle whose projection on ℓ is smaller than $2\pi/n - 1$. Hence $\Omega(\sqrt{n})$ vertices on the outer circle are projected to points between the projections of p_1 and p_2. The bandwidth obtained is $\Omega(\sqrt{n})$, instead of $O(1)$. $\square$

5 Proofs

5.1 Tree volume versus volume

We prove Theorem 3, that tree volume approximates volume.

Proof: (**Theorem 3**.) In theorem 2 we have seen that if (S, d) is a tree metric space, then $\mathrm{Vol}(S) = \mathrm{Tvol}(S)/(k-1)!$. If (S, d) is not a tree metric space, we can modify it to (S, d') that is a tree metric space, where $d' \geq d$, by letting $d'(u, v)$ be the distance between u and v as measured by going along edges of the minimum spanning tree of S. The minimum spanning tree and the value of $\mathrm{Tvol}(S)$ do not change, whereas $\mathrm{Vol}(S, d') \geq \mathrm{Vol}(S, d)$, by the fact that volume increases monotonically with distance. Hence $\mathrm{Vol}(S) \leq \mathrm{Tvol}(S)/(k-1)!$.

We now show that $\mathrm{Tvol}(S) \leq (k-1)!\mathrm{Vol}(S)2^{(k-2)/2}$. Use the minimum spanning tree to embed $S = \{v_1, \ldots, v_k\}$ in k dimensional space as follows. Assume w.l.o.g. that (v_1, v_2) is the longest edge in the spanning tree (breaking ties arbitrarily), and direct all spanning tree edges towards v_1. Assume without loss of generality that for every v_i, its parent vertex v_j has $j < i$, and let δ_i be the length of the tree edge from v_i to v_j. We now describe the embedding $\phi : S \to R^k$. For all $v_i \in S$, $\phi(v_i)$ will have only nonnegative coordinates and l_1 norm of $\frac{\sqrt{2}}{2}\delta_2$. The first coordinate of $\phi(v_1)$ is $\frac{\sqrt{2}}{2}\delta_2$, and the other coordinates are 0. The second coordinate of $\phi(v_2)$ is $\frac{\sqrt{2}}{2}\delta_2$, and the other coordinates are 0. For every new vertex v_i, $\phi(v_i)$ is obtained from $\phi(v_j)$ for its parent vertex by the following process. We subtract a total of $\frac{\sqrt{2}}{2}\delta_i$ from the l_1 norm of $\phi(v_j)$, starting with the highest nonzero coordinate, and working backwards only with nonzero coordinates. In addition, coordinate i is set to $\frac{\sqrt{2}}{2}\delta_i$. For example, if $\phi(v_j) = (0, 3, 1, 1, 0, 0)$ and $\delta_i = 2$ then $\phi(v_i) = (0, 3, 2 - \sqrt{2}, 0, \sqrt{2}, 0)$ (assuming that v_j is the parent of v_i).

It is not hard to see that $\mathrm{Evol}(\phi(S)) \geq \delta_2(\Pi_{i=3}^{k}\frac{\sqrt{2}}{2}\delta_i)/(k-1)!$. This follows from the fact that $|\phi(v_1) - \phi(v_2)| = \delta_2$, and for every $i \geq 3$, $\phi(v_i)$ is at distance $\frac{\sqrt{2}}{2}\delta_i$ in a direction orthogonal to the subspace spanned by $\{\phi(v_1), \ldots, \phi(v_{i-1})\}$.

Using Pythagoras' theorem it can be seen that this embedding is a contraction with respect to tree edges. To see that it is also a contraction with respect to other intervertex distances, use the fact that any nontree edge (v, v') closes a cycle in which all other edges are tree edges. Each of these tree edges is not longer than edge (v, v'). Let u be the highest index common ancestor of v and v' with respect to the minimum spanning tree. Let δ be the length of the longest tree edge between u and v, and let δ' be the length of the longest tree edge between u and v'. W.l.o.g., assume that $\delta \geq \delta'$. Then it can be verified that the l_1 distance between $\phi(v)$ and $\phi(v')$ is at most $\frac{\sqrt{2}}{2}(\delta + \delta' + (\delta - \delta'))$, where each of the three components in the summation is in a subspace orthogonal to the others. It follows that the l_2 distance is at most $\delta \leq d(v, v')$. □

5.2 Using the Ellipsoid algorithm

We prove Theorem 4.

Proof: (**Theorem 4.**) Let (S, d) with $S = \{v_1, \ldots v_k\}$ be a finite metric space for which we want to find a contraction $\phi : S \to R^k$ such that Evol$(\phi(S))$ is maximized, and hence compute Vol(S). W.l.o.g., we may assume that $\phi(v_k) = 0$, and determine only the values $\phi(v_i)$, for $1 \leq i \leq k-1$. This corresponds to finding the matrix M described in Section 3.3. Rather than find M directly, we can first find the matrix $Q = MM^T$, and then obtain M from Q, using (for example) Cholesky decomposition [29].

To show that Vol$(S) \geq W$, for some $W > 0$, the matrix Q has to satisfy several constraints.

1. The matrix Q is positive definite. (Q is equal to MM^T.)
2. $Det(Q) \geq ((k-1)!W)^2$. (We use Evol$(\phi(S)) = \sqrt{Det(Q)}/(k-1)!$.)
3. For $1 \leq i \leq k-1$, $Q_{ii} \leq (d(v_i, v_k))^2$. ($\phi$ needs to be a contraction, and $Q_{ii} = |\phi(v_i) - \phi(v_k)|^2$, because v_k is translated to the origin.)
4. For $1 \leq i < j \leq k-1$, $Q_{ii} + Q_{jj} - 2Q_{ij} \leq (d(v_i, v_j))^2$. (Observe that $Q_{ii} + Q_{jj} - 2Q_{ij} = |\phi(v_i) - \phi(v_j)|^2$.)

Finding a matrix Q that satisfies the above conditions can be solved using the Ellipsoid algorithm with a polynomial time separation oracle [13]. For any matrix X, we can check in polynomial time whether one of the above constraints is violated. Finding a separating hyperplane for constraints 3 and 4 is straightforward, as the constraints themselves define a separating hyperplane. A separating hyperplane for constraint 1 is standard in positive definite programming: for a matrix X that is not positive definite, find an eigenvector y that corresponds to a nonpositive eigenvalue of X, and add the constraint $y^T Q y > 0$. The following separating hyperplane for constraint 2 was suggested by Lovasz. The set of order $k-1$ matrices satisfying the constraint $Det(Q) \geq ((k-1)!W)^2$ is not convex, but with the additional requirement that Q is positive definite the set of feasible solutions is convex. Given a matrix X that satisfies constraints 1, 3, and 4, but not constraint 2, trace$(QX^{-1}) > k-1$ is a linear constraint that can serve as a separating hyperplane. This can be seen as follows. Observe that trace$(XX^{-1}) = k-1$, whereas for any positive definite matrix Q with $Det(Q) \geq ((k-1)!W)^2$, $Det(QX^{-1}) > 1$, because $Det(X^{-1}) > ((k-1)!W)^{-2}$. As X is positive definite, so is X^{-1}. The matrix QX^{-1} is a product of two positive definite matrices. As such, it need not be symmetric, but its eigenvalues are real and positive. (More generally, let $C = AB$ where A and B are positive definite matrices. Let $Cv = \lambda v$ for some complex eigenvector v and eigenvalue λ. Then $(Bv)^*ABv$ is real and positive, where * denotes conjugate transpose. As $B = B^*$, we have that $(Bv)^*ABv = \lambda v^*Bv$. But as v^*Bv is positive real, so is λ.) As trace(QX^{-1}) is the sum of the eigenvalues and $Det(QX^{-1}) > 1$ is their product, trace$(QX^{-1}) > k-1$ follows from the fact that for positive numbers, arithmetic mean is not smaller than geometric mean. □

The above arguments can be generalized to prove Theorem 6, finding an embedding with near optimal distortion η.

Proof:(**Theorem 6**.) We need to find a matrix M_n that has n rows and L columns, where row i gives the embedding of vertex v_i in R^L. Rather than find M_n directly, we shall find $Q = M_n(M_n)^T$, and obtain M_n by Choleskey decomposition. Use the Ellipsoid algorithm with a separation oracle to find an n by n matrix Q with the following properties:

1. The matrix Q is positive semidefinite.
2. For $1 \leq i < j \leq n$, $Q_{ii} + Q_{jj} - 2Q_{ij} \leq (d(v_i, v_j))^2$. (The embedding is a contraction.)
3. For $1 \leq i \leq n$, $n^4 \leq Q_{ii} \leq n^4 + n^2$. (Here we imagine that vertices are embedded in a hyperplane at distance n^2 from the origin.)
4. For every k-subset S, $Det(Q_S) \geq (\mathrm{Vol}(S))^2 n^4 (k!)^2/\eta^{2(k-1)}$, where Q_S is the k by k submatrix of Q in which rows and columns are indexed by members of S. (Recall that $\sqrt{Det(Q_S)}/(k-1)!$ is the volume of the simplex that is defined by the origin and the points $\phi(S)$. This is a factor of n^2/k larger than $\mathrm{Evol}(\phi(S))$, when the origin is at distance n^2 from the hyperplane containing $\phi(S)$.)

Property 4 gives a set of $\binom{n}{k}$ volume constraints, and this number is polynomial in n for any fixed k. If any of the above constraints is violated, a separating hyperplane can be found in polynomial time in a way similar to that for computing $\mathrm{Vol}(S)$, described above. □

Remark: When k is not constant, it is not possible to list out explicitly in polynomial time all $\binom{n}{k}$ volume constraints. In this case we do not know if it is possible to find in polynomial time a volume respecting embedding with distortion $(1+\epsilon)\eta$ for *every* k-subset. However, it is possible to use an algorithm as in Theorem 6 to find in random polynomial time an embedding that has distortion $(1+\epsilon)\eta$ for most k-subsets. Given an embedding

ϕ, we select a random k-set S and compute $\mathrm{Vol}(S)/\mathrm{Evol}(\phi(S))$. If this ratio is larger than $((1+\epsilon)\eta)^{k-1}$, we have found a violated volume constraint and can do one more iteration of the Ellipsoid algorithm to obtain a new embedding. If this ratio is smaller than $((1+\epsilon)\eta)^{k-1}$ for sufficiently many randomly sampled k-subsets, we stop. We note that the random k-subset S need not be chosen from the uniform distribution. Depending on the intended application, we may choose S from a different distribution. For example, in the context of running an algorithm similar to PLOGBAND, it is better to choose S only from those subsets whose projection happens to be small. Using such an approach, the Ellipsoid algorithm can replace the random subsets embedding in algorithm PLOGBAND even when $k = \log n$.

5.3 The random subsets embedding

Let $S \subset V$ be a set of k vertices in graph G. For vertex v and a nonnegative integer t, let $\rho(v, t)$ be the distance (in G) between v and its tth nearest neighbor in G. We set $\rho(v, 0) = 0$, and note that $\rho(v, 1) = 1$, as the nearest neighbor of v is at distance one from v. We define an *elimination order* for the vertices of S. For vertex $v_i \in S$, let $d_i = \min_{v_j \in S\setminus\{v_i\}}[d(v_i, v_j)]$. Pick the minimum t such that there is some $v_i \in S$ with $\rho(v_i, t) \geq d_i/2$. For this t, pick arbitrarily one $v_i \in S$ for which $\rho(v_i, t) \geq d_i/2$ is realized, and rename v_i as u_k (the kth vertex in the elimination order) and $d_i/2$ as q_k (the kth *critical distance*). Remove $u_k = v_i$ from the set S to obtain a set S'. Now iterate the above process with S', eliminating vertex by vertex, renaming the vertices as $u_{k-1}, \ldots, u_1$, and obtaining critical distances $q_{k-1}, \ldots, q_2$ (there is no critical distance associated with u_1, so we may take $q_1 = 1$).

Proposition 14 *For the elimination order as above, $q_2 \ldots q_k \geq Tvol(S)/2^{k-1}$.*

Proof: For each vertex v_i that is eliminated from S, its corresponding d_i represents the distance of it to a vertex that was not yet removed from S. Hence the product of the d_i is the product of edge lengths of some spanning tree of S, and the proof follows from the definition of Tvol. $\square$

We use the above elimination order in the proof of Theorem 5, that the random subsets embedding is volume respecting.

Proof: (**Theorem 5.**) Let S be a set of k vertices in the graph G, and let ϕ be a random subsets embedding. Let K be the simplex given by $\phi(S)$, and recall that K is a $(k-1)$-dimensional simplex lying in a space of dimension L, where $L \simeq (k \log n)^2 \log k$. We bound $\mathrm{Evol}(K)$ from below. Our method of doing this is as follows.

We consider a sequence of simplexes, $K_1, K_2, \ldots, K_k$, where K_i is the $(i-1)$-dimensional simplex composed of the vertices $u_1, \ldots, u_i$ in the elimination order for S. Let $\mathrm{Evol}(K_1) = 1$.

Lemma 15 *For arbitrary S and $2 \leq i \leq k$, and using the notation above (including that of the elimination order). With probability at least $1 - n^{-k}$ (over the choice of ϕ)*

$$Evol(K_i) \geq \frac{2q_i Evol(K_{i-1})}{(i-1)\eta}$$

where η is as in Theorem 5.

The proof of Lemma 15 appears shortly. We now show that Lemma 15 implies Theorem 5.

From Lemma 15 it follows that with probability at least $1 - (k-1)/n^k$, $\mathrm{Evol}(K) \geq 2^{k-1} q_2 \ldots q_k/(k-1)!\eta^{k-1}$. ¿From Proposition 14 it follows that then $\mathrm{Evol}(K) \geq \mathrm{Tvol}(S)/(k-1)!\eta^{k-1}$. ¿From Theorem 3 it follows that then $\mathrm{Evol}(K) \geq \mathrm{Vol}(S)/\eta^{k-1}$. As this event happens with probability at least $1-(k-1)/n^k$, and there are only $\binom{n}{k} \leq n^k/2(k-1)$ different choices of S, we have that with probability at least $1/2$ the volume distortion for every S is simultaneously less than $\eta^{(k-1)/(k-1)} = \eta$. □

We now prove Lemma 15.

Proof:(Lemma 15.) The inequality $\mathrm{Evol}(K_i) \geq \frac{2q_i \mathrm{Evol}(K_{i-1})}{(i-1)\eta}$ is equivalent to showing that the distance from $\phi(u_i)$ to the $(i-2)$-dimensional subspace of K_{i-1} is at least $2q_i/\eta$. Let x be an arbitrary point in this subspace, expressed as $x = \sum_{j=1}^{i-1} \lambda_j \phi(u_j)$, where $\sum_{j=1}^{i-1} \lambda_j = 1$. The main technical content of our proof is in showing the following proposition:

Proposition 16 *With probability at least $1 - n^{-3k}$, $|\phi(u_i) - x| \geq 3q_i/\eta$.*

The proof of Proposition 16 will be given shortly. We now show that Proposition 16 implies Lemma 15.

Recall that ϕ is a contraction. Then $d(u_i, u_1) < n$ implies $|\phi(u_i) - \phi(u_1)| < n$. Hence the point y closest to $\phi(u_i)$ in the subspace of K_{i-1} is at distance at most n from $\phi(u_1)$. Consider a large ball of radius n centered at u_1 and lying in the $(k-2)$-dimensional subspace of K_{i-1}. In the large ball we can pack at most $(\eta n)^{i-2} < n^{2k}$ small balls of radius $1/\eta$ (when n is large enough so that $\eta < n$). Now rather than consider one x, let the center of each small ball be a potential x. From Proposition 16 it follows that with probability at least $1 - n^{-k}$, the distance of $\phi(u_i)$ from each such center is at least $3q_i/\eta$. Hence for an arbitrary point y in the large ball,

$$|\phi(u_i) - y|^2 \geq (\frac{3q_i}{\eta})^2 - |x-y|^2 \geq (\frac{3q_i}{\eta})^2 - (\frac{2}{\eta})^2 \geq 5(\frac{q_i}{\eta})^2$$

implying $|\phi(u_i) - y| \geq \frac{2q_i}{\eta}$, as desired. □

It now remains to prove Proposition 16.

Proof: (**Proposition 16.**) So as to simplify notation and not carry around the $\sqrt{L}$ factor, we ignore the scaling factor of $1/\sqrt{L}$ in the definition of ϕ. The index i is reserved throughout the proof to be that of the vertex u_i referred to in the statement of Proposition 16. We shall be sloppy in our treatment of constants (e.g., not making distinctions between log and ln), so that minor technical errors are left in the proof. However, it should be clear to the reader that these errors are easily corrected by a more pedantic treatment of the constants.

For the proof of the proposition, we would like to view the location of x as fixed, and the embedding of u_i is random. Then $\phi(u_i)$ is not likely to be located at the same location as x. A quantitative version of this argument would say that $\phi(u_i)$ is "far away" from x. However, the location of x itself also depends on ϕ, as $x = \sum_{j=1}^{i-1} \lambda_j \phi(u_j)$. Hence care is needed in bounding the correlation between the locations of x and $\phi(u_i)$.

To determine the location of x, it suffices to first determine $\phi(u_1), \ldots, \phi(u_{i-1})$. We shall show that even after $\phi(u_1), \ldots, \phi(u_{i-1})$ are determined, there is sufficient randomness left in ϕ to send u_i away from x.

Recall that ϕ specifies L coordinates. For a coordinate l, $\phi_l(v_j) = d(S_l, v_j)$, where $S_l \subset V$ is a random subset of vertices, each chosen independently with probability p_l. In determining S_l, we shall think of the vertices of V as being listed in the order $w_1, w_2, \ldots$, where for $z = a(i-1) + j$ (where $1 \leq j \leq i-1$), w_z is the ath closest vertex to u_j (where the 0th closest vertex is u_j itself). We shall go over the sorted list of vertices one by one, and for each vertex independently decide with probability p_l whether it belongs to S_l or not. (The list may contain each vertex $i-1$ times. Inclusion in S_l is determined at the first time the vertex is encountered. The rest of its occurrences are ignored.) To determine ϕ_l completely, we must go over the whole sorted list. However, as a thought experiment, we shall only partially determine ϕ_l. The rule is that we stop going over the list once all the values $\phi_l(u_1), \ldots, \phi_l(u_{i-1})$ are determined. This will happen when for each j there is some a such that $w_{a(i-1)+j} \in S_l$.

Repeating this process for every l we get $\phi(u_j)$ for every $1 \leq j \leq i-1$, and hence the location of x is fixed. The value of $\phi(u_i)$ is not necessarily fixed by this process, because the vertices left unscanned in the lists may well be those closest to u_i, and their inclusion/noninclusion in the respective S_ls effects the value of $\phi_l(u_i)$.

Recall the value of t associated with u_i in the definition of the elimination order. There are t vertices (including u_i) at distance less than q_i from u_i, all other vertices are at distance at least q_i from u_i, and the t vertices nearest u_i are disjoint from the t vertices nearest u_j, for all $j < i$.

Recall that with each coordinate of ϕ we associated a value t_l. We shall call a coordinate *eligible* if $t_l \leq t$. Our lower bound on the distance between $\phi(u_i)$ and x is based only on eligible coordinates. An eligible coordinate l is *good* if when we partially determined ϕ_l we did not look at any of the t vertices nearest u_i.

Claim 1 *For an eligible coordinate, the probability that it is good is at least* $1/2$.

Proof: Observe that an eligible coordinate is good if for every $j \le i-1$, at least one of the t vertices nearest u_j was put in S_l. For a specific u_j, the probability of this happening is at least $1-(1-p_l)^t \ge 1-(1-p_l)^{t_l} \ge 1-1/2k$. Hence the probability that this fails for at least one u_j is at most $(i-1)/2k < 1/2$. $\square$

Recall that in the random subsets embeddings we arranged the coordinates in blocks sharing the same value of t_l. As each such block has $\Omega(k^2 \log n)$ coordinates, Chernoff bounds imply that with probability at least $1-n^{-4k}$, at least one third of the coordinates of an eligible block are good.

Consider now l for which the respective t_l is the largest satisfying $t_l \le t$.

Claim 2 *For a good coordinate with t_l as above, $\phi_l(u_i) \ge q_i$ with probability at least $1/5k$, where the probability is taken over the random completion of the partial ϕ_l to a complete ϕ_l.*

Proof: Observe that $t \le t_l(1+1/\log k)$. Hence the probability that $\phi_l(u_i) \ge q_i$ is

$$(1-p_l)^t \ge (1-p_l)^{t_l(1+1/\log k)} \ge (1/2k)^{1+1/\log k} \ge 1/5k$$

proving the claim. $\square$

Now our analysis breaks into two subcases. We start with the simpler of the two.

Case 1: $Pr[\phi_l(u_i) \le q_i/2] > 1/5k$**.** Together with Claim 2 this implies that for every value b, $Pr[|\phi_l(u_i) - b| \ge q_i/4] \ge 1/5k$. Hence regardless of the value of x in coordinate l, $\phi_l(u_i)$ has probability at least $1/5k$ of differing from it by at least $q_i/4$. Call a coordinate in which this is realized *lucky*. Then standard Chernoff bounds show that of the $\Omega(k^2 \log n)$ coordinates in the block, $\Omega(k \log n)$ will be lucky (with probability at least $1-n^{-4k}$). In this case we shall have (after also putting back the missing factor of $1/\sqrt{L}$) that

$$|\phi(u_i) - x| = \Omega(q_i\sqrt{k \log n/L}) = \Omega(q_i/\sqrt{k \log k \log n})$$

Case 2: $Pr[\phi_l(u_i) \le q_i/2] \le 1/5k$**.** The analysis of Case 2 is more complicated than that of Case 1. In order to get best possible bounds, we shall consider several blocks of coordinated, rather than just one block. In fact, we shall consider several *superblocks*. The lower bound on the distance between u_i and x will be obtained as a sum of distances in several superblocks.

Let t' denote the number of vertices of V that are at distance up to $q_i/2$ from u_k. Then by the assumption of Case 2, $(1-p_l)^{t'} \ge 1-1/5k$. As $(1-p_l)^{t_l} \le 1/2k$, this implies that $t' < t_l/k \log k$.

Recall that the coordinates of ϕ are partitioned into $\log n \log k$ blocks. Merge every consecutive $\log k$ blocks into one superblock. Hence we now have $\log n$ superblocks of coordinates, and within each superblock, the values of the respective t_ls are within a factor of two of each other. Let B be a superblock for which the largest of its $\log k$ defining probabilities, which we denote by p_h, satisfies $(1-p_h)^{t'} \leq 1/2$, for t' as above.

Claim 3 *Every coordinate of a block B as above is eligible.*

Proof: Let p_s be the smallest defining probability in B, and let t_s be its respective t-value. We need to show that $t_s < t$. Let t_h be associated with p_h. Then $t_s \leq 2t_h$, because the different t values within a block differ by a factor of at most two. Also, $t_h < 2t' \log k$, because $(1-p_h)^{t_h} = 1/2k$ and $(1-p_h)^{t'} \leq 1/2$. Above we have observed that $t' < t_l/k \log k$. Also, we know that $t_l < t$. Putting everything together we get that $t_s \leq 4t/k$, proving the claim (for $k \geq 4$). (The case $k = 3$ can be proved by paying more attention to the constants in our construction. Details are omitted.) $\Box$

Chernoff bounds and Claim 1 imply that a constant fraction of the eligible coordinates are good. This holds for every eligible superblock. Denote the eligible superblocks by $B_1, B_2, \ldots$, in order of decreasing defining probabilities. For an eligible superblock B_m, let p_m be its smallest defining probability, and let t_m be the largest value satisfying $(1-p_m)^{t_m} \geq 1/2$. Then associated with the sequence of superblocks we have a sequence of values $t_1, t_2, \ldots$, where the last of these values $t_m \geq t'/2$.

Claim 4 *For some universal constant δ and for every defining probability p in a superblock B_m we have $(1-p)^{t_m} > \delta$ and $1-(1-p)^{t_{m-1}} > \delta$.*

Proof: Let t_p be such that $(1-p)^{t_p} = 1/2$. Then $t_{m-1} \leq t_p \leq t_m$. Because $t_{m-1} \simeq t_m/2$ it follows that $(1-p)^{t_m} \geq 1/4$ and $(1-p)^{t_{m-1}} \leq \sqrt{2}/2$. $\Box$

Let d_m be the distance of the t_mth vertex from u_i. Then for every coordinate l in block B_m, and for every possible value x_l of x in coordinate l, we have that $Pr[|\phi_l(u_i) - x_l| \geq (d_m - d_{m-1})/2] \geq \delta$. Chernoff bounds imply that indeed for a constant fraction of the coordinates in a superblock (i.e., $\Omega(L/\log n)$ coordinates), $|\phi_l(u_i) - x_l| \geq (d_m - d_{m-1})/2$. The l_1 distance between x and $\phi(u_i)$ is lower bounded by a telescoping sum over all eligible superblocks, giving $\Omega(q_i L/\log n)$ with overwhelming probability. For l_2 norm, this implies $|\phi(u_i) - x| \geq \Omega(q_i\sqrt{L}/\log n)$. Scaling back by $1/\sqrt{L}$ completes the treatment of **Case 2**.

Combining Cases 1 and 2 completes the proof of Proposition 16. $\Box$

5.4 Volumes of random projections

We first prove proposition 7, which gives bounds on the length of the projection of a random unit vector.

Proof: (**Proposition 7.**) Our proof follows the techniques in [10].

Let A_i denote the surface area of the unit sphere in R^i. Then for $L > k > 0$,

$$A_L = \int_0^{\pi/2} A_k(\cos\theta)^{k-1} A_{L-k}(\sin\theta)^{L-k-1} d\theta$$

The above formula (for which a formal proof is given in [10]) can be understood intuitively as follows. Consider R^L as the product of two orthogonal subspaces, R^k and R^{L-k}. Consider an arbitrary unit vector r in R^L (i.e., a point on the unit sphere in R^L). Then for $0 \le \theta \le \pi/2$, if its projection r_k on R^L is $\cos\theta$, then its projection r_{L-k} on R^{L-k} is $\sin\theta$. Now fix one value of θ, and consider all unit vectors $r \in R^L$ with $|r_k| = \cos\theta$. In R^k they form a ball of radius $\cos\theta$, and hence of surface area $A_k(\cos\theta)^{k-1}$. In R^{L-k} they form a ball of radius $\sin\theta$, and hence of surface area $A_{L-k}(\sin\theta)^{L-k-1}$. Hence their contribution to the area of the unit ball in R^L is $A_k(\cos\theta)^{k-1}A_{L-k}(\sin\theta)^{L-k-1}$. Integrating over all values of θ we get the area of the unit ball.

Using similar arguments we have:

$$Pr_r[|r_k| \le \cos\alpha] = \frac{1}{A_L}\int_\alpha^{\pi/2} A_k(\cos\theta)^{k-1} A_{L-k}(\sin\theta)^{L-k-1} d\theta$$

We now set $\cos\alpha = \epsilon\sqrt{k/L}$ for some small $\epsilon < 1$ and upper bound the above integral. The range of the integration satisfies

$$\pi/2 - \alpha \le (\pi/2)\cos\alpha = O(\epsilon\sqrt{k/L})$$

The integrand (ignoring the constants) satisfies

$$(\cos\theta)^{k-1}(\sin\theta)^{L-k-1} \le (\cos\theta)^{k-1} \le (\cos\alpha)^{k-1} = (\epsilon\sqrt{k/L})^{k-1}$$

Using $A_i = 2\pi^{i/2}/\Gamma(i/2)$, we have that the constant term satisfies

$$\frac{A_k A_{L-k}}{A_L} = \frac{2\Gamma(L/2)}{\Gamma(k/2)\Gamma((L-k)/2)}$$

From the definition of the Γ function it follows that $\Gamma(L/2)/\Gamma((L-k)/2) = O((L/2)^{k/2})$. Also, $\Gamma(k/2) \ge ((k/2-1)/e)^{k/2-1}$. (If $k \le 2$, then replace the last inequality by $\Gamma(k/2) \ge 1/2$.) It follows that

$$\frac{A_k A_{L-k}}{A_L} = k(O(L/k))^{k/2}$$

Putting all the above together we obtain:

$$Pr_r[|r_k| \le \epsilon\sqrt{k}/\sqrt{L}] \le O(\epsilon\sqrt{k/L})(\epsilon\sqrt{k/L})^{k-1} k(O(L/k))^{k/2} \le O(k\epsilon^k) \le (\beta\epsilon)^k$$

for some universal $\beta > 0$. This completes the proof of the first part of the proposition.

Inequalities such as the second part of the proposition (upper bound on the probability of a large projection when $k = 1$) are well known. See Chapter 2 in [23], for example. A nice proof is given in [3], which we repeat here for completeness.

We choose a random point on the unit sphere in R^L and want to bound the probability of the event that its projection on R^1 is larger than $\sqrt{c/L}$. This event implies that the projection on the orthogonal subspace R^{L-1} is of length at most $\sqrt{1-c/L}$. Cut from the unit ball the two caps that correspond to points where this event happens, and join them together. The convex body that we get is entirely contained in a ball of radius $\sqrt{1-c/L}$, and has smaller surface area than this bounding ball. As surface area of a ball is proportional to its radius taken to the power of $L-1$, the measure of points giving a large projection on R^1 is at most

$$(\sqrt{1-c/L})^{L-1} = (1-c/L)^{(L-1)/2} \leq (1/e)^{c(L-1)/2L} \leq (1/e)^{c/4}$$

When L is large, the above probability tends to $(1/e)^{c/2}$. $\square$

We can now prove Theorem 9, which bounds the length of the projection of a convex body on a random line.

Proof: (**Theorem 9**.) Let R^k be the subspace of R^L that contains K. The projection r' of the random unit vector r on R^k remains spherically symmetric, though r' is no longer a unit vector. We observe that if the length of r' is ℓ, then by scaling K by a factor of ℓ we can use Lemma 8 to conclude that $Pr[P(K,r') \leq 1] < (O(1)/k)^{k/2}/(\ell)^k \mathrm{Evol}(K)$. The expected length of r' is $\Theta(\sqrt{k/L})$. Observe that when $|r'| \geq \sqrt{k/L}$ then

$$Pr[P(K,r') \leq 1] \leq \left(\sqrt{L/k}\right)^k \frac{(O(1))^{k/2}}{k^{k/2}\mathrm{Evol}(K)} \leq \frac{(O(1)L)^{k/2}}{k^k \mathrm{Evol}(K)}$$

as desired. Hence it remains to deal with the case that $|r'| < \sqrt{k/L}$. The probability that the length of r' is a factor of $c > 1$ below $\sqrt{k/L}$ is $(\beta/c)^k$, for some $\beta > 0$ (see Proposition 7). Consider all possible lengths ℓ for r', starting from $\sqrt{k/L}$, and decreasing by factors of two until $(O(1)/k)^{k/2}/(\ell)^k \mathrm{Evol}(K) = 1$. There are at most $\log(\mathrm{Evol}(K))/k + \log k$ such values of ℓ. For each value of ℓ,

$$Pr[(\ell/2 \leq |r'| \leq \ell) \wedge (P(K,r') \leq 1)]$$

$$\leq (\beta\ell\sqrt{L/k})^k \frac{(O(1)/k)^{k/2}}{(\ell/2)^k \mathrm{Evol}(K)} = (O(1)L)^{k/2}/k^k \mathrm{Evol}(K)$$

Summing up over $\log(\mathrm{Evol}(K))/k + \log k$ values of ℓ, and noting that terms such as $1/k$ and $\log k$ are suppressed by the term $(O(1))^{k/2}$, the theorem is proved. $\square$

5.5 Tree volume versus local density

To prove Theorem 10 we use the following lemma:

Lemma 17 *For any finite metric space (S, d) with $S = \{v_1, \ldots, v_k\}$,*

$$\frac{2^{k-1}}{Tvol(S)} \leq \sum_{\pi} \frac{1}{d(v_{\pi(1)}, v_{\pi(2)}) \ldots d(v_{\pi(k-1)}, v_{\pi(k)})}$$

where the summation is taken over all permutations $\pi : \{1, \ldots, k\} \to \{1, \ldots, k\}$.

Proof: The proof is by induction on k. For $k = 2$, there is only one spanning tree and two possible permutations π and hence the inequality reads $\frac{2}{\text{Tvol}(S)} \leq \frac{2}{d(v_1,v_2)}$, which is clearly true. Assume now that the lemma holds for all sets of up to k points, and prove for a set $S = \{v_1, \ldots v_{k+1}\}$. Consider a minimum spanning tree T on S and remove a leaf (w.l.o.g., assume that it is v_{k+1}) from this tree to obtain a minimum spanning tree T' on $S' = \{v_1, \ldots, v_k\}$. Clearly, $\text{Tvol}(S) \geq \text{Tvol}(S') \min_i[d(v_{k+1}, v_i)]$. So to prove the lemma, it suffices to show that

$$\min_i[d(v_{k+1}, v_i)] \sum_{\sigma} \frac{1}{d(v_{\sigma(1)}, v_{\sigma(2)}) \ldots d(v_{\sigma(k)}, v_{\sigma(k+1)})}$$
$$\geq \sum_{\pi} \frac{2}{d(v_{\pi(1)}, v_{\pi(2)}) \ldots d(v_{\pi(k-1)}, v_{\pi(k)})}$$

where π is a permutation on k elements, and σ is a permutation on $k+1$ elements. Every permutation π leads to $k+1$ different permutations σ, depending on where v_{k+1} is inserted in π. Let π be (w.l.o.g.) the identity permutation, let $X = \frac{1}{d(v_1,v_2)\ldots d(v_{k-1},v_k)}$, and let v_i be the vertex for which $\min_i[d(v_{k+1}, v_i)]$ is achieved. Then it suffices to show that

$$X\left(\frac{1}{d(v_{k+1}, v_1)} + \frac{1}{d(v_{k+1}, v_k)} + \sum_{j=1}^{k-1} \frac{d(v_j, v_{j+1})}{d(v_{k+1}, v_j)d(v_{k+1}, v_{j+1})}\right)$$
$$\geq \frac{2X}{d(v_{k+1}, v_i)}$$

The above inequality can be proved by using the triangle inequality to bound each term $d(v_j, v_{j+1})$ from below. When $j < i$ use $d(v_j, v_{j+1}) \geq d(v_j, v_{k+1}) - d(v_{k+1}, v_{j+1})$, and when $j \geq i$ use $d(v_j, v_{j+1}) \geq -d(v_j, v_{k+1}) + d(v_{k+1}, v_{j+1})$. The inequality is a result of two telescopic sums. $\Box$

We now prove Theorem 10, that $\sum_{\{S \subset V, |S|=k\}} \frac{1}{\text{Tvol}(S)} \leq n(D(1 + \ln(n/D)))^{k-1}$.

Proof:(**Theorem 10.**) Rather than bound $\sum_{\{S\subset V, |S|=k\}} \frac{1}{\mathrm{Tvol}(S)}$, we shall show

$$\sum_S \sum_\pi \frac{1}{d(v_{\pi(1)}, v_{\pi(2)}) \dots d(v_{\pi(k-1)}, v_{\pi(k)})}$$
$$\leq n(2D(1+\ln(n/D)))^{k-1}$$

and then the theorem follows from Lemma 17.

To prove the above inequality we use induction on k, the cardinality of S. The base case for $k=1$ reads $n \leq n$. For the inductive step, observe that when going from k to $k+1$ the right hand side of the inequality increases by a factor of $(2D(1+\ln(n/D)))$. To see by how much the left hand side increases, consider an arbitrary k-set S and an arbitrary permutation π. When increasing to $k+1$, this contributes $n-k$ terms to the left hand side, one for each possible choice of v_{k+1}. Quantitatively, the increase is by a factor of $\sum_{v_{k+1}\in V\setminus S} \frac{1}{d(v_k, v_{k+1})}$. As there are at most $2dD$ vertices at distance d from v_k, we can bound $\sum_{v_{k+1}\in V\setminus S} \frac{1}{d(v_k, v_{k+1})}$ from above by $2D\sum_{i=1}^{n/2D} 1/i$. $\square$

5.6 A generalization of Pythagoras' theorem

Let K be a simplex of k points in R^L, where $L > k$. What can we say about the volume of K by considering $(k-1)$ dimensional projections of it? This question came up in the proof of Theorem 5, given in previous versions of the paper. The current proof of Theorem 5 does not make use of the contents of this section. Anticipating future applications, this section is left in the paper.

To obtain some intuition, consider the following simple case in which there is a k-dimensional cube K in R^{2k}, the projection of K on the first k coordinates gives a cube K_1 with sidelength 1, and the projection of K on the last k coordinates gives a cube K_2 with sidelength 1. Then $\mathrm{Evol}(K_1) = \mathrm{Evol}(K_2) = 1$. From Pythagoras' theorem one can infer that the side length of K is $\sqrt{2}$, and hence the volume is $(\sqrt{2})^k$. This shows that the volume of K is very much larger (a factor of $2^{k/2}$) than each of its two projections, and suggests the equality $(\mathrm{Evol}(K))^{2/k} = (\mathrm{Evol}(K_1))^{2/k} + (\mathrm{Evol}(K_2))^{2/k}$. It can be verified that whenever K, K_1 and K_2 are k-dimensional cubes then the above equality is correct. We note that when k is 1, a cube becomes a line segment, and the equality becomes Pythagoras' theorem.

Motivated by the above discussion, we have the following lemma.

Lemma 18 *Let $T_1, \dots, T_t$ be t subspace R^L, where each T_i is of $(k-1)$ dimensions and every two subspaces T_i and T_j are orthogonal. Let K be a k-point simplex in R^L, and let K_i be the projection of K on subspace T_i, for $1 \leq i \leq t$. Then*

$$(Evol(K))^{\frac{2}{k-1}} \geq (Evol(K_1))^{\frac{2}{k-1}} + \dots + (Evol(K_t))^{\frac{2}{k-1}}$$

Proof: Denote the vertices of K by $p_1, \ldots, p_k$. For $2 \leq i \leq k$, let q_i be the point closest to p_i (i.e., the orthogonal projection of p_i) on the $(i-2)$-dimensional plane containing the points $p_1, \ldots, p_{i-1}$. This implies for example that $|p_3 - q_3|$ is the height of the triangle (p_1, p_2, p_3) (where (p_1, p_2) is the base). Then

$$\mathrm{Evol}(K) = \frac{1}{(k-1)!}\Pi_{i=2}^{k}|p_i - q_i|$$

Let p_i^j denote the projection of p_i on T_j and let q_i^j denote the projection of q_i on T_j. Then

$$\mathrm{Evol}(K_j) \leq \frac{1}{(k-1)!}\Pi_{i=2}^{k}|p_i^j - q_i^j|$$

with equality only if for every i, the vector $p_i^j - q_i^j$ is orthogonal to the $(i-2)$ dimensional plane containing $(p_1^j, \ldots, p_{i-1}^j)$.

We note that from Pythagoras' theorem

$$|p_i - q_i|^2 \geq \sum_{j=1}^{t} |p_i^j - q_i^j|^2$$

with equality only if the line $(p_i - q_i)$ has zero projection outside of the subspaces T_j.

Let $\alpha_i^j = |p_i^j - q_i^j|/|p_i - q_i|$. Then

$$(\mathrm{Evol}(K_j)/\mathrm{Evol}(K))^{2/(k-1)} = (\Pi_{i=2}^{k}\alpha_i^j)^{2/(k-1)} \leq \frac{1}{k-1}\sum_{i=2}^{k}(\alpha_i^j)^2$$

with equality of geometric and arithmetic mean only if for every fixed j, all α_i^j are the same as i ranges from 2 to k. Using $\sum_{j=1}^{t}(\alpha_i^j)^2 \leq 1$ we obtain

$$\sum_{j=1}^{t}(\mathrm{Evol}(K_j)/\mathrm{Evol}(K))^{2/(k-1)} \leq \sum_{j=1}^{t}\left(\frac{1}{k-1}\sum_{i=2}^{k}(\alpha_i^j)^2\right) \leq \frac{1}{k-1}\sum_{i=2}^{k}\sum_{j=1}^{t}(\alpha_i^j)^2 \leq 1$$

implying

$$(\mathrm{Evol}(K))^{2/(k-1)} \geq (\mathrm{Evol}(K_1))^{2/(k-1)} + \ldots + (\mathrm{Evol}(K_t))^{2/(k-1)}$$

□

Remark: Exact equality in Lemma 18 is obtained only if K is contained in the subspace that is the direct sum of the T_is (namely, the projection of K outside of the T_is is a single point), and all K_i are similar (can be obtained from K by scaling). This follows by checking which conditions make all inequalities into equalities in the proof above.

6 Discussion

Volume respecting embeddings were developed as a tool for obtaining algorithms that approximate the bandwidth within polylogarithmic factors. They can also be used for some other vertex ordering problems (see [3]), and for problems related to VLSI layout (see [32]). Preliminary experiments were made so as to compare the bandwidth output by algorithm PLOGBAND to the bandwidth output by known heuristics that order the vertices by breadth first search numbering. On most graphs tried, known heuristics outperformed algorithm PLOGBAND. However, on some graphs designed specifically so as to fool the known heuristics, algorithm PLOGBAND found linear orders that had bandwidth a factor of ten smaller than the bandwidth found by known heuristics. It is fair to point out that on graphs with thousands of vertices, algorithm PLOGBAND is order of magnitudes slower than known heuristics, as it runs breadth first search several hundred times, rather than just once. (A more detailed report may perhaps be available at a later date.)

Many questions remain open. We present several open problems on which we believe it is possible to make progress.

- Determine the best possible approximation ratio for the bandwidth. A similar problem, of a combinatorial rather than computational nature, is to determine the worst possible ratio between bandwidth and local density. As an intermediate step, one may try to shave off a factor of $\Omega(\sqrt{\log n})$ from our analysis for algorithm PLOGBAND.

- Determine the smallest possible distortion for k-volume respecting embeddings. For the random subsets embedding we proved that the distortion is $O(\log n)$ whenever $k = O(\log n / \log\log n)$. For $k = \log n$, the distortion proved here is $O(\log n \sqrt{\log\log n})$, and it would be nice to get rid of the $\sqrt{\log\log n}$ factor. For large values of k, the term $\sqrt{k}$ begins to dominate our bounds. Rao [26] outlines an embedding for which one can prove bounds on the distortion that are better than those proved in the current manuscript, whenever $k = \Omega((\log n)^2)$ (and also for smaller values of k, for some special families of graphs).

- Let P_n be a path on n vertices. Embed P_n in R^3 via a contraction that minimizes area distortion. Can the area distortion be polylogarithmic in n? In other words, is there a $(3, O((\log n)^c))$-volume respecting embedding of P_n in R^3?

Acknowledgements

This work was motivated by open questions presented at the Dagstuhl Seminar 9734 on Combinatorial Approximation Algorithms, August 18 to 22, 1997. Discussions with Reuven Bar-Yehuda, Moses Chariker, Robert Krauthgamer, Michael Langberg, Moni Naor, Satish

Rao, Yuri Rabinovich, Gideon Schechtman and Oded Schramm helped in bringing this work to its current form. The separation oracle used in the proof of Theorem 4 was suggested by an oracle – Laci Lovasz. Algorithm PLOGBAND was programmed and experimented with by Tzvika Hartman. Part of this work was done while the author was visiting Compaq Systems Research Center, Palo Alto, California.

References

[1] S. Assman, G. Peck, M. Syslo, J Zak. "The bandwidth of caterpillars with hairs of length 1 and 2". *SIAM J. Algebraic Discrete Methods* 2 (1981), 387–393.

[2] G. Blache, M. Karpinski, J. Wirtgen. "On approximation intractability of the bandwidth problem". Manuscript, 1998.

[3] A. Blum, G. Konjevod, R. Ravi, S. Vempala. "Semidefinite relaxations for minimum bandwidth and other vertex-ordering problems". *Proc. of 30th STOC, 1998, 100–105.*

[4] H. Bodlaender, M. Fellows, M. Hallet. "Beyond NP-completeness for problems of bounded width: hardness for the W Hierarchy". *Proc. of 26th STOC, 1994, 449–458.*

[5] J. Bourgain. "On Lipschitz embedding of finite metric spaces in Hilbert space". *Israel J. Math. 52 (1985) 46–52.*

[6] P. Chinn, J. Chvatalova, A. Dewdney, N. Gibbs. "The bandwidth problem for graphs and matrices – a survey". *Journal of Graph Theory, 6 (1982), 223–254.*

[7] F. Chung, P. Seymour. "Graphs with small bandwidth and cutwidth". *Discrete Mathematics 75 (1989) 113–119.*

[8] J. Chvatalova. On the bandwidth problem for graphs, Ph.D. dissertation, University of Waterloo, 1980.

[9] E. Cuthill, J. McKee. "Reducing the bandwidth of sparse symmetric matrices". *ACM National Conference Proceedings, 24, 1969, 157–172.*

[10] P. Frankl, H. Maehara. "The Johnson-Lindenstrauss lemma and the sphericity of some graphs". *Journal of Combinatorial Theory, Series B 44, 355–362 (1988).*

[11] M. Garey, R. Graham, D. Johnson, D. Knuth. "Complexity results for bandwidth minimization". *SIAM J. Appl. Math. 34 (1978), 477-495.*

[12] A. George, J. Liu. Computer Solution of Large Positive Definite Systems. Prentice-Hall, 1981.

[13] M. Grotschel, L. Lovasz, A. Schrijver. Geometric Algorithms and Combinatorial Optimization. Springer-Verlag, Berlin, 1987.

[14] E. Gurari, I. Sudborough. "Improved dynamic programming algorithms for bandwidth minimization and the min-cut linear arrangement problems". *J. Algorithms, 5 (1984), 531–546.*

[15] J. Haralambides, F. Makedon, B. Monien. "Bandwidth minimization: an approximation algorithm for caterpillars". *Math Systems Theory 24, 169-177 (1991).*

[16] M. Karpinski, J. Wirtgen, A. Zelikovsky. "An approximation algorithm for the bandwidth problem on dense graphs". *ECCC TR97-017.*

[17] D. Kleitman, R. Vohra. "Computing the bandwidth of interval graphs". *SIAM J. Discrete Math.*, 3 (1990), 373–375.

[18] T. Kloks, D. Kratsch, H. Muller. "Approximating the bandwidth for asteroidal triple-free graphs". In *Algorithms – ESA '95, Paul Spirakis (Ed.), Lecture Notes in Computer Science 979, 434–447, Springer.*

[19] D. Knuth. *The Art of Computer Programming, Vol. 2, Seminumerical Algorithms.* Addison Wesley, Second Edition, 1981.

[20] P. M. Gruber, C. G. Lekkerkerker. Geometry of Numbers. Second edition, 1987. North Holland Mathematical Library, Volume 37.

[21] J. Lagarias. "Point lattices". In Handbook of Combinatorics, Volume 1, R. Graham, M. Grotschel, L. Lovasz (editors), MIT Press, 1995.

[22] N. Linial, E. London, Y. Rabinovich. "The geometry of graphs and some of its algorithmic applications". *Combinatorica 15 (2) (1995) 215–245.*

[23] V. Milman, G. Schechtman. Asymptotic Theory of Finite Dimensional Normed Spaces. *Lecture Notes in Mathematics, 1200, Springer-Verlag, 1986.*

[24] B. Monien. "The bandwidth-minimization problem for caterpillars with hairlength 3 is NP-complete". *SIAM J. Algebraic Discrete Methods* 7 (1986), 505–512.

[25] C. Papadimitriou. "The NP-completeness of the bandwidth minimization problem". *Computing 16 (1976), 263–270.*

[26] S. Rao. "Small distortion and volume preserving embeddings for Planar and Euclidean metrics". *Proc. of 15th Annual ACM Symposium on Computational Geometry*, Miami Beach, Florida, 1999.

[27] L. A. Santalo. Integral Geometry and Geometric Probability. Encyclopedia of Mathematics and its Applications, Volume 1, Addison Wesley, 1976.

[28] J. Saxe. "Dynamic programming algorithms for recognizing small-bandwidth graphs in polynomial time". *SIAM Journal on Algebraic Methods 1 (1980), 363–369.*

[29] G. Strang. Linear Algebra and its Applications, Third Edition. Saunders College Publishing, Harcourt Brace Jovanovich College Publishers, 1988.

[30] J. Turner. "On the probable performance of heuristics for bandwidth minimization". *SIAM J. Comput., 15, (1986), 561–580.*

[31] W. Unger. "The complexity of the approximation of the bandwidth problem". In *Proc. 39th Annual IEEE Symposium on Foundations of Computer Science*, 1998, 82–91.

[32] S. Vempala. "Random projection: a new approach to VLSI layout". In *Proc. 39th Annual IEEE Symposium on Foundations of Computer Science*, 1998, 389–395.