The geometry of graphs and some of its algorithmic applications *

Nathan Linial †
Institute of Computer Science
Hebrew University
Jerusalem 91904
Israel
nati@cs.huji.ac.il

Eran London
Institute of Mathematics
Hebrew University
Jerusalem 91904
Israel
eran@math.huji.ac.il

Yuri Rabinovich
Department of Computer Science
University of Toronto
Toronto, Ontario M5S 1A4
Canada
yuri@cs.toronto.edu

Abstract

In this paper we explore some implications of viewing graphs as *geometric* objects. This approach offers a new perspective on a number of graph–theoretic and algorithmic problems. There are several ways to model graphs geometrically and our main concern here is with geometric representations that respect the *metric* of the (possibly weighted) graph. Given a graph G we map its vertices to a normed space in an attempt to (i) Keep down the dimension of the host space and (ii) Guarantee a small *distortion*, i.e., make sure that distances between vertices in G closely match the distances between their geometric images.

In this paper we develop efficient algorithms for embedding graphs low–dimensionally with a small distortion. Further algorithmic applications include:

- A simple, unified approach to a number of problems on multicommodity flows, including the Leighton–Rao Theorem [36] and some of its extensions. We solve an open question in this area, showing that the max–flow vs. min–cut gap in the k-commodities problem is $O(\log k)$. Our new deterministic polynomial–time algorithm finds a (nearly tight) cut meeting this bound.
- For graphs embeddable in low–dimensional spaces with a small distortion, we can find low–diameter decompositions (in the sense of [6] and [41]). The parameters of the decomposition depend only on the dimension and the distortion and not on the size of the graph.
- In graphs embedded this way, small balanced *separators* can be found efficiently.

Given faithful low–dimensional representations of statistical data, it is possible to obtain meaningful and efficient *clustering*. This is one of the most basic tasks in pattern–recognition. For the (mostly heuristic) methods used in the practice of pattern–recognition, see [19], especially chapter 6.

Our studies of multicommodity flows also imply that every embedding of (the metric of) an n-vertex, constant–degree expander into a Euclidean space (of any dimension) has distortion $\Omega(\log n)$. This result is tight, and closes a gap left open by Bourgain [11].

1 Introduction

Many combinatorial and algorithmic problems concern either directly or implicitly the distance, or *metric* on the vertices of a possibly weighted graph. It is, therefore, natural to look for connections

*AMS Classification: 05C, 46B, 52B, 68Q. A preliminary version of this paper appeared in FOCS 35 (1994), 577-591.

†Supported in part by grants from the Israeli Academy of Sciences, the Binational Science Foundation Israel–USA and the Niedersachsen–Israel Research Program.

between such questions and classical geometric structures. The approach taken here is to model graph metrics by mapping the vertices to a real normed spaces in such a way that the distance between any two vertices is close to the distance between their geometric images. Embeddings are sought so as to minimize (i) The *dimension* of the host space, and (ii) The *distortion*, i.e., the extent to which the combinatorial and geometric metrics disagree. Specifically, we ask:

1. How small can the dimension and the distortion be for given graphs? Is it computationally feasible to find good embeddings?
2. Which graph–algorithmic problems are easier for graphs with favorable (low–dimensional, small distortion) embeddings? What are the combinatorial implications of such embeddings?
3. The above discussion extends to embeddings of general finite metric spaces. What are the algorithmic and combinatorial implications in this more general context?

Here are some of the answers provided in this paper:

1. We develop a randomized polynomial–time algorithm that embeds every finite metric space in a Euclidean space with minimum distortion. Bourgain [11] had shown that every n-point metric can be embedded in an $O(\log n)$-dimensional Euclidean space with a logarithmic distortion, and our algorithm finds embeddings which are at least as good. Better embeddings are provided for particular families of graphs.
2. The geometry of graphs offers an excellent framework for studying multicommodity flow problems. In particular, our methods yield a very short and conceptual proof for several theorems that relate the value of the multicommodity flow in a network to the minimum cut in it (e.g., the Leighton–Rao Theorem [36]). An open question in this area is solved by the very same argument.

   A low–diameter decomposition of a graph G is a (not necessarily proper) coloring of the vertices with few colors, so that all monochromatic connected subgraphs have small diameters. In [41] the precise tradeoff between the number of colors and the diameter was found for general graphs. In particular, both parameters can be made logarithmic in the order of the graph, which is, in general, essentially tight. Here we establish a better tradeoff that depends only on the dimension and the distortion at which the graph is embeddable, not on the number of vertices. For many algorithmic applications of low–diameter decompositions, see [6] and [15].

   Graphs embeddable in a d-dimensional normed space with a small distortion have balanced separators of only $O(d \cdot n^{1-\frac{1}{d}})$ vertices. If the embedding is given, such separators can be found in random polynomial time. For closely related work see [47].
3. *Clustering* of statistical data calls for grouping sample points in sets (clusters), so that distances between members of the same cluster are significantly smaller than between points from distinct clusters. If data points reside in a high–dimensional Euclidean space, or even worse, if distances between points do not conform with any norm, then clustering is notoriously problematic. Our algorithms provide one of the few non–heuristic approaches to this fundamental problem in pattern–recognition.

Many problems in this area remain open, some of which appear in Section 8.

A preliminary version of this paper appeared in FOCS 35 (1994), 577-591. The present paper is (hopefully) better written and contains a number of new results, especially in Sections 4 and 8. In the first version of this article (January 1994) we conjectured that the present methods are applicable to the study of multicommodity flow problems. This plan materialized in July 1994 and lead to the revised version of September 1994.

2 Related work and overview of results

During the past few years a new area of research has been emerging, an area which may be aptly called *the geometry of graphs*. This is the study of graphs from a geometric perspective. A comprehensive survey of this fascinating area is well beyond our scope, so we restrict ourselves to a few brief remarks and examples. Geometric models of graphs can be classified as either (i) Topological models, (ii) Adjacency models, or (iii) Metric models. The *topological* approach is mainly concerned with graph planarity and embeddability of graphs on other 2-dimensional manifolds (see [33] for a recent survey). It also deals with 3-dimensional embeddings of a graph, mostly in the context of knot theory. (See Welsh's book [58] for some of these developments.)

Particularly fascinating is the way in which geometric/topological considerations come up in the theory of forbidden minors [54]. The *double cycle cover* conjecture (surveyed in [29]) says that in every bridgeless graph there is a collection of cycles which cover every edge exactly twice. It is not hard to see that this is equivalent to the statement that for every graph G there is a 2-dimensional manifold on which G can be embedded so that every edge is incident with two distinct cells (2-dimensional faces) of the embedding.

In an adjacency model, the geometry respects only the relation of adjacency/nonadjacency of vertices, but not necessarily the actual distance. A prime example for this approach is the Koebe–Andreev–Thurston Theorem (see [35], [2], [3], and [57]) that every planar graph is the contact graph of openly disjoint planar discs. Higher–dimensional results in the same vein appear in [20], [21], [47], and [53].

Another noteworthy adjacency model calls for mapping the vertices of a graph to a Euclidean unit sphere where adjacent vertices are to be mapped to *remote* points on the sphere. This approach, initiated in [43], has recently found further interesting algorithmic applications (see [25] and [32]).

Let X be a set of k vertices in a graph G. An embedding φ of G in $\mathbb{R}^{k-1}$ is *X-convex* if the vertices of X are mapped to the vertices of a $(k-1)$-dimensional simplex and $\varphi(x) \in \mathrm{conv}\{\varphi(y) | y \text{ is a neighbor of } x\}$ for every $x \notin X$. It is shown in [39] that G is k-connected iff for every such X there is an X-convex φ under which all points are in general position. This result can be put together with *orthogonal embeddings* (see [43]), an adjacency model where adjacent vertices are mapped to perpendicular vectors. The combination of these two approaches is discussed in [44].

We now turn to problems where graph metrics have a role. A *low–diameter decomposition* of a graph (see [41], [6], [15]) is a (not necessarily proper) coloring of the vertices with $\le k$ colors, so that every connected monochromatic subgraph has diameter $\le D$. In [41] it is shown that an n-vertex graph has such a decomposition whenever $k^D > n$ and $D^k > n$, the conditions being essentially tight. The first condition is necessary, e.g., in decomposing triangulations of Euclidean spaces. The second condition is necessary for expander graphs. The similarity between this dichotomy and that of positive/negative curvature in geometry still awaits a satisfactory explanation.

Low–dimensional models for finite metric spaces have previously been studied mostly by functional analysts (see [4], [8], [11], [18], [27], [30], [31], and [46]). Study of graph metrics has also led to the notion of spanners [49] and hop–sets [14]. Local–global considerations, which are commonplace in geometry, arose for graphs as well (see [38] and the references therein, [40], and the recent [52]).

All the referenced work notwithstanding, this short discussion leaves out large amounts of relevant work, for example, on embedding graphs in particular families of graphs such as d-dimensional lattices, cubes, squashed cubes etc. (see [26], [59]). Particularly relevant are notions of dimension that emerge from such considerations, see, e.g., chapter 5 in [7]. The possibility of embedding graphs in spaces other than Euclidean and spheric geometry is very appealing, and hardly anything has been done in this direction (but see [56]). We have also said nothing about modeling geometric objects with graphs, which is a related vast area.

To initiate our technical discussion, recall that a *norm* $\|\cdot\|$ associates a real number $\|x\|$ with every point x in real d-space, where (i) $\|x\| \geq 0$, with equality only if $x = 0$, (ii) $\|\lambda x\| = |\lambda|\,\|x\|$, for every $x \in \mathbb{R}^d$ and every real λ, and (iii) For every $x, y \in \mathbb{R}^d$, $\|x+y\| \leq \|x\| + \|y\|$. The metric associated with the norm $\|\cdot\|$ is $d(x,y) = \|x-y\|$. Particularly important are l_p norms: $\|(x_1,\ldots,x_n)\|_p = (\sum |x_i|^p)^{1/p}$. We denote $\mathbb{R}^n$ equipped with l_p norm by l_p^n. *Euclidean* norm is, of course, l_2.

An *isometry* is a mapping φ from a metric [1] space (X,d) to a metric space (Y,ρ) which preserves distances, i.e., $\rho(\varphi(x),\varphi(y)) = d(x,y)$ for all $x,y \in X$. Given a connected graph $G = (V,E)$, we often denote by G also the metric space induced on V by distances in the graph. Isometries are often too restricted and much flexibility is gained by allowing the embedding to distort the metric somewhat. This leads to the main definition in this article:

Definition 2.1: Metric Dimension: For a finite metric space (X,d) and $c \geq 1$, define $\dim_c(X)$ to be the least dimension of a real normed space $(Y,\|\cdot\|)$, such that there is an embedding φ of X into Y where every two points $x_1, x_2 \in X$ satisfy

$$d(x_1,x_2) \geq \|\varphi(x_1) - \varphi(x_2)\| \geq \frac{1}{c} \cdot d(x_1,x_2).$$

Such an embedding is said to have *distortion* $\leq c$. The *(isometric) dimension* of X is defined as $\dim(X) = \dim_1(X)$. ∎

We recall (Lemma 5.1) that $\dim(X) \leq n$ for every n-point metric space (X,d), whence this definition makes sense.

We list some of the main findings on isometric and near–isometric embeddings. Unless stated otherwise, n is always the number of vertices in the graph G or the number of points in a finite metric space (X,d).

The least distortion with which X may be embedded in l_p (of any dimension) is denoted by $c_p(X)$.

Finding good embeddings

1. In Theorem 3.2 we show:
   There is a deterministic polynomial time algorithm that embeds every n-point metric space (X,d) in a Euclidean space with distortion arbitrarily close to the optimum $c_2(X)$. By [11], $c_2(X) = O(\log n)$.
   In random polynomial time X may be embedded in $l_2^{O(\log n)}$ with distortion $\leq (1+\epsilon) \cdot c_2(X)$ (for any $\epsilon > 0$).
   For every $1 \leq p \leq 2$, X may be embedded in $l_p^{O(\log n)}$ with distortion $O(c_2(X))$. Such an embedding may be found in random polynomial time. The same distortion can also be attained in deterministic polynomial time, but the dimension is $O(n^2)$.
   In random polynomial time X may be also embedded in $l_p^{O(\log^2 n)}$ for every $p > 2$ with distortion $O(\log n)$.
   If X is the metric of a constant–degree expander graph, then $c_p(X) = \Omega(\log n)$ for every $2 \geq p \geq 1$.

[1] Technically, we are discussing semi–metrics, as we allow two distinct points to have distance zero.

2. For every metric space (X, d) on n points there is a $\gamma > 0$ such that the metric $\gamma \cdot d$ can be embedded in a *Hamming* metric with an $O(\log n)$ distortion (Corollary 3.8). The bound is tight.

Structural consequences

1. The gap between the maximum flow and the minimum cut in a multicommodity flow problem is majorized by the least distortion with which a particular metric can be embedded in l_1. This metric is defined via the Linear Programming dual of a program for the maximum flow. This is the basis for a unified and simple proof to a number of old and new results on multicommodity flows (Section 4).

2. Low–dimensional graphs have small separators: A d-dimensional graph G has a set S of $O(d \cdot n^{1-\frac{1}{d}})$ vertices which separates the graph, so that no component of $G \setminus S$ has more than $\left(1 - \frac{1}{d+1} + o(1)\right) n$ vertices (Theorem 6.1).

3. The vertices of any d-dimensional graph can be $(d+1)$-colored so that each monochromatic connected component has diameter $\leq 2d^2$ (compare with [41]).
   They can also be covered by "patches" so that each r-sphere (r – any positive integer) in the graph is contained in at least one patch, while no vertex is covered more than $d+1$ times. The diameter of each such patch is $\leq (6d+2)dr$ (compare with [6]). Moreover, the patches may be $(d+1)$-colored so that equally colored patches are at distance ≥ 2.
   That is, there exist low–diameter decompositions with parameters depending on the dimension alone (Theorem 7.1).

4. Low–dimensional graphs have large diameter, $\mathrm{diam}(G) \geq n^{O(\frac{1}{\dim(G)})}$ (Lemma 5.6).

Algorithmic consequences

1. Near–tight cuts for multicommodity flow problems can be found in deterministic polynomial time (Section 4).

2. Given an isometric embedding of G in d dimensions, a balanced separator of size $O(d \cdot n^{1-\frac{1}{d}})$ can be found in random polynomial time (Theorem 6.1).

3. Low–dimensional, small–distortion representation of statistical data offers a new approach to clustering which is a key problem in pattern–recognition (Section 3.2).

Isometric dimensions

- All trees have dimension $O(\log n)$. The bound is tight (Theorem 5.3).
- $\dim(K_n) = \lceil \log_2 n \rceil$ (Proposition 5.4). (This result essentially goes back to [16].)
- $\sum_{i=1}^{k} \lceil \log_2 n_i \rceil \geq \dim(K_{n_1,\ldots,n_k}) \geq \sum_{i=1}^{k} \lceil \log_2 n_i \rceil - 1$, where $K_{n_1,\ldots,n_k}$ is the complete k-partite graph, and $n_i \geq 2$ (Theorem 5.8).
- For cycles: $\dim(2m\text{-Cycle}) = m$, and $m+1 \geq \dim((2m+1)\text{-Cycle}) \geq \frac{m}{2} - 1$. Consequently, $\dim(G) \geq \frac{\mathrm{girth}(G)}{4} - 1$. However, $\dim_{\frac{\pi}{2}}(n\text{-Cycle}) = 2$ (Proposition 5.10 and Remark 5.11).
- $\dim(d\text{-Cube}) = d$ (Corollary 5.12).

3 Low–distortion low–dimensional embeddings

3.1 Good embeddings

We start by quoting:

Theorem 3.1: (**Johnson–Lindenstrauss [30]**, *see also [20]*) *Any set of n points in a Euclidean space can be mapped to $\mathbb{R}^t$ where $t = O(\frac{\log n}{\epsilon^2})$ with distortion $\leq 1+\epsilon$ in the distances. Such a mapping may be found in random polynomial time.*

Proof: (Rough sketch) Although the original paper does not consider computational issues, the proof is algorithmic. Namely, it is shown that an orthogonal projection of the original space (which can be assumed to be n-dimensional) on a random t-dimensional subspace, almost surely produces the desired mapping. This is because the length of the image of a unit vector under a random projection is strongly concentrated around $\sqrt{t/n}$. ∎

Our general results on near–isometric embeddings are summarized in the following theorem:

Theorem 3.2:

1. (**Bourgain [11]**, *see also [31], [46]*). *Every n-point metric space (X,d) can be embedded in an $O(\log n)$-dimensional Euclidean space with an $O(\log n)$ distortion.*

2. *There is a deterministic polynomial–time algorithm that for every $\epsilon > 0$ embeds (X,d) in a Euclidean space with distortion $< c_2(X)+\epsilon$. In random polynomial–time (X,d) may be embedded in $l_2^{O(\log n)}$ with distortion $O(c_2(X))$ (By Claim 1, $c_2(X) = O(\log n)$).*

3. *In random polynomial–time (X,d) may be embedded in $l_p^{O(\log n)}$ (for any $1 \leq p \leq 2$), with distortion $O(c_2(X))$.*

4. *In deterministic polynomial time (X,d) may be embedded in $l_p^{O(n^2)}$ (for any $1 \leq p \leq 2$), with distortion $O(c_2(X))$.*

5. *In random polynomial–time (X,d) may be embedded in $l_p^{O(\log^2 n)}$ (for any $p > 2$), with distortion $O(\log n)$.*

6. *Every embedding of an n-vertex constant–degree expander into an l_p space, $2 \geq p \geq 1$, of* any *dimension, has distortion $\Omega(\log n)$.*

Proof: Claim 1 appears mostly for future reference, but is seen to be an immediate corollary of Claims 3 and 5 and Theorem 3.1.

To prove Claim 2, let the rows of the matrix M be the images of the points of X under a distortion-c embedding in some Euclidean space. Let further $A = MM^t$. Clearly, A is positive semidefinite, and for every $i \neq j$,

$$\frac{1}{c^2} d_{i,j}^2 \leq a_{i,i} + a_{j,j} - 2a_{i,j} \leq d_{i,j}^2.$$

As in [43], [25], and [32] the ellipsoid algorithm can be invoked to find an ϵ–approximation of c in polynomial time.

The dimension is reduced to $O(\log n)$ by applying Theorem 3.1.

To prove Claim 3, use Claim 2 and recall that for any m, l_2^m may be embedded in l_p^{2m}, for every $1 \leq p \leq 2$, with constant distortion (see [50], chapter 6). This embedding may be found in random polynomial–time. In fact, it is enough to map l_2^m isometrically into a random m-dimensional subspace of l_p^{2m}.

For the deterministic algorithm in Claim 4, start with the algorithm in Claim 2. Once an optimal embedding into Euclidean space (of dimension at most n) is found, proceed with the following explicit embedding of l_2^m to $l_p^{O(m^2)}$ (see [9]):

Lemma 3.3: *Let $\mathbb{F} \subset \{-1,+1\}^m$ be a 4-wise independent family of vectors, let $x \in \mathbb{R}^m$ and $1 \le p \le 2$. Then,*

$$\|x\|_2 \le \left(\frac{1}{|\mathbb{F}|} \sum_{\epsilon_i \in \mathbb{F}} | < x, \epsilon_i > |^p \right)^{\frac{1}{p}} \le \sqrt{3}\|x\|_2.$$

There are explicit constructions of such families $\mathbb{F}$ with $O(m^2)$ vectors. ∎

Therefore, having embedded X into n-dimensional Euclidean space, map this space to $l_p^{O(n^2)}$ via

$$x \in l_2^n \to \frac{1}{|\mathbb{F}|^{\frac{1}{p}}}(< x, \epsilon_1 >, < x, \epsilon_2 >, \ldots, < x, \epsilon_{O(n^2)} >) \in l_p^{O(n^2)}.$$

By the Lemma 3.3 this embedding adds only a constant distortion.

We now turn to the Claim 5.

The overall structure of the algorithm follows the general scheme of Bourgain's proof: For each cardinality $k < n$ which is a power of 2, randomly pick $O(\log n)$ sets $A \subseteq V(G)$ of cardinality k. Map every vertex x to the vector $(d(x, A))$ (where $d(x, A) = \min\{d(x, y) | y \in A\}$) with one coordinate for each A selected. It is shown that this mapping to $l_1^{O(\log^2 n)}$ has almost surely an $O(\log n)$ distortion.

We now turn to the actual proof. Let $B(x, \rho) = \{y \in X | d(x, y) \le \rho\}$ and $\mathring{B}(x, \rho) = \{y \in X | d(x, y) < \rho\}$ denote the closed and open balls of radius ρ centered at x. Consider two points $x \ne y \in X$. Let $\rho_0 = 0$, and let ρ_t be the least radius ρ for which both $|B(x, \rho)| \ge 2^t$ and $|B(y, \rho)| \ge 2^t$. We define ρ_t as long as $\rho_t < \frac{1}{4}d(x, y)$, and let $\hat{t}$ be the largest such index. Also let $\rho_{\hat{t}+1} = \frac{d(x,y)}{4}$. Observe that $B(y, \rho_j)$ and $B(x, \rho_i)$ are always disjoint.

Notice that $A \cap \mathring{B}(x, \rho_t) = \emptyset \iff d(x, A) \ge \rho_t$, and $A \cap B(y, \rho_{t-1}) \ne \emptyset \iff d(y, A) \le \rho_{t-1}$. Therefore if both conditions hold, then $|d(y, A) - d(x, A)| \ge \rho_t - \rho_{t-1}$.

Let us assume that $|\mathring{B}(x, \rho_t)| < 2^t$ (otherwise we argue for y). On the other hand $|B(y, \rho_{t-1})| \ge 2^{t-1}$. Therefore, a random set of size $\Theta(\frac{n}{2^t})$ has a constant probability to both intersect $B(y, \rho_{t-1})$ and miss $\mathring{B}(x, \rho_t)$.

We randomly select $q = O(\log n)$ sets of cardinality 2^l, the least nonnegative power of two that is $\ge \frac{n}{2^{t+1}}$. Then, with high probability, for each pair x, y, at least $\frac{q}{10}$ of the sets chosen will intersect $B(y, \rho_{t-1})$ and miss $\mathring{B}(x, \rho_t)$.

Note that the same applies to $\rho_{\hat{t}+1} - \rho_{\hat{t}}$, since again we wish to miss a set of size $< 2^{\hat{t}+1}$, and to intersect a set of size $\ge 2^{\hat{t}}$.

Therefore for almost every choice of $A_1, \ldots, A_q$:

$$\sum_{i=1}^{q} |d(x, A_i) - d(y, A_i)| \ge \frac{\log n \cdot (\rho_t - \rho_{t-1})}{10}.$$

We do this now for every $l = 1, 2, \ldots, \lfloor \log n \rfloor$ and obtain $Q = O(\log^2 n)$ sets for which:

$$\sum_{i=1}^{Q} |d(x,A_i)-d(y,A_i)| \geq \frac{\log n}{10} \cdot \sum_{i=1}^{\hat{t}+1}(\rho_i - \rho_{i-1}) = \frac{\log n}{10} \cdot \rho_{\hat{t}+1} \geq \log n \cdot \frac{d(x,y)}{40}.$$

The reverse inequality is obtained by observing that $|d(x,A_i)-d(y,A_i)| \leq d(x,y)$ for every A_i, whence:

$$\sum_{i=1}^{Q} |d(x,A_i)-d(y,A_i)| \leq C \cdot \log^2 n \cdot d(x,y).$$

Thus, the mapping which sends every vertex x to the vector $(\frac{d(x,A_i)}{Q} \mid i = 1,2,\ldots,Q)$, embeds G in an $O(\log^2 n)$-dimensional space endowed with the l_1-norm and has distortion $O(\log n)$. In fact, for every $p \geq 1$, a proper normalization of this embedding satisfies the same statement with respect to the l_p norm.

The only modification is that now x gets mapped to $(\frac{d(x,A_i)}{Q^{1/p}} \mid i = 1,2,\ldots,Q)$: Let $\tau(x,y)$ denote the l_p distance between the image of x and the image of y, i.e.,

$$\tau(x,y) = \left(\frac{\sum_{i=1}^{Q} |d(x,A_i)-d(y,A_i)|^p}{Q}\right)^{\frac{1}{p}}.$$

But for every i, $|d(x,A_i)-d(y,A_i)| \leq d(x,y)$ whence $\tau(x,y) \leq d(x,y)$. On the other hand:

$$d(x,y) \leq \frac{40 \cdot \sum_{i=1}^{Q} |d(x,A_i)-d(y,A_i)|}{\log n} = O\left(\log n \cdot \frac{\sum_{i=1}^{Q} |d(x,A_i)-d(y,A_i)|}{Q}\right) \leq O(\log n \cdot \tau(x,y)),$$

by the monotonicity of p-th moment averages.

The proof of Claim 6 is deferred to Proposition 4.2. ∎

The following technical corollary will be needed in Section 4.

Corollary 3.4:

1. *Let (X,d) be a finite metric space and let $Y \subseteq X$. There exists a randomized polynomial–time algorithm that finds an embedding $\varphi : X \to l_1^{O(\log^2 |Y|)}$, so that $d(x,y) \geq \|\varphi(x)-\varphi(y)\|$ for every $x,y \in X$, and if x,y are both in Y, then also $\|\varphi(x)-\varphi(y)\| \geq \Omega(\frac{1}{\log |Y|}) \cdot d(x,y)$.*

2. *Let (X,d) be a finite metric space and $\{(s_i,t_i) \mid i = 1,2,\ldots,k\} \in X \times X$. There exists a deterministic polynomial time algorithm that finds an embedding $\varphi : X \to l_1^{O(n^2)}$, so that $d(x,y) \geq \|\varphi(x)-\varphi(y)\|$ for every $x,y \in X$, and $\|\varphi(s_i)-\varphi(t_i)\| \geq \Omega(\frac{1}{\log k}) \cdot d(s_i,t_i)$ for every $i = 1,2,\ldots,k$.*

Proof: The proof of Claim 1 follows the proof of Claim 5 of Theorem 3.2:
Map x to $(d(x,A_i) \mid i = 1,2,\ldots,O(\log^2 |Y|))$, where the sets A_i are randomly chosen subsets of Y.

The second part of the Corollary follows the proof of Claim 2 in Theorem 3.2. Next, apply the embedding to $l_1^{O(n^2)}$ mentioned in proving Claim 4. ∎

Claim 2 of Theorem 3.2 leads to a characterization of $c_2(X)$, the least distortion with which (X,d) may be embedded in a Euclidean space of any dimension.

The acronym $PSD = PSD_n$ denotes the cone of real symmetric $n \times n$ positive semidefinite matrices.

Corollary 3.5: *An n-point metric space (X,d) may be embedded in a Euclidean space with distortion $\leq c$ iff for every matrix $Q \in PSD$ which satisfies $Q \cdot \vec{1} = \vec{0}$ the following inequality holds:*

$$\sum_{q_{i,j}>0} q_{i,j} \cdot d_{i,j}^2 + c^2 \sum_{q_{i,j}<0} q_{i,j} \cdot d_{i,j}^2 \leq 0.$$

In particular, this inequality holds for any metric d and any $Q \in PSD$ that satisfies $Q \cdot \vec{1} = \vec{0}$ with $c = O(\log n)$.

Proof: We retain the notation used in the proof of Claim 2. As observed in that proof, X has such an embedding iff there is matrix $A \in PSD_n$ with

$$d_{i,j}^2 \leq a_{i,i} + a_{j,j} - 2a_{i,j} \leq c^2 d_{i,j}^2.$$

for all i and j. The consistency of such a system of conditions can be decided using convex programming duality. The analysis is facilitated since we know a complete (infinite) list of linear inequalities that define the cone PSD, namely:

Proposition 3.6: *A matrix P belongs to PSD_n iff it satisfies $\sum_{i,j} p_{i,j} \cdot y_{i,j} \geq 0$ for every $Y \in PSD_n$.*

Proof: If Y has rank one, say $y_{i,j} =< v_i, v_j >$ for some vector v, then $\sum p_{i,j} \cdot y_{i,j} = vPv^t$. Therefore, the condition that $\sum_{i,j} p_{i,j} \cdot y_{i,j} \geq 0$ for every $Y \in PSD$ implies that $P \in PSD$. On the other hand, if $P \in PSD$, then for the same reason, $\sum_{i,j} p_{i,j} \cdot y_{i,j} \geq 0$ whenever Y has rank one. The general case follows, since every $Y \in PSD$ is a nonnegative combination of matrices of rank one. ∎

We return to the proof of Corollary 3.5. By convex programming duality and the fact that a nonnegative combination of PSD matrices is again in PSD, it follows that the conclusion of the Corollary is incorrect, i.e., no such matrix A exists, iff there is a matrix $Q \in PSD$ such that the inequality $\sum_{i,j} a_{i,j} \cdot q_{i,j} \geq 0$ contradicts some nonnegative combination of the inequalities

$$J_{i,j}: \quad c^2 d_{i,j}^2 \geq a_{i,i} + a_{j,j} - 2a_{i,j} \geq d_{i,j}^2.$$

For the combination in question to be a contradiction, all terms $a_{i,j}$ have to be eliminated. In particular, inequality $J_{i,j}$ must be multiplied by $q_{i,j}/2$. More accurately, if $q_{i,j} \geq 0$, the right inequality in $J_{i,j}$ is taken, and otherwise we take the left part of $J_{i,j}$ multiplied by $-q_{i,j}/2$. The term involving $a_{i,i}$ disappears only if its coefficient $\sum_j q_{i,j}$ vanishes, i.e., only if $Q \cdot \vec{1} = \vec{0}$ holds. The Corollary follows. ∎

Remark 3.7: The case $c = 1$, i.e., the characterization of metric spaces that isometrically embed in Euclidean space is classical (see [10]). ∎

A *Hamming space* is a metric space which consists of $\{0,1\}$ vectors of the same length, equipped with the Hamming metric.

Corollary 3.8: *For every n-point metric space (X,d) there is a Hamming space (Y,ρ) and a $\gamma > 0$ such that (X,d) can be embedded in $(Y, \gamma \cdot \rho)$ with distortion $O(\log n)$. The bound is tight.*

Proof: It suffices to find a constant–distortion embedding for every finite subset of l_1^m into a Hamming space. Nothing is changed by adding the same number to all values at some coordinate. Also recall that multiplication by a fixed factor is allowed. Therefore, at the cost of an arbitrarily small distortion the entries of the i-th coordinate may be assumed to be integers, the smallest of which is 0. If the largest i-th coordinate is r, this coordinate is replaced by r new ones, where $x_i = s$ is replaced by s new coordinates of 1 followed by $r - s$ coordinates of 0. This latter step adds no distortion, being an isometry into a Hamming space.
Again, the tightness result follows from Proposition 4.2. ∎

3.2 Applications to clustering

It is a recurring situation in pattern–recognition, where one is given a large number of sample points, which are believed to fall into a small number of categories. It is therefore desired to partition the points into a few *clusters* so that points in the same cluster tend to be much closer than points in distinct clusters. When sampling takes place in a low–dimensional Euclidean space, clustering is reasonably easy, but when the dimension is high, or worse still, if the metric is non–Euclidean, reliable clustering is a notoriously difficult problem. See Duda and Hart [19], in particular chapter 6, for a standard reference in this area.

The algorithms we have just described offer a new approach to clustering. These algorithms are currently being practically tested [37] in a project to search for patterns among protein sequences. One pleasing aspect that already emerges is this – the second algorithm in Theorem 3.2 assumes that the distance between any pair of points in the space can be evaluated in a single time unit. In this particular application, the metric space consists of all presently known proteins. Molecular biologists have developed a number of measures to estimate the (functional, structural, evolutionary etc.) distance between protein sequences, and some widely available software packages (e.g., FASTA, BLAST) calculate them. At this writing about 40,000 proteins of average length ca. 350 have already been sequenced. For our purposes, a protein is a word in an alphabet of 20 letters (amino–acids) and it takes about a quarter of a second to compute a single distance according to any of the common metrics, using standard software on a typical workstation. A straightforward implementation of the algorithm is therefore infeasible. The difficulty stems from having to compute $d(x, A)$ for large A's. A closer observation of the proof shows that if we fail to include the coordinates which correspond to large A's, the effect is that the distance between *close* pairs of points (protein sequences) is *reduced* in the mapping. This is definitely a welcome effect in a clustering algorithm, so what seems to be a problem turns out as a kind of a blessing.

4 Multicommodity flows via low-distortion embeddings

We briefly recall some definitions about multicommodity network flows. G is an undirected n-vertex graph, with a *capacity* $C_e \geq 0$ associated with every edge e. There are k pairs of (source–sink) nodes (s_μ, t_μ), and for each such pair a distinct *commodity* and a *demand* $D_\mu \geq 0$ are associated. For simplicity of notation we let $C_{i,j} = 0$ for all non–edges (i, j). As usual, flows have to satisfy conservation of matter, and the total flow through an edge must not exceed the capacity of the edge. Maximal multicommodity flow problems come in a number of flavors, and we concentrate on the following version: Find *maxflow* – the largest λ for which it is possible to simultaneously flow λD_μ between s_μ and t_μ for all μ.

A trivial upper bound on λ is attained by considering *cuts* in G. For $S \subset V$ let $Cap(S)$ be the sum of the capacities of the edges connecting S and $\overline{S}$. Also let $Dem(S)$ be the sum of the demands between source–sink pairs separated by S (i.e., $|S \cap \{s_\mu, t_\mu\}| = 1$). Obviously, $\lambda \leq \frac{Cap(S)}{Dem(S)}$ for every S.

In the case of a single commodity ($k = 1$), the max–flow min–cut theorem is easily seen to say that λ equals the minimum of $\frac{Cap(S)}{Dem(S)}$ over $S \subset V$.

It came as a pleasant surprise when Leighton and Rao showed that in some cases the gap between the max–flow and the min–cut cannot be too big. We show that this gap is bounded by the least distortion with which a certain metric associated with the network can be embedded in l_1. This fact yields a unified approach to Leighton–Rao's [36], and the subsequent [34], [23], and [51]. The result for non–unit demands is new.

Theorem 4.1: [2] *There is a deterministic polynomial–time algorithm which given a network G =*

[2] A randomized version of this Theorem appeared in the preliminary version of this paper and in the paper of Yonatan

(V, E, C) *and the demands for* k *source–sink pairs, finds an* $S \subset V$ *for which*

$$\frac{Cap(S)}{Dem(S)} \leq O(\log k) \cdot \mathit{maxflow}.$$

Proof: The optimal λ is the maximum of a certain linear program. By LP duality (e.g., [23]):

$$\lambda = \min \frac{\sum_{i \neq j} C_{i,j} \cdot d_{i,j}}{\sum_{\mu=1}^{k} D_\mu \cdot d_{s_\mu, t_\mu}},$$

where the minimum is over all metrics d on G. We first apply Claim 2 of Corollary 3.4 to the minimizing metric with $X = V(G)$. The vertices of G are thus mapped to points $\{x_1, \ldots, x_n\}$ in $\mathbb{R}^m$ where $\|x_i - x_j\|_1 \leq d_{i,j}$ for all i, j and $\|x_{s_\mu} - x_{t_\mu}\|_1 \geq \Omega(d_{s_\mu, t_\mu} / \log k)$ for $\mu = 1, \ldots, k$. Therefore,

$$\frac{\sum_{i \neq j} C_{i,j} \cdot \|x_i - x_j\|_1}{\sum_{\mu=1}^{k} D_\mu \cdot \|x_{s_\mu} - x_{t_\mu}\|_1} \leq O(\lambda \cdot \log k).$$

Since we are dealing with the l_1^m metric we may conclude:

$$\frac{\sum_{i \neq j} C_{i,j} \cdot \|x_i - x_j\|_1}{\sum_{\mu=1}^{k} D_\mu \cdot \|x_{s_\mu} - x_{t_\mu}\|_1} = \frac{\sum_{r=1}^{m} \sum_{i \neq j} C_{i,j} \cdot |x_{i,r} - x_{j,r}|}{\sum_{r=1}^{m} \sum_{\mu=1}^{k} D_\mu \cdot |x_{s_\mu, r} - x_{t_\mu, r}|} \geq \min_{1 \leq r \leq m} \frac{\sum_{i \neq j} C_{i,j} \cdot |x_{i,r} - x_{j,r}|}{\sum_{\mu=1}^{k} D_\mu \cdot |x_{s_\mu, r} - x_{t_\mu, r}|}.$$

Let $\tilde{r}$ be an index where the minimum is achieved. We claim that no generality is lost in assuming that all $x_{i,\tilde{r}}$ are in $\{0, 1\}$, whence

$$\frac{\sum_{i \neq j} C_{i,j} \cdot |x_{i,\tilde{r}} - x_{j,\tilde{r}}|}{\sum_{\mu=1}^{k} D_\mu \cdot |x_{s_\mu, \tilde{r}} - x_{t_\mu, \tilde{r}}|} = \frac{Cap(S)}{Dem(S)},$$

where $S = \{i | x_{i,\tilde{r}} = 1\}$.

To justify the assumption $x_{i,\tilde{r}} \in \{0, 1\}$ for every i, we argue that for any real $a_{ij} = a_{ji}$, $b_{ij} = b_{ji}$, $1 \leq i \neq j \leq n$ the minimum of $\frac{\sum_{i \neq j} a_{ij} |z_i - z_j|}{\sum_{i \neq j} b_{ij} |z_i - z_j|}$ (over real z's) can be attained with all $z_i \in \{0, 1\}$. This is shown by a variational argument: If the z's take exactly two values, one value can be replaced by zero and the other by one without affecting the expression. Otherwise, let $s > t > u$ be three values taken by z's. Fixing all other values, and letting t vary over the interval $[u, s]$, the expression is the ratio of two linear functions in t. Therefore, all z's which equal t can be changed to either s or u without increasing the expression. This procedure is applied repeatedly until only two values remain.

The algorithm is a straight implementation of the proofs. First solve the linear program $\lambda = \min\{\sum_{i \neq j} C_{i,j} \cdot d_{i,j}$ under the condition $\sum_{\mu=1}^{k} D_\mu \cdot d_{s_\mu, t_\mu} = 1$, where d is a metric on $G\}$. Approximate the optimizing metric in l_1 as in the second part of Corollary 3.4. Consider the index $\tilde{r}$ which minimizes $\frac{\sum_{i \neq j} C_{i,j} \cdot |x_{i,\tilde{r}} - x_{j,\tilde{r}}|}{\sum_{\mu=1}^{k} D_\mu \cdot |x_{s_\mu, \tilde{r}} - x_{t_\mu, \tilde{r}}|}$. Finally optimize this expression using the above variational procedure to find a near–optimal cut. Note that instead of this last step it suffices to consider only the cuts to an initial segment and a final segment of the one–dimensional embedding.∎

The proof shows that the max–flow min–cut gap is accounted for by the distortion in approximating a certain metric by l_2 norm. In those cases where distortion smaller than $\log k$ will do, better bounds follow for the multicommodity flow problem. For example – suppose the n-point metric space defined by the optimal d is isometrically embeddable in $\mathbb{R}^s$ for some small s. Then (as mentioned in the proof

Aumann and Yuval Rabani [5]. We were recently informed by Naveen Garg [22] that he, too, managed to derandomize a variant of the algorithm that appears in our FOCS '94 paper.

of Theorem 7.1) d may be approximated by l_2^s with distortion $\sqrt{s}$, yielding a better bound than in the general case.

A *cut metric* on n points is defined by picking $S \subset [n]$ and defining $d(x,y) = 1$ if $|S \cap \{x,y\}| = 1$, and as zero otherwise. A simple but useful fact is that a metric on $[n]$ is realizable in l_1 iff it is a nonnegative combination of cut metrics. This fact explains much of what happens in the proof of Theorem 4.1.

Proposition 4.2: *Every embedding of an n-vertex constant–degree expander into an l_p space, $2 \geq p \geq 1$, of any dimension, has distortion $\Omega(\log n)$. The metric space of such a graph cannot be embedded with constant distortion in any normed space of dimension $o(\log^2 n)$.*

Proof: As observed in [36] the max–flow min–cut gap is $\Omega(\log n)$ for the all–pairs, unit–demand flow problem on a constant–degree expander, where all capacities are one. Consider the corresponding expression $\Phi = \frac{\sum_{[i,j]\in E} d_{i,j}}{\sum_{i\neq j} d_{i,j}}$ for certain metrics d. When d is the expander's own metric $\Phi = \Theta(\frac{1}{n \log n})$. On the other hand, the minimum of Φ over one-dimensional metrics d is the expander's min–cut, i.e., $\Omega(\frac{1}{n})$. Consequently $\min \Phi$ over d in l_1 is also $\Omega(\frac{1}{n})$. This gap implies that every embedding of the expander's metric in l_1 (of any dimension) has distortion $\Omega(\log n)$.

The same conclusion holds also for embeddings into l_p for $2 \geq p \geq 1$, because in this range, every finite dimensional l_p space can be embedded in l_1 with a constant distortion ([50], chapter 6).

Every d-dimensional norm may be approximated with distortion $\sqrt{d}$ by an affine image of the Euclidean norm (e.g., [50]). Therefore, an embedding of constant distortion into any $o(\log^2 n)$-dimensional normed space translates into an $o(\log n)$-distortion embedding into l_2, which is impossible. ∎

Two commodities

That max–flow=min–cut for two commodities ([28] and [55]) can be shown as follows: Let d be the metric for which $\lambda = \frac{\sum_{i\neq j} C_{i,j}\cdot d_{i,j}}{\sum_{\mu=1}^{k} D_\mu \cdot d_{s_\mu,t_\mu}}$. Map every vertex x to the point (d_{x,s_1}, d_{x,s_2}) and let D be the l_∞ metric among these points. If we replace d by D, the numerator can only decrease, while the denominator stays unchanged, whence there is no loss in assuming d to be (a restriction of) the l_∞^2 metric. It is not hard to see that the linear mapping $\tau(z_1,z_2) = (\frac{z_1+z_2}{2}, \frac{z_1-z_2}{2})$ satisfies $\|\tau(z_1,z_2) - \tau(w_1,w_2)\|_1 = \|(z_1,z_2)-(w_1,w_2)\|_\infty$. An application of τ thus allows us to assume that the metric d is, in fact (a restriction of) l_1. From here on, the proof of Theorem 4.1 can be followed to derive our claim.

5 Isometries

5.1 General results

All logarithms are to base 2. $G = (V,E)$ is always a connected graph and n is the number of its vertices. Unless otherwise stated, embeddings are into $\mathbb{R}^d$. [3] No distinction is made between a vertex x and its image under the embedding. If G can be embedded in $(X, \|\cdot\|)$ we also say that X *realizes* G.

The *unit ball* of a d-dimensional real normed space is: $\mathbb{B} = \{x \in \mathbb{R}^d \text{ with } \|x\| \leq 1\}$. This is a convex body which is centrally symmetric around the origin. Every centrally symmetric convex body $\mathbb{Q}$ induces a norm, called the Minkowski norm: $\|x\|_{\mathbb{Q}} = \inf\{\lambda > 0 \text{ such that } \frac{x}{\lambda} \in \mathbb{Q}\}$. Thus, normed spaces are denoted either as $(\mathbb{R}^d, \|\cdot\|)$ or as $(\mathbb{R}^d, \mathbb{Q})$. The boundary of $\mathbb{B}$ is $\partial\mathbb{B} = \{x \in \mathbb{R}^d \text{ with } \|x\| = 1\}$. In an isometric embedding of G in $\mathbb{R}^d$ the set $\{\frac{x-y}{d_G(x,y)} | x \neq y \in V\}$ is contained in $\partial\mathbb{B}$. It is not hard to see

[3] However, d may also denote a metric, and so will be paid extra, [12] page 223.

that there is no loss of generality in assuming $\mathbb{B} = \mathrm{conv}\{\frac{x-y}{d_G(x,y)} \mid x \neq y \in V\}$. That is, we may assume $\mathbb{B}$ is a *centrally symmetric convex polytope.* A copy of $\mathbb{B}$ centered at x is denoted $\mathbb{B}(x)$.

The following well–known lemma shows that the notion of dimension is well–defined.

Lemma 5.1: *A metric space (X, d) with n points can be isometrically embedded into l_∞^n.*

Proof: Let $X = \{x_1, \ldots, x_n\}$ with $d_{ij} = d(x_i, x_j)$. Map x_i to a point $z^i \in \mathbb{R}^n$ whose k-th coordinate is $z_k^i = d_{ik}$. Then $||z^i - z^j||_\infty = \max_k |z_k^i - z_k^j| \geq |z_j^i - z_j^j| = |d_{ij} - d_{jj}| = d_{ij}$. On the other hand, for all k, $|z_k^i - z_k^j| = |d_{ik} - d_{jk}| \leq d_{ij}$ by the triangle inequality. ∎

We show later some examples of graphs with isometric dimension $\frac{n}{2}$. The highest isometric dimension among n-vertex graphs is currently unknown (but see [8] and [60]).

The l_∞ norm is *universal* in that, as Lemma 5.1 shows, it realizes all graphs (in fact, all finite metric spaces). In some sense, it is the *only* universal norm, as the l_∞ norm defines a graph on $\mathbb{Z}^d$, which must be realized by any other universal norm. Therefore any universal norm must contain a copy of l_∞. (Lest the reader suspects that l_1 is universal as well, we remark that $K_{2,3}$ is not embeddable in this norm, viz. [17].) It is therefore of interest to also study $\overline{\dim}(G)$ (and $\overline{\dim}_c(G)$), the least k such that G can be isometrically embedded in l_∞^k (with distortion c). Clearly $\overline{\dim}(G) \geq \dim(G)$, and this gap can be exponential (e.g., $\dim(d\text{-Cube}) = d$ while $\overline{\dim}(d\text{-Cube}) = 2^{d-1}$; see Corollary 5.12 and comments following Theorem 5.15). One simplification in studying isometric embeddings into l_∞^d is that all vertices may be assumed to map to $\mathbb{Z}^d$: Given any embedding, round all coordinates up and isometry is preserved.

Geodetic paths in G and the face lattice of the unit sphere $\mathbb{B}$ are related via

Proposition 5.2: *Let $P = (x_1, \ldots, x_k)$ be a geodetic path in G. In an isometric embedding of G in $(\mathbb{R}^d, \mathbb{B})$, all vectors $\frac{x_j - x_i}{j-i}$, $k \geq j > i \geq 1$, lie on the same face of $\mathbb{B}$.* ∎

J is an *isometric subgraph* of G if distances within J are the same as in the whole graph G. The dimension of such a subgraph provides a lower bound on $\dim(G)$. Examples of isometric subgraphs of G include cliques, induced subgraphs of diameter 2, geodetic paths and irreducible cycles. In particular, since $\dim(C_n) \geq \frac{n}{4} - 1$ (Proposition 5.10 and Remark 5.11), if G has finite girth, then $\dim(G) \geq \frac{\mathrm{girth}(G)}{4} - 1$.

To get the reader initiated on methods for estimating isometric dimensions, we present a bound on the dimension of trees:

Theorem 5.3: *For every n-vertex tree T, $\dim(T) = O(\log n)$. Moreover, if T has l leaves, then $\dim(T) = O(\log l)$. The bounds are tight.*

Proof: The proof shows that T can be isometrically embedded in $l_\infty^{c \cdot \log n}$ with $c = \frac{1}{\log 3 - 1}$. It is well-known that every n-vertex tree T has a "central" vertex O, such that each component of $T \setminus \{O\}$ has at most $\frac{2n}{3}$ vertices. Let R, L be two subtrees of T of size $\leq \frac{2n}{3}$ whose union is T, sharing only the vertex O. Find isometric embeddings, one for L and one for R in $l_\infty^{c \cdot \log n - 1}$. Such an embedding remains an isometry if all points are translated the same amount. We may thus assume that in the embeddings of L and R the vertex O is mapped to the origin in $\mathbb{R}^{c \cdot \log n - 1}$. The whole tree is isometrically embedded in a space of one more dimension. The coordinates used to embed R and L are maintained, and the value of the new coordinate is set as follows: For O it is zero, for $x \in V(L)$ this new coordinate is $-d_T(O, x)$ and for $x \in V(R)$ it is $d_T(O, x)$. It is not hard to verify that this is an isometric embedding, as claimed.

The upper bound in terms of the number of leaves is obtained by splitting T into two subtrees with a single common vertex, neither of which contains more than $\frac{2}{3}$ of the leaves of T (see [13] Theorem 2.1').

Both bound are tight for stars, $\dim(K_{1,n-1}) = \Omega(\log n)$ by Proposition 5.5. ∎

5.2 Estimating the dimension – volume considerations

Although stated in a different context, [16] shows that $\dim(K_n) \geq \lceil \log_2 n \rceil$. Here is a sketch of an argument based on the original ideas and translated to our language:

Proposition 5.4: $\dim(K_n) = \lceil \log_2 n \rceil$.

Proof: (Sketch) Let K_n be isometrically mapped to $\{x_1, \ldots, x_n\}$ in $(\mathbb{R}^d, \mathbb{B})$. Let $D = \text{conv}\{x_1, \ldots, x_n\}$. We claim that the sets $D + x_i$ $(i = 1, \ldots, n)$ have disjoint interiors. Assuming this for a moment, notice that since D is convex, $D + x_i \subseteq 2D$ for all i. Hence $n \cdot \text{vol}(D) = \text{vol}(\cup_i (D + x_i)) \leq \text{vol}(2D) = 2^d \cdot \text{vol}(D)$ and the conclusion follows. To complete the proof, suppose for contradiction that $(D + x_i) \cap (D + x_j)$ has a nonempty interior. This implies that $x_i - x_j$ is an internal point of $D - D$. But $D - D = \text{conv}\{x_\alpha - x_\beta \mid n \geq \alpha \neq \beta \geq 1\}$ which are all vectors of norm 1, whence $D - D \subseteq \mathbb{B}$. It follows that $x_i - x_j$, a vector of norm 1, is in $\text{int}(\mathbb{B})$, a contradiction.

On the other hand $\dim(K_n) \leq \lceil \log_2 n \rceil$ follows by mapping the vertices of K_n to the vertices of the $\lceil \log_2 n \rceil$-dimensional cube, under l_∞ norm. ∎

Volume considerations yield upper bounds on degrees and a lower bound on the diameter:

Proposition 5.5: *All vertex degrees in a d-dimensional graph do not exceed $3^d - 1$. This bound is tight.*

Proof: Place a copy of $\frac{1}{2}\mathbb{B}$ around a vertex v, and around each of its neighbors. The interiors of all these balls are disjoint, and their union is contained in $\frac{3}{2}\mathbb{B}(v)$: If w is a neighbor of v and $z \in \frac{1}{2}\mathbb{B}(w)$ then $\|z - w\| \leq \frac{1}{2}$ since $z \in \frac{1}{2}\mathbb{B}(w)$, and $\|w - v\| = 1$ by the isometry. Hence $\|z - v\| \leq \|z - w\| + \|w - v\| \leq \frac{3}{2}$. Comparing volumes we get the desired result. Equality is attained by the grid points under l_∞ norm. ∎

Lemma 5.6: $\text{diam}(G) \geq \frac{1}{2}(n^{\frac{1}{d}} - 1)$.

Proof: Surround each vertex of G by $\frac{1}{2}\mathbb{B}$. These balls have disjoint interiors and are contained in $(\text{diam}(G) + \frac{1}{2})\mathbb{B}$, centered at an arbitrary vertex of G. The conclusion follows as before. The bound is nearly tight, as we know of graphs with $\text{diam}(G) \sim \frac{1}{\sqrt{2}}(n^{\frac{1}{d}} - 1)$. ∎

5.3 Estimating the dimension – ranks

Isometric dimensions are related to linear algebra via an alternative definition of a norm: Recall that we are only concerned with normed spaces $(X, \mathbb{B})$ where $\mathbb{B}$ is a centrally symmetric polytope. Associate with each pair of opposite facets $\{F, -F\}$ of $\mathbb{B}$, a linear functional l_F which is identically 1 on F, and -1 on $-F$. Then $\forall x \in X,\ \|x\| = \max_F |l_F(x)|$.

An $n \times r$ matrix M *implements* a graph G if for all i, j $\max_k |m_{ik} - m_{jk}| = d_G(v_i, v_j)$. The following result offers a characterization of $\dim(G)$ in terms of matrix ranks:

Theorem 5.7: $\dim(G) = \min \text{rank}(M)$*, where the minimum is over all matrices M that implement G.*

Proof: If G is isometrically embeddable in $(\mathbb{R}^d, \mathbb{B})$, define M via: $m_{iF} = l_F(v_i)$, where l_F is the functional corresponding to the pair of opposite facets $\{F, -F\}$ of $\mathbb{B}$. Clearly, M implements G, and its rank is $\leq \dim(G)$.

On the other hand, suppose that $M_{n \times s}$ implements G and $d = \text{rank}(M)$. Mapping the vertices of G to the rows of M is an embedding of G to the d-dimensional space L, spanned by the rows of M. The norm is induced by the unit sphere $\mathbb{B} = L \cap [-1, 1]^s$, the intersection of L with the unit cube in $\mathbb{R}^s$. The fact that M implements G implies that the above mapping is an isometry of G into the normed space $(L, \mathbb{B})$. ∎

Here are some applications of the theorem:

Theorem 5.8: *If* $n_1, n_2, \ldots, n_k \geq 2$, *then* $\dim(K_{n_1,n_2,\ldots,n_k}) \geq \sum_{i=1}^{k} \lceil \log n_i \rceil - 1$.

Proof: Let A_i be the i-th part of V and M be a matrix implementing G. For $a, b \in A_i$, consider a column j where $|m_{aj} - m_{bj}| = 2$. If $x \notin A_i$, then $d(x,a) = d(x,b) = 1$, whence $m_{xj} = \frac{1}{2} \cdot (m_{aj} + m_{bj})$. It follows that if $\frac{1}{2} \cdot (m_{aj} + m_{bj}) \cdot \vec{1}$ is subtracted from the j-th column, all entries not in the rows of A_i become zero. Repeat this step for all columns j that implement the distance between two points from the same part. Next, eliminate all other columns. The resulting matrix Q has $\text{rank}(Q) \leq \text{rank}(M) + 1$ (elementary operations with a single column can increase the rank by ≤ 1 and eliminating columns can only decrease it). Q is a direct sum $Q = \oplus Q_i$, where $\frac{1}{2} Q_i$ implements K_{n_i}, whence $\text{rank}(Q_i) \geq \lceil \log n_i \rceil$. The theorem follows. ∎

The upper bound $\dim(K_{n_1,n_2,\ldots,n_k}) \leq \sum_{i=1}^{k} \lceil \log n_i \rceil$ is shown easily, using l_∞ norm, as suggested by the proof of the lower bound.

Corollary 5.9: *Let G be a clique K_{2n} minus a perfect matching. Then* $\dim(G) = n$.

Proof: Here, it is obvious that the the elementary operations on the columns do not cause a loss of one in the dimension. ∎

Proposition 5.10: $\dim(C_{2m}) = m$.

Proof: The following construction gives an upper bound, with l_1 norm: The vertices are mapped to the following $2m$ points:

$$\begin{aligned} \text{for } i = 1, \ldots, m: \ & x_i = \textstyle\sum_{t=1}^{i} e_t; \\ \text{for } i = m+1, \ldots, 2m-1: \ & x_i = \textstyle\sum_{t=i-m+1}^{m} e_t; \\ & x_{2m} = 0, \end{aligned}$$

where e_t is the t-th unit vector.

For the lower bound, let the matrix A implement C_{2m}, and let $v_1, \ldots, v_{2m}$ be the vertices in cyclic order. All indices are taken modulo $2m$. Consider a column t where $d_G(v_j, v_{m+j}) = m$ is realized. It has the form $a_{j,t} = \delta$; $a_{j+1,t} = a_{j-1,t} = \delta + \epsilon$; $a_{j+2,t} = a_{j-2,t} = \delta + 2\epsilon$; $\ldots, a_{j+m,t} = \delta + m\epsilon$, for some real δ and $\epsilon \in \{-1, 1\}$. By elementary operations with the column vector $\vec{1}$ it can be transformed so that $a_{j-i,t} = a_{j+i,t} = i$ for $i = 0, \ldots, m$. We thus obtain an $m \times m$ minor whose (r,s)-entry is $|r - s|$. This matrix is non–singular, being the distance matrix of a path, which is known to be non-singular (e.g., [42] pp. 64-65). This implies $\text{rank}(A) \geq m - 1$. A more careful analysis shows that the elementary operations with the all–one vector can be avoided, yielding $\text{rank}(A) \geq m$. ∎

Remark 5.11: For cycles of odd length: $m + 1 \geq \dim(C_{2m+1}) \geq \frac{m}{2} - 1$. The upper bound is achieved with l_∞ norm: Let $\{w_1, w_2, \ldots, w_{m+1}\}$ be $m+1$ consecutive vertices in the cycle. Map each vertex x to an $(m+1)$-vector, whose i-th coordinate is $d(x, w_i)$. As for a lower bound, the above argument yields only $\dim(C_{2m+1}) \geq \frac{m}{2} - 1$, though probably $\dim(C_n) = \lceil \frac{n}{2} \rceil$ for all n. ∎

Consequently:

Corollary 5.12: $\dim(m\text{-Cube}) = m$.

Proof: The m-Cube embeds isometrically in l_1^m. The lower bound follows from the fact that the m-Cube contains a $2m$-Cycle as an isometric subgraph. ∎

Consequently, the infinite cubic grid in $\mathbb{R}^m$ has dimension m. Moreover, for the part of the grid $G = [1, \ldots, n]^m$, not only does $\dim(G) = m$, but also $\dim_c(G) = m$ for any $c \leq n^{\frac{1}{m}}$. This bound can be obtained using volume arguments as described in a previous section.

The *stabbing dimension* of a finite family of convex bodies $\mathcal{K}$ in $\mathbb{R}^d$ is the least dimension of a linear space L which intersects every $K \in \mathcal{K}$.

Associate with a connected graph G on n vertices, the polyhedron $P = P_G \subseteq \mathbb{R}^n = \{x \in \mathbb{R}^n$ with $|x_i - x_j| \leq d(v_i, v_j)$ for all $1 \leq i, j \leq n\}$. Alternatively, $P = \{x \in \mathbb{R}^n$ with $|x_i - x_j| \leq 1, \forall [v_i, v_j] \in E(G)\}$. Clearly, P is a centrally symmetric prism. For each i, j the facets $F^+_{i,j}$ and $F^-_{i,j}$ are determined by the equation $x_i - x_j = \pm\, d(v_i, v_j)$. Let $\mathcal{F} = \mathcal{F}_G$ consist of all such faces. (In fact, because of the central symmetry, it would suffice to consider $F^+_{i,j}$ alone.)

Theorem 5.13: *The stabbing dimension of $\mathcal{F}_G$ coincides with the isometric dimension of G.*

Proof: Suppose that L "stabs" all the faces in $\mathcal{F}_G$, and choose for each pair i, j a vector in $L \cap F_{i,j}$. Let the matrix M have these vectors as columns. Clearly M implements G; by Theorem 5.7, $\dim(G) \leq \dim(L)$.

On the other hand, given a matrix of minimum rank implementing G, define L to be the span of its columns. Clearly L meets all the required facets, and $\dim(L) = \dim(G)$. ∎

The case when G is a clique has an interesting geometric implication:

Theorem 5.14: *Let C be the cube $[-\frac{1}{2}, \frac{1}{2}]^m$. The stabbing dimension of any family of n pairwise disjoint faces $\{F_1, \ldots, F_n\}$ of C is at least $\lceil \log_2 n \rceil$.*

Proof: Let L be a linear space that meets all F_i. Choose points $v_i \in L \cap F_i$, and form an $n \times m$ matrix M whose rows are the v_i's. Since the faces are disjoint, for each two rows $i \neq j$ there is a column l where $v_{i,l} = -\frac{1}{2}$ and $v_{j,l} = \frac{1}{2}$, or vice versa. Since all entries of M are in $[-\frac{1}{2}, \frac{1}{2}]$, M implements the n-Clique.

By Theorem 5.7 and Proposition 5.4 $\text{rank}(M) \geq \lceil \log_2 n \rceil$, whence $\dim L = \text{rank} M \geq \lceil \log_2 n \rceil$ as claimed. ∎

The remaining part of the section concerns $\overline{\dim}(G)$.

Theorem 5.15:

- $\overline{\dim}(G)$ *is the least number of columns in a matrix implementing G.*
- $\overline{\dim}(G)$ *equals half the least number of faces of the unit ball of a normed space in which G can be isometrically embedded.* ∎

Here are some examples demonstrating the convenience of working with $\overline{\dim}(G)$:

(i) $\overline{\dim}(K_n) = \lceil \log_2 n \rceil$.

Let M realize K_n, and recall that M may be assumed to have integer entries. Hence, each column in M realizes the distance between every $x \in A$ and $y \notin A$ for some $A \subset V$. Namely, an isometric embedding of K_n into l^d_∞ is equivalent to covering $E(K_n)$ by d complete bipartite graphs and it is well known that $d \geq \lceil \log_2 n \rceil$, as claimed.

A similar argument shows that $\overline{\dim}\,(\text{linegraph}(K_n)) = \Theta(\log n)$.

(ii) $\overline{\dim}(d\text{-Cube}) = 2^{d-1}$.

Let M implement the d-Cube and let column t realize $d(x, y) = d$ for some antipodal pair x, y. Without loss of generality $m_{tz} = d(x, z)$. Hence every antipodal pair requires a separate column. At the same time this is just a description of a matrix implementing the d-Cube with 2^{d-1} columns.

6 Separators

Theorem 6.1: *Let $\dim_c(G) = d$ and assume that $c \cdot d = o(n^{\frac{1}{d}})$. Then G has a set S of $O(c \cdot d \cdot n^{1-\frac{1}{d}})$ vertices which separates the graph, so that no component of $G \setminus S$ has more than $(1 - \frac{1}{d+1} + o(1))n$ vertices.*

Proof: Given a distortion-c embedding of G in $(\mathbb{R}^d, \mathbb{B})$, our approach is this: Find two parallel hyperplanes H_1, H_2 in $\mathbb{R}^d$ at distance 1 (distance is taken in the $\mathbb{B}$-norm). S is the set of vertices which are embedded in the closed *slab* between the two hyperplanes. V_1 is the set of vertices embedded strictly above H_1 and in V_2 are those embedded strictly below H_2. That S separates V_1 and V_2 is obvious. What we need is to construct H_1 and H_2, so that,

$$|S| = O(c \cdot d \cdot n^{1-\frac{1}{d}})$$

and

$$|V_1|, |V_2| \leq \left(1 - \frac{1}{d+1} + o(1)\right) n.$$

The proof uses a beautiful idea from [48] which starts from the following well known consequence of Helly's Theorem [61]:

Proposition 6.2: *For any set V of n points in $\mathbb{R}^d$ there exists a* centroid *O such that every closed halfspace determined by a hyperplane passing through O contains at least $\frac{n}{d+1}$ points of V.*

By translation, O may be assumed to be the origin.

Partition the points in V according to their distance from O. First we will show that no slab contains "too many" points of V, that are "near" the origin. Then we show that on a random choice of H_1, and $H_2 = -H_1$, the expected number of points in V which are "far" from the origin and fall in the slab is small.

It is a well known fact (e.g., [50]) that every d-dimensional norm may be approximated with distortion $\sqrt{d}$ by an affine image of the Euclidean norm. That is, we may assume that G is embedded in l_2^d with distortion $c\sqrt{d}$.

So we should look for a closed slab $\mathbb{H}$ of (Euclidean) width $c\sqrt{d}$ containing "not too many" points of V. Let $n_1 = \#\{x : x \in V \cap \mathbb{H}, \|x\|_2 < R_0\}$.

Lemma 6.3: *Let $R_0 = \Theta(c \cdot \sqrt{d})$. Then $n_1 \leq O(c \cdot d \cdot (2R_0 + 1)^{d-1})$.*

Proof: All distances in G are ≥ 1, so the same holds also for the Euclidean distance of the images. Therefore, if we locate a $\frac{1}{2}\mathbb{B}$ around each point in $V \cap \mathbb{H}$ we get n_1 spheres with disjoint interiors. Since $R_0 = \Theta(c \cdot \sqrt{d})$, they all reside inside a cylinder of height $\Theta(c \cdot \sqrt{d})$ and base a $(d-1)$-dimensional ball of radius $(R_0 + \frac{1}{2})$. Comparing volumes we obtain:

$$n_1 \cdot v_d \cdot \left(\frac{1}{2}\right)^d \leq v_{d-1} \cdot \left(R_0 + \frac{1}{2}\right)^{d-1} \cdot O(c \cdot \sqrt{d}).$$

Where v_t is the volume of the unit ball in $\mathbb{R}^t$. Recall that $v_{2t} = \frac{\pi^t}{t!}$ and $v_{2t+1} = \frac{2^{t+1} \cdot \pi^t}{1 \cdot 3 \cdots (2t+1)}$, and in particular, $\frac{v_t}{v_{t+1}} = \Theta(\sqrt{t})$.

Consequently: $n_1 \leq O(c \cdot d \cdot (2R_0 + 1)^{d-1})$. ∎

Now, we wish to estimate the probability for a *remote* point x (i.e., $\|x\|_2 \geq R_0$) to belong to a randomly chosen slab. A slab is determined by the unit vector perpendicular to its boundary and our choice is by the uniform distribution on $\mathbb{S}^{d-1}$.

Lemma 6.4: *Let $x \in \mathbb{R}^d$, $\|x\|_2 \geq R_0$. Then $Pr(x \in \mathbb{H}) \leq O\left(\frac{c \cdot d}{R_0}\right)$.*

Proof: Associate with each slab $\mathbb{H}$ the two points on $\mathbb{S}^{d-1}$ in the directions of the two unit vectors perpendicular to the hyperplanes of $\mathbb{H}$. Slabs have width $c\sqrt{d}$, so the points associated to slabs containing

x form a symmetric stripe of width $2c\frac{\sqrt{d}}{\|x\|_2}$, on $\mathbb{S}^{d-1}$. Therefore, the desired probability is the ratio between the surface area of this stripe and the surface area of the whole sphere. We recall the following fact:

Remark 6.5: Let C be a measurable subset of $\mathbb{S}^{d-1}$, and let α be the ($(d-1)$-dimensional) measure of C. Let $\sigma(C) = \{y \mid y = \lambda x \text{ for some } x \in C \text{ and } 1 \geq \lambda \geq 0\}$, the cone with base C and apex at the origin. Then the (d-dimensional) measure of $\sigma(C)$ is $\frac{\alpha}{d}$. In particular, the surface area of $\mathbb{S}^{d-1}$ is $d \cdot v_d$. ∎

We need to evaluate the surface area of C, the part of the stripe of width $2c\frac{\sqrt{d}}{\|x\|_2}$, that is on $\mathbb{S}^{d-1}$. By the previous remark, this surface area equals $d \cdot \mathrm{vol}(\sigma(C))$. Assume that C is symmetric with respect to the hyperplane $z_d = 0$. Then $\sigma(C) \subseteq \{(z_1, \ldots, z_d) \text{ with } \sum_1^{d-1} z_i^2 \leq 1 \text{ and } |z_d| \leq \frac{c\cdot\sqrt{d}}{\|x\|_2}\}$, and as the volume of this cylinder is $2\frac{c\cdot\sqrt{d}}{\|x\|_2} v_{d-1}$:

$$Pr(x \in \mathbb{H}) \leq 2\frac{c \cdot d^{3/2}}{\|x\|_2} \cdot \frac{v_{d-1}}{d \cdot v_d} = O\left(\frac{c \cdot d}{R_0}\right). \blacksquare$$

Thus n_2, the expected number of remote points $x \in V$ which belong to $\mathbb{H}$, satisfies $n_2 = O(\frac{n \cdot c \cdot d}{R_0})$. We optimize, by selecting R_0 so that $n_1 \sim n_2$. Then

$$\frac{n \cdot c \cdot d}{R_0} \sim c \cdot d \cdot {R_0}^{d-1}$$

or

$$R_0 = \Theta(n^{\frac{1}{d}}).$$

This yields $n_1 + n_2 = O(c \cdot d \cdot n^{1-\frac{1}{d}})$, for the expected number of points in the slab.

The requirements $R_0 = \Theta(c \cdot \sqrt{d})$ and $R_0 = O(n^{\frac{1}{d}})$ are consistent, since $c \cdot d = o(n^{\frac{1}{d}})$ was assumed. ∎

The centroid can be found in time linear in n and d^d (see [45]). Therefore, given an embedding of G, the proof translates to a randomized polynomial time algorithm to find such a separator, provided that $d = O(\frac{\log n}{\log\log n})$. It is interesting to observe that in [47] an essentially similar separation is obtained, although both the setting and the methods are different.

It is not difficult to see that for G that is a product of d paths of length $n^{1-\frac{1}{d}}$, the theorem is essentially tight.

7 Low–diameter decompositions of graphs

Following [41] a *decomposition* of a graph $G = (V, E)$ is a partition of the vertex set into subsets (called *blocks*). The *diameter* of the decomposition is the least δ such that any two vertices belonging to the same connected component of a block are at distance $\leq \delta$ in the graph. Modifying this definition in the spirit of [6], we consider *coverings* of G wherein distinct blocks may have nonempty intersections. Diameters of coverings are defined as for decompositions. The *degree* of a covering is the largest number of blocks to which any vertex may belong. A covering is *r-subsuming* if every r-ball in G is contained in some block of the covering.

Theorem 7.1: *Let G be a graph with $\overline{d} = \overline{\dim}_c(G)$, $d = \dim_c(G)$, and let r be a positive integer. Then*

1. *G can be decomposed to $\overline{d} + 1$ blocks, each of diameter $\leq 2c\overline{d}$.*

2. *G can be decomposed to $d+1$ blocks, each of diameter $\leq 2cd^2$.*

3. *G has a covering by $\overline{d}+1$ blocks, each of diameter $\leq (6\overline{d}+2)cr$, that is r-subsuming, and the cover has degree $\leq (\overline{d}+1)$.*

4. *G has a covering by $d+1$ blocks, each of diameter $\leq (6d+2)dcr$, that is r-subsuming, and the cover has degree $\leq (d+1)$.*

Remark 7.2: Combined with Theorem 3.2 we obtain a decomposition to $O(\log n)$ blocks of diameter $O(\log^3 n)$, and a covering of diameter $O(\log^3 n)$ with degree $O(\log n)$, results slightly inferior to the optimal $(O(\log n), O(\log n))$ from [41] and [6]. ∎

Proof: We prove the case $c = 1$, the general case then follows easily. Throughout the proof we use l_∞ norm. The key to the proof is the following universal tiling of $\mathbb{R}^{\overline{d}}$: Consider $\mathbb{Z}^{\overline{d}}$, and define K_i, $i = 0, \ldots, \overline{d}-1$ as the $\frac{2\overline{d}-(2i+1)}{4\overline{d}}$ neighborhood of the i-dimensional faces of its cubes (i.e., K_0 consists of radius $\frac{2\overline{d}-1}{4\overline{d}}$ cubes centered at the grid points, K_1 is the $\frac{2\overline{d}-3}{4\overline{d}}$ neighborhood of the edges of the grid, etc.). Define $T_0 = K_0$, $T_i = K_i - \cup_{j=0}^{i-1} K_j$ for $i = 1, \ldots, \overline{d}-1$. Finally, $T_{\overline{d}}$ is the remaining part of $\mathbb{R}^{\overline{d}}$. It is not hard to check now that each T_i is a union of disjoint "bricks", each of diameter < 1, and that the distance between any two such bricks is $\geq \frac{1}{2\overline{d}}$.

Claim 1 is now immediate: embed the graph in $\mathbb{R}^{\overline{d}}$, and consider the tiling as above, magnified by factor $2\overline{d}+\epsilon$. Each $G \cap T_i$ is a proper block.

The proof of Claim 2 is slightly more complicated, since one has to correlate between the d-dimensional grid and an arbitrary d-dimensional norm. We need a small distortion approximation of $\mathbb{B}$ (or a linear transformation thereof) by a cube. Distortion d is attainable: first approximate $\mathbb{B}$ by a Euclidean unit ball (actually, by its Löwner–John ellipsoid, see [50]), then approximate the unit ball by a unit cube, both distortions being $\leq \sqrt{d}$. By recent work of Giannopoulos [24] an $O(d^{0.859})$ distortion is attainable.

As before, we embed G in $\mathbb{R}^d$, and superimpose on it the (unit) lattice that approximates $\mathbb{B}$ to a factor of d, magnified by $2d^2+\epsilon$. It is easy to check that, again, each T_i defines a proper block.

If $c > 1$, the only change is that the diameter of sets covered by a single brick may be multiplied by c.

To prove Claim 3: Magnify the tiling by a factor of $6\overline{d}+\epsilon$. Turn the tiling into a covering by defining new blocks as the 1-neighborhoods of the old blocks. Since different blocks of T_i were at least 3 apart, no point in $\mathbb{R}^d$ is covered more than $d+1$ times. Furthermore – each 1-neighborhood in $\mathbb{R}^d$ is covered by at least one new block, since $\mathbb{R}^d$ is tiled by the old blocks (identify each 1-neighborhood with its center). The diameter of each new block is $\leq 6\overline{d}+2$, since any two connected components of the same block are at least one unit apart.
In order to finish the proof – given $r > 0$ – magnify the covering by a factor of r, and embed G in $\mathbb{R}^d$.

The proof of Claim 4 follows easily from the same arguments. ∎

Remark 7.3: We have covered $\mathbb{R}^d$ with $\{T_0, \ldots, T_d\}$, where each T_i is a union of compact sets of diameter < 1, any two of which are at least $\frac{1}{2d}$ apart. The construction is nearly optimal in two respects:

It is impossible to cover $\mathbb{R}^d$ with fewer than $d+1$ sets each of which is the disjoint union of compact sets whose diameters are bounded from above and whose mutual distances are bounded away from zero. This follows, e.g., from Lemma 3.4 in [41].

We next show that for any cover $\{T_i\}_{i=0}^d$ as above, there are two sets in the same family whose distance does not exceed $O(\frac{\log d}{d})$. Indeed, there must be a T_i with upper density at least $\frac{1}{d+1}$. Let $T^* = T_i + m\mathbb{B}$ (Minkowski Sum), where $2m$ is the least distance between two connected components of T_i. If K_α is a typical connected component of T_i, then the sets $K_\alpha + m\mathbb{B}$ have disjoint interiors. Also T_i has upper density at least $\frac{1}{d+1}$ in $\mathbb{R}^d$, and therefore also in $K^* = \cup(K_\alpha + m\mathbb{B})$. By the Brunn–Minkowski inequality (see [50]):

$$\frac{\mathrm{vol}(K_\alpha + m\mathbb{B})}{\mathrm{vol}(K_\alpha)} \geq \frac{(\mathrm{vol}(K_\alpha)^{1/d} + \mathrm{vol}(m\mathbb{B})^{1/d})^d}{\mathrm{vol}(K_\alpha)} =$$

$$= \left(1 + \left(\frac{\mathrm{vol}(m\mathbb{B})}{\mathrm{vol}(K_\alpha)}\right)^{1/d}\right)^d.$$

We used the fact that in any norm $\mathrm{vol}(\mathbb{B}) \geq \mathrm{vol}(K)$ whenever $\mathrm{diam}(K) \leq 2$, which can be shown as follows: Consider the symmetrization $Q = (K - K)/2$ (Minkowski sum). Clearly, Q is centrally symmetric with respect to the origin, and $\mathrm{diam}(Q) \leq \mathrm{diam}(K) \leq 2$. Also, by Brunn–Minkowski $\mathrm{vol}(Q) \geq \mathrm{vol}(K)$. But the symmetry of Q and the bound on its diameter imply that $Q \subseteq \mathbb{B}$ and the conclusion follows. By the density properties of T_i, there is an index α for which $d \geq \frac{\mathrm{vol}(K_\alpha + m\mathbb{B})}{\mathrm{vol}(K_\alpha)} \geq (1+m)^d$ which implies $m = O(\frac{\log d}{d})$, as claimed.

We do not know what the bound on m is for various norms. A plausible guess would be that $m = O(1/d)$ for every norm, and that this is tight for l_∞. ∎

8 Further problems

Many of the questions addressed in this paper can be considered for directed graphs as well as for undirected ones. To get started in this direction, let us define a *directed metric* on X as a nonnegative real function d on $X \times X$, which satisfies the *directed* triangle inequality: $d(x,y) + d(y,z) \geq d(x,z)$ for every $x, y, z \in X$. A *directed norm* in $\mathbb{R}^n$ satisfies the same set of requirements as does a norm, except that $||\lambda x|| = \lambda ||x||$ is to hold only for nonnegative λ. The "unit ball" in such a space, i.e., the set of those vectors whose directed norm does not exceed one is a convex set which contains the origin in its interior (central symmetry is no longer required) and any such set $\mathbb{B}$ defines a directed norm in the obvious way: $||x|| = \inf\{\lambda > 0 \text{ such that } \frac{x}{\lambda} \in \mathbb{B}\}$.

It is pleasing to observe that an analogue of Lemma 5.1 holds in this more general context:

Proposition 8.1: *Every n-point directed metric space (X,d) can be isometrically embedded into $\mathbb{R}^n$ equipped with an appropriate directed norm.*

Proof: Let $X = \{x_1, \ldots, x_n\}$ with $d_{i,j} = d(x_i, x_j)$. Pick n linearly independent vectors $\{z_1, \ldots, z_n\}$. Let the unit ball be $\mathbb{B} = \mathrm{conv}\{\frac{z_i - z_j}{d_{i,j}} | i \neq j\}$, and map x_i to z_i. (According to this definition $\mathbb{B}$ is not full dimensional. This may be overcome by using $\mathbb{B}'$, a small height bi-pyramid over $\mathbb{B}$. Alternatively, project all z_i to the subspace spanned by $\mathbb{B}$.) The only way the claim could fail is that for some pair, say $\{1, n\}$, the point $\frac{z_1 - z_n}{d_{1,n}}$ is in the interior of $\mathrm{conv}\{\frac{z_i - z_j}{d_{i,j}} | i \neq j\}$.

If this is the case, then

$$\mu \cdot \frac{z_1 - z_n}{d_{1,n}} = \sum_{i,j} \alpha_{i,j} \frac{z_i - z_j}{d_{i,j}}$$

for some $\mu > 1$ and some nonnegative $\alpha_{i,j}$ whose sum is 1. Pick such an equality with μ as large as possible. Let H be the directed graph on vertex set $[n]$ where (i,j) is an edge iff $\alpha_{i,j} > 0$, and observe that:

- H is acyclic: Associated with every directed cycle C in H is a linear dependency with positive coefficients: $\sum_{(i,j)\in E(C)} \delta_{i,j} \frac{z_i - z_j}{d_{i,j}} = 0$ (with $\delta_{i,j} > 0$). It is possible to replace $\alpha_{i,j}$ by $\alpha_{i,j} - t \cdot \delta_{i,j}$ for all $(i,j) \in E(C)$ for some $t > 0$ and renormalize the sum, so as to retain $\alpha_{i,j} \geq 0$ while increasing μ, contrary to our assumption.

- By linear independence of the z_i, the only source in H is 1, and the only sink is n.

- Without loss of generality no two directed paths in H have the same starting vertex and the same final vertex. Otherwise, it is possible to shift weight from one to the other without decreasing μ.

Consequently, H is a single directed path from 1 to n and the above equality has the form:

$$\mu \cdot \frac{z_1 - z_n}{d_{1,n}} = \sum_{i=1}^{n-1} \alpha_{i,i+1} \frac{z_i - z_{i+1}}{d_{i,i+1}}$$

with $\alpha_{i,i+1} > 0$ and $\sum_{i=1}^{n-1} \alpha_{i,i+1} = 1$.

The linear independence of the z_i's implies that:

$$\frac{\mu}{d_{1,n}} = \frac{\alpha_{1,2}}{d_{1,2}} = \frac{\alpha_{2,3}}{d_{2,3}} = \ldots = \frac{\alpha_{n-1,n}}{d_{n-1,n}}.$$

So

$$1 = \sum \alpha_{i,i+1} = \frac{\mu}{d_{1,n}} \sum d_{i,i+1} \geq \mu$$

by the directed triangle inequality, a contradiction. ∎

It is an intriguing idea that large diameters in graphs can be essentially attributed to low-dimensionality. The easy converse is our Lemma 5.6. Attempts to make this statement precise were, in fact, the initial motivation for this research. A plausible conjecture along these lines was formulated with the help of L. Levin. We are grateful for his permission to include it here. Let $\mathbb{Z}_\infty^d$ denote the graph of the d-dimensional lattice with l_∞ metric. Define the *growth rate* $\rho(G)$ of a graph G as the maximum (over all choices of r and x) of $\frac{\log|B(x,r)|}{\log(r+1)}$ (where $B(x,r)$ is the r-ball around x).

Conjecture 8.2: Let G have growth rate $\rho = \rho(G)$. Then $\mathbb{Z}_\infty^{O(\rho)}$ contains a (not necessarily induced) subgraph isomorphic to G. ∎

By a standard counting argument, fewer than $\rho(G)$ dimensions will not suffice.

The conjecture is true for $G = d$-Cube and here is a sketch of a proof: First one checks that $\rho = \rho(d\text{-Cube}) = \Theta(\frac{d}{\log d})$. Let $y_1, \ldots, y_{O(\rho)}$ be randomly selected d-dimensional $(-1,1)$-vectors. Each vertex x of the d-Cube, (considered as a d-dimensional $(0,1)$-vector) gets mapped to the vector of (real) inner products $(< x, y_i > | i = 1, \ldots, O(\rho))$. Neighboring vertices in the d-Cube are mapped to adjacent vertices in $\mathbb{Z}_\infty^{O(\rho)}$, since the y_i's are $(-1,1)$-vectors. Also, with high probability this mapping is one-to-one: Let x, z be two vertices of the cube, at Hamming distance $t = \Omega(\sqrt{d})$. Their images agree in the i-th coordinate iff $< x, y_i > = < z, y_i >$. The probability of this event is exactly the chance for a one–dimensional random walk to be in the origin at time t, i.e., $O(t^{-1/2})$. Therefore, the probability that x and z have the same image is $t^{-\Omega(\rho)} = o(2^{-d})$. Consequently, almost surely no collision occurs among points of Hamming distance $\Omega(\sqrt{d})$. On the other hand, if the Hamming distance between x and z is $O(\sqrt{d})$, their probability of collision is $\leq 2^{-\Omega(\rho)}$, but $2^{-\Omega(\rho)} \sum_{s=O(\sqrt{d})} \binom{d}{s} = o(1)$. Therefore with high probability this mapping is one-to-one, as claimed. The proof for regular trees follows by similar arguments.

Note that $\rho(G)$ may be much smaller than any dimension in which G may be embedded nearly isometrically:

Example 8.3: Construct a tree of depth $2m$ as follows: For $m \geq i \geq 1$ level i contains exactly $(i+1)^2$ vertices. Each vertex in level i has at least one child and the $2i+3$ vertices in a randomly selected set have two children. Vertices in the last m levels have exactly one child each. The number of vertices in this tree is $n = \Theta(m^3)$. Now, while $\rho(G)$ is a constant, a near isometric embedding of G requires dimension $\Omega(\log n)$, since the distance between any two of the $(m+1)^2$ leaves is between $2m$ and $4m$. The conclusion follows from the standard volume argument. ∎

A related notion is the *bandwidth*, $\mathrm{bw}(G)$, of a graph G: It is the least w such that there is a bijection $f : V(G) \to [n]$ for which $|f(u) - f(v)| \leq w$ for every $(u, v) \in E(G)$.

Conjecture 8.4: If $\dim(G) = d$, then $\mathrm{bw}(G) \leq O(k(d) \cdot n^{1-\frac{1}{d}} \cdot \mathrm{polylog}(n))$. ∎

This bound is tight as the bandwidth of the discrete cube $([1, n^{\frac{1}{d}}]^d)$ is $\Omega(n^{1-\frac{1}{d}})$. The case $d = 2$ can be solved using [1].

A question raised in [11] and [4] is to estimate the least $\psi = \psi_c(n)$ so that every n-point metric space can be embedded in a ψ-dimensional normed space with distortion c. In particular – what is the least dimension needed to embed a constant–degree expander graph with constant distortion (we have shown that it is $\Omega(\log^2 n)$)? It is not impossible that a constant–distortion embedding of such graphs requires $\Omega(n)$ dimensions.

An obvious general question is

Problem 8.5: What is the computational complexity of deciding whether $\dim(G) = d$? Similarly for $\overline{\dim}(G)$ and $\dim_c(G)$. ∎

It is not hard to see, e.g., by Proposition 5.5, that almost all n-vertex graphs have dimension at least $\Omega(\log n)$. We would like to know

Problem 8.6: What is the typical dimension of an n-vertex graph? ∎

We suspect the answer to be linear or nearly linear. The situation for near–isometries is quite different. Since almost all n-vertex graphs have diameter 2, and since $\dim(K_n) = \lceil \log_2 n \rceil$, almost all graphs satisfy $\dim_2(G) = O(\log n)$. Furthermore, by the method of [4], $\dim_{1+\epsilon}(G) \leq c_\epsilon \log n$ for almost all graphs and every $\epsilon > 0$. On the other hand $\overline{\dim}(G) = \Omega(n)$ for most graphs, because almost surely no single coordinate can implement the distance between more than n pairs x, y with $d(x, y) = 2$, but almost surely there are $\Omega(n^2)$ such distances in a random G.

We are only starting to understand the role of girth in this field (but see [52]), and offer:

Problem 8.7: Let all vertices in G have degree ≥ 3. Does every embedding of G in a Euclidean space (of any dimension) have distortion $\Omega(girth(G))$? ∎

9 Acknowledgments

The authors wish to thank Mike Saks, Micha A. Perles, Gil Kalai, Yoav Benyamini, Joram Lindenstrauss, Alex Samorodnitsky, Leonard Schulman, and Tomek Szankowski for stimulating discussions. We are particularly thankful to Leonid Levin for formulating Conjecture 8.2.

References

[1] N. Alon, M. Katchalski, and W.R. Pulleyblank, Cutting disjoint disks by straight lines, Discrete and Computational Geometry 4 (1989), 239-243.

[2] E. M. Andreev, Convex polyhedra in Lobačevskiĭ spaces, Mat. Sb. (N.S.) 81 (123) (1970), 445-478. English translation: Math. USSR Sb. 10 (1970), 413-440.

[3] E. M. Andreev, Convex polyhedra of finite volume in Lobačevskiĭ space, Mat. Sb. (N.S.) 83 (125) (1970), 256-260. English translation: Math. USSR Sb. 12 (1970), 255-259.

[4] J. Arias-de-Reyna and L. Rodrígues-Piazza, Finite metric spaces needing high dimension for Lipschitz embeddings in Banach spaces, Israel J. Math. 79 (1992), 103-111.

[5] Y. Aumann and Y. Rabani, An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm, preprint, 1994.

[6] B. Awerbuch and D. Peleg, Sparse partitions, FOCS 31 (1990), 503-513.

[7] L. Babai and P. Frankl, "Linear Algebra Methods in Combinatorics", Preliminary Version 2, Department of Computer Science, The University of Chicago, Chicago, 1992.

[8] K. Ball, Isometric embedding in l_p-spaces, Europ. J. Combinatorics 11 (1990), 305-311.

[9] B. Berger, The fourth moment method, SODA 2 (1991), 373-383.

[10] L. M. Blumenthal, "Theory and Applications of Distance Geometry", Chelsea, New York, 1970.

[11] J. Bourgain, On Lipschitz embedding of finite metric spaces in Hilbert space, Israel J. Math. 52 (1985), 46-52.

[12] L. Carroll, "Through the Looking–Glass and what Alice Found There", Chapter 6, Pan Books, London, 1947.

[13] F. R. K. Chung, Separator theorems and their applications, in "Paths, Flows, and VLSI–Layout" (B. Korte, L. Lovász, H. J. Prömel, and A. Schrijver eds.) Springer, Berlin – New York, 1990, 17-34.

[14] E. Cohen, Polylog–time and near–linear work approximation scheme for undirected shortest paths, STOC 26 (1994), 16-26.

[15] L. J. Cowen, On local representations of graphs and networks, PhD dissertation, MIT/ LCS/ TR-573, 1993.

[16] L. Danzer and B. Grünbaum, Über zwei Probleme bezüglich konvexer Körper von P. Erdös und von V. L. Klee, Math. Zeitschr. 79 (1962), 95-99.

[17] M. Deza and M. Laurent, Applications of cut polyhedra, Liens - 92 - 18, September 1992.

[18] M. Deza and H. Maehara, Metric transforms and Euclidean embeddings, Trans. AMS 317 (1990), 661-671.

[19] R.O. Duda and P.E. Hart, “Pattern Classification and Scene Analysis”, John Wiley and Sons, New York, 1973.

[20] P. Frankl and H. Maehara, The Johnson–Lindenstrauss lemma and the sphericity of some graphs, J. Comb. Th. B 44 (1988), 355-362.

[21] P. Frankl and H. Maehara, On the contact dimension of graphs, Discrete and Computational Geometry 3 (1988), 89-96.

[22] N. Garg, A deterministic $O(\log k)$-approximation algorithm for the sparsest cut, preprint, 1995.

[23] N. Garg, V. V. Vazirani, and M. Yannakakis, Approximate max-flow min-(multi)cut theorems and their applications, STOC 25 (1993), 698-707.

[24] A. A. Giannopoulos, On the Banach–Mazur distance to the cube, to appear in “Geometric Aspects of Functional Analysis”.

[25] M. X. Goemans and D. P. Williamson, .878-Approximation algorithms for MAX CUT and MAX 2SAT, STOC 26 (1994), 422-431.

[26] R. L. Graham and P. M. Winkler, On isometric embeddings of graphs, Trans. AMS 288 (1985), 527-536.

[27] W. Holsztynski, $\mathbb{R}^n$ as a universal metric space, Abstract 78T-G56, Notices AMS 25 (1978), A-367.

[28] T. C. Hu, Multicommodity network flows, Operations Research 11 (1963), 344-360.

[29] F. Jaeger, A survey of the cycle double cover conjecture, Annals of Discrete Math. 27 (1985), 1-12.

[30] W. B. Johnson and J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space, Contemporary Mathematics 26 (1984), 189-206.

[31] W. B. Johnson, J. Lindenstrauss, and G. Schechtman, On Lipschitz embedding of finite metric spaces in low dimensional normed spaces, in “Geometric Aspects of Functional Analysis”, (J. Lindenstrauss and V. Milman eds.) LNM 1267, Springer, Berlin – New York, 1987, 177-184.

[32] D. Karger, R. Motwani, and M. Sudan, Approximate graph coloring by semidefinite programming, FOCS 35 (1994), 2-13.

[33] A. K. Kelmans, Graph planarity and related topics, Contemporary Mathematics 147 (1993), 635-667.

[34] P. Klein, A. Agrawal, R. Ravi, and S. Rao, Approximation through multicommodity flow, FOCS 31 (1990), 726-737.

[35] P. Koebe, Kontaktprobleme der konformen Abbildung, Berichte Verhande. Sächs. Akad. Wiss. Leipzig, Math.-Phys. Klasse 88 (1936), 141-164.

[36] T. Leighton and S. Rao, An approximate max–flow min–cut theorem for uniform multicommodity flow problems with applications to approximation algorithms, FOCS 29 (1988), 422-431.

[37] M. Linial, N. Linial, H. Margalit, N. Tishby, and G. Yona, Work in progress, 1994.

[38] N. Linial, Local–global phenomena in graphs, Combinatorics, Probability and Computing 2 (1993), 491-503.

[39] N. Linial, L. Lovász, and A. Wigderson, Rubber bands, convex embeddings and graph connectivity, Combinatorica 8 (1988), 91-102.

[40] N. Linial, D. Peleg, Yu. Rabinovich, and M. Saks, Sphere packing and local majorities in graphs, The 2nd Israel Symp. on Theory and Computing Systems (1993), 141-149.

[41] N. Linial and M. Saks, Low diameter graph decompositions, SODA 2 (1991), 320-330. Journal version: Combinatorica 13 (1993), 441-454.

[42] J. H. van Lint and R. M. Wilson, "A Course in Combinatorics", Cambridge University Press, Cambridge, 1992.

[43] L. Lovász, On the Shannon capacity of a graph, IEEE Trans. Inf. Th. 25 (1979), 1-7.

[44] L. Lovász, M. Saks, and A. Schrijver, Orthogonal representations and connectivity of graphs, Linear Algebra Appl. 114-115 (1989), 439-454.

[45] J. Matoušek, Computing the center of planar point sets, in "Discrete and computational Geometry: papers from the DIMACS special year", (J. E. Goodman, R. Pollack, and W. Steiger eds.) AMS, Providence, 1991, 221-230.

[46] J. Matoušek, Note on bi-Lipschitz embeddings into normed spaces, Comment. Math. Univ. Carolinae 33 (1992), 51-55.

[47] G. L. Miller, S–H. Teng, and S. A. Vavasis, A unified geometric approach to graph separators, FOCS 32 (1991), 538-547.

[48] G. L. Miller and W. Thurston, Separators in two and three dimensions, STOC 22 (1990), 300-309.

[49] D. Peleg and A. Schäffer, Graph spanners, J. Graph Theory 13 (1989), 99-116.

[50] G. Pisier, "The Volume of Convex Bodies and Banach Space Geometry", Cambridge University Press, Cambridge, 1989.

[51] S. A. Plotkin and E. Tardos, Improved bounds on the max-flow min-cut ratio for multicommodity flows, STOC 25 (1993), 691-697.

[52] Yu. Rabinovich and R. Raz, On embeddings of finite metric spaces in graphs with a bounded number of edges, in preparation.

[53] J. Reiterman, V. Rödl, and E. Šiňajová, Geometrical embeddings of graphs, Discrete Mathematics 74 (1989), 291-319.

[54] N. Robertson and P. D. Seymour, Graph Minors I–XX, J. Comb. Th. B (1985 - present).

[55] B. Rothschild and A. Whinston, Feasibility of two–commodity network flows, Operations Research 14 (1966), 1121-1129.

[56] D. D. Sleator, R. E. Tarjan, and W. P. Thurston, Rotation distance, triangulations, and hyperbolic geometry, J. AMS 1 (1988), 647-681.

[57] W. P. Thurston, The finite Riemann mapping theorem, invited address, International Symposium in Celebration of the Proof of the Bieberbach Conjecture, Purdue University, 1985.

[58] D. J. A. Welsh, "Complexity: Knots, Colourings and Counting", LMS Lecture Note Series 186, Cambridge University Press, Cambridge, 1993.

[59] P. M. Winkler, Proof of the squashed cube conjecture, Combinatorica 3 (1983), 135-139.

[60] H. S. Witsenhausen, Minimum dimension embedding of finite metric spaces, J. Comb. Th. A 42 (1986), 184-199.

[61] I. M. Yaglom and V. G. Boltyanskiĭ, "Convex Figures", Holt, Rinehart and Winston, New York, 1961.