Chapter 8
Metric representations

Nachdiplomvorlesung by László Lovász
ETH Zürich, Spring 2014

Given a graph, we would like to embed it in a Euclidean space so that the distances between nodes in the graph should be the same as, or at least close to, the geometric distance of the representing vectors. Our goal is to illustrate how to embed graphs into euclidean spaces (or other standard normed spaces), and, mainly, how to use such embeddings to prove graph-theoretic theorems and to design algorithms.

It is not hard to see that every embedding will necessarily have some distortion in non-trivial cases. For example, the "claw" $K_{1,3}$ cannot be embedded isometrically in any dimension. We get more general and useful results if we study embeddings where the distances may change, but in controlled manner. To emphasize the difference, we will distinguish *distance preserving* (isometric) and *distance respecting* embeddings.

The complete k-graph can be embedded isometrically in a euclidean space with dimension $k-1$, but not in lower dimensions. There is often a trade-off between dimension and distortion. This motivates our concern with the dimension of the space in which we represent our graph.

Several of the results are best stated in the generality of finite metric spaces. Recall that a metric space is a set V endowed with a distance function $d: V \times V \to \mathbb{R}_+$ such that $d(u,v) = 0$ if and only if $u = v$, $d(v,u) = d(u,v)$, and $d(u,w) \le d(u,v) + d(v,w)$ for all $u,v,w \in V$. There is a large literature of embeddings of one metric space into another so that distances are preserved (isometric embeddings) or at least not distorted too much (we call such embeddings, informally, "distance respecting"). These results are often very combinatorial, and have important applications to graph theory and combinatorial algorithms (see [6, 14]).

Here we have to restrict our interest to embeddings of (finite) metric spaces into basic normed spaces like euclidean or ℓ_1 spaces. We discuss some examples of isometric embeddings, then we construct important distance respecting embeddings, and then we introduce the fascinating notion of volume-respecting embeddings. We show applications to important graph-theoretic algorithms. Several facts from high dimensional geometry are used, which are collected in the last section.

1 Isometric embeddings

1.1 Supremum norm

Theorem 1.1 (Frechet) *Any finite metric space can be embedded isometrically into* $(\mathbb{R}^k, \ell_\infty)$ *for some finite* k.

The theorem is valid for infinite metric spaces as well (when infinite dimensional L_∞ spaces must be allowed for the target space), with a slight modification of the proof. We describe the simple proof for the finite case, since its idea will be used later on.

Proof. Let (V, d) be a finite metric space, and consider the embedding

$$\mathbf{u}_v = (d(v,i) : \ i \in V) \qquad (v \in V).$$

Then by the triangle inequality,

$$\|\mathbf{u}_u - \mathbf{u}_v\|_\infty = \max_i |d(u,i) - d(v,i)| \le d(u,v).$$

On the other hand, the coordinate corresponding to $i = v$ gives

$$\|\mathbf{u}_u - \mathbf{u}_v\|_\infty \ge |d(u,v) - d(v,v)| = d(u,v),$$

showing that the embedding is isometric. □

1.2 Manhattan distance

Another representation with special combinatorial significance is an representation in $\mathbb{R}^n$ with the ℓ_1-distance (often called Manhattan distance). It is convenient to allow *semimetrics*, i.e., symmetric functions $d : \ V \times V \to \mathbb{R}_+$ satisfying $d(i,i) = 0$ and $d(i,k) \le d(i,j) + d(j,k)$ (the triangle inequality), but for which $d(u,v) = 0$ is allowed even if $u \ne v$. An ℓ_1-representation of a finite semimetric space (V,d) is a mapping $\mathbf{u} : \ V \to \mathbb{R}^m$ such that $d(i,j) = \|\mathbf{u}_i - \mathbf{u}_j\|_1$. We say that (V,d) is ℓ_1*-representable*, or shortly an ℓ_1*-semimetric*.

For a fixed underlying set V, all ℓ_1-semimetrics form a closed convex cone: if $i \mapsto \mathbf{u}_i \in \mathbb{R}^m$ is an ℓ_1-representation of (V,d) and $i \mapsto \mathbf{u}'_i \in \mathbb{R}^{m'}$ is an ℓ_1-representation of (V,d'), then

$$i \mapsto \begin{pmatrix} \alpha \mathbf{u} \\ \alpha' \mathbf{u}' \end{pmatrix}$$

is an ℓ_1-representation of $\alpha d + \alpha' d'$ for $\alpha, \alpha' \ge 0$.

An ℓ_1-semimetric of special interest is the semimetric defined by a subset $S \subseteq V$,

$$\nabla_S(i,j) = \mathbb{1}(|\{i,j\} \cap S| = 1),$$

which I call a 2*-partition semimetric*. (These are often call “cut-semimetrics”, but I don’t want to use this term since “cut norm” and “cut distance” are used in this book in a different sense.)

Lemma 1.2 *A finite semimetric space (V, d) is an ℓ_1-semimetric if and only if it can be written as a nonnegative linear combination of 2-partition semimetrics.*

Proof. Every 2-partition semimetric ∇_S can be represented in the 1-dimensional space by the map $\mathbb{1}_S$. It follows that every nonnegative linear combination of 2-partition semimetrics on the same underlying set is an ℓ_1-semimetric. Conversely, every ℓ_1-semimetric is a sum of ℓ_1-semimetrics representable in $\mathbb{R}^1$ (just consider the coordinates of any ℓ_1-representation). So it suffices to consider ℓ_1-semimetrics (V, d) for which there is a representation $i \mapsto u_i \in \mathbb{R}$ such that $d(i, j) = |u_i - u_j|$. We may assume that $V = [n]$ and $u_1 \le \cdots \le u_n$, then for $i < j$,

$$d(i, j) = \sum_{k=1}^{n-1} (u_{k+1} - u_k) \nabla_{\{1,\dots,k\}}$$

expresses d as a nonnegative linear combination of 2-partition semimetrics. □

2 Distance respecting representations

As we have seen, the possibility of a geometric representation that reflects the graph distance exactly is limited, and therefore we allow distortion. Let $F : V_1 \to V_2$ be a mapping of the metric space (V_1, d_1) into the metric space (V_2, d_2). We define the *distortion* of F as

$$\max_{u,v \in V_1} \frac{d_2(F(u), F(v))}{d_1(u, v)} \Big/ \min_{u,v} \frac{d_2(F(u), F(v))}{d_1(u, v)}.$$

Note that to have finite distortion, the map F must be injective. The distortion does not change if all distances in one of the metric spaces are scaled by the same factor. So if we are looking for embeddings in a Banach space, then we may consider embeddings that are contractive, i.e., $d_2(F(u), F(v)) \le d_1(u, v)$ for all $u, v \in V_1$.

2.1 Dimension reduction

In many cases, we can control the dimension of the ambient space based on general principles, and we start with a discussion of these. The dimension problem is often easy to handle, due to a fundamental lemma [8].

Lemma 2.1 (Johnson–Lindenstrauss) *For every $0 < \varepsilon < 1$, every n-point set $S \subset \mathbb{R}^n$ can be mapped into $\mathbb{R}^d$ with $d < (80 \ln n)/\varepsilon^2)$ with distortion at most ε.*

More careful computation gives a constant 8 instead of 80.

Proof. Orthogonal projection onto a random d-dimensional subspace L does the job. First, let us see what happens to the distance of a fixed pair of points $\mathbf{x}, \mathbf{y} \in S$. Instead of projecting the fixed segment of length $|\mathbf{x} - \mathbf{y}|$ on a random subspace, we can project a random vector of the same length on a fixed subspace. Then Lemma 4.1 implies that with probability at least $1-4e^{-\varepsilon^2 d/4}$, the projections $\mathbf{x}'$ and $\mathbf{y}'$ of every fixed pair of points $\mathbf{x}$ and $\mathbf{x}'$ satisfy

$$\frac{1}{1+\varepsilon}\sqrt{\frac{d}{n}} \le \frac{|\mathbf{x}'-\mathbf{y}'|}{|\mathbf{x}-\mathbf{y}|} \le (1+\varepsilon)\sqrt{\frac{d}{n}}.$$

It follows that with probability at least $1-\binom{n}{2}4e^{-\varepsilon^2 d/4}$, this inequality holds simultaneously for all pairs $x, y \in S$, and then the distortion of the projection is at most $(1+\varepsilon)^2 < 1+3\varepsilon$. Replacing ε by $\varepsilon/3$ and choosing d as in the Theorem, this probability is positive, and the lemma follows. □

2.2 Embedding with small distortion

We describe embedding of a general metric space into a euclidean space, due to Bourgain [1], which has low distortion (and other very interesting and useful properties, as we will see later).

Theorem 2.2 *Every metric space with n points can be embedded into the euclidean space $\mathbb{R}^d$ with distortion $O(\log n)$ and dimension $d = O(\log n)$.*

To motivate the construction, let us recall Frechet's Theorem 1.1: we embed a general metric space isometrically into ℓ_∞ by assigning a coordinate to each point w, and considering the representation

$$\mathbf{x}: \ i \mapsto (d(w,i): \ w \in V).$$

If we want to represent the metric space by euclidean metric with a reasonably small distortion, this construction will not work, since it may happen that all points except u and v are at the same distance from u and v, and once we take a sum instead of the maximum, the contribution from u and v will become negligible. The remedy will be to take distances from sets rather than from points; it turns out that we need sets with sizes of all orders of magnitude, and this is where the logarithmic factor is lost.

We start with describing a simpler version, which is only good "on the average". For a while, it will be more convenient to work with ℓ_1 distances instead of euclidean distances. Both of these deviations from our goal will be easy to fix.

Let (V,d) be a metric space on n elements. Let $m = \lceil \log n \rceil$, and for every $1 \le i \le m$, choose a random subset $\mathbf{A}_i \subseteq V$, putting every element $v \in V$ into $\mathbf{A}_i$ with probability 2^{-i}. Let $d(v,\mathbf{A}_i)$ denote the distance of node v from the set $\mathbf{A}_i$ (if $\mathbf{A}_i = \emptyset$, we set $d(v,\mathbf{A}_i) = K$ for some very large K, the same for every v). Consider the mapping $\mathbf{x}: \ V \to \mathbb{R}^N$ defined by

$$\mathbf{x}_u = \big(d(u,\mathbf{A}_1), \dots, d(u,\mathbf{A}_m)\big). \tag{1}$$

We call this the *random subset representation* of (V, d). Note that this embedding depends on the random choice of the sets A_i, and so we have to make probabilistic statements about it.

Lemma 2.3 *The random subset representation* (1) *satisfies*

$$\|\mathbf{x}_u - \mathbf{x}_v\|_1 \le md(u,v), \quad \textit{and} \quad \mathsf{E}(\|\mathbf{x}_u - \mathbf{x}_v\|_1) \ge \frac{1}{16} d(u,v).$$

for every pair of points $u, v \in V$.

Proof. By the triangle inequality

$$|d(u,\mathbf{A}_i) - d(v,\mathbf{A}_i)| \le d(u,v), \tag{2}$$

and hence

$$\|\mathbf{x}_u - \mathbf{x}_v\|_1 = \sum_{i=1}^{m} |d(u,\mathbf{A}_i) - d(v,\mathbf{A}_i)| \le md(u,v).$$

To prove the lower bound, fix two points u and v. Let $(u_0 = u, u_1, \dots, u_{n-1})$ be the points in V ordered by increasing distance from u (so that $u_0 = u$), and let $(v_0 = v, v_1, \dots, v_{n-1})$ be defined analogously. We break ties arbitrarily. Define $\rho_i = \max(d(u, u_{2^i}), d(v, v_{2^i}))$ as long as $\rho_i < d(u,v)/2$. Let k be the first i with $\max(d(u, u_{2^k}), d(v, v_{2^k})) \ge d(u,v)/2$, and define $\rho_k = d(u,v)/2$.

Focus on any $i < k$, and let (say) $\rho_i = d(u, u_{2^i})$. If $\mathbf{A}_i$ does not select any point from $(u_0, u_1, \dots, u_{2^i-1})$, but selects a point from $(v_0, v_1, \dots, v_{2^{i-1}})$, then

$$|d(u,S_i) - d(v,S_i)| \ge d(u,u_{2^i}) - d(v,,v_{2^{i-1}}) \ge \rho_i - \rho_{i-1}.$$

The sets $(u_0, u_1, \dots, u_{2^i-1})$ and $(v_0, v_1, \dots, v_{2^{i-1}})$ are disjoint, and so the probability that this happens is

$$\left(1 - \left(1 - \frac{1}{2^i}\right)^{2^{i-1}}\right)\left(1 - \frac{1}{2^i}\right)^{2^i} \ge \frac{1}{8}.$$

So the contribution of $\mathbf{A}_i$ to $\mathsf{E}\big(|d(u,\mathbf{A}_i) - d(v,\mathbf{A}_i)|\big)$ is at least $(\rho_i - \rho_{i-1})/8$. It is easy to check that this holds for $i = k$ as well. Hence

$$\mathsf{E}(\|\mathbf{x}_u - \mathbf{x}_v\|_1) = \sum_{i=1}^{m} \mathsf{E}|d(u,\mathbf{A}_i) - d(v,\mathbf{A}_i)| \ge \frac{1}{8}\sum_{i=0}^{k}(\rho_i - \rho_{i-1}) = \frac{1}{8}\rho_k = \frac{1}{16} d(u,v), \tag{3}$$

which proves the Lemma. $\square$

Proof of Theorem 2.2. The previous lemma only says that the random subset representation does not shrink distances too much "on the average". Equation 3 suggests the remedy:

the distance $\|\mathbf{x}_u - \mathbf{x}_v\|_1$ is the sum of m independent bounded random variables, and hence it will be close to its expectation with high probability, if m is large enough. Unfortunately, our m as chosen is not large enough to be able to claim this simultaneously for all u and v; but we can multiply m basically for free. To be more precise, let us generate N independent copies $\mathbf{x}^1, \dots, \mathbf{x}^N$ of the representation (1), and consider the mapping $\mathbf{y}: V \to \mathbb{R}^{Nm}$, defined by

$$\mathbf{y}_u = (\mathbf{x}_u^1, \dots, \mathbf{x}_u^N). \tag{4}$$

Then trivially

$$\|\mathbf{y}_u - \mathbf{y}_v\|_1 \le Nmd(u,v) \quad \text{and} \quad \mathsf{E}(\|\mathbf{y}_u - \mathbf{y}_v\|_1) = N\mathsf{E}(\|\mathbf{x}_u - \mathbf{x}_v\|_1) \ge N\frac{1}{16}d(u,v).$$

For every u and v, $(1/N)\|\mathbf{y}_u - \mathbf{y}_v\|_1 \to \mathsf{E}(\|\mathbf{x}_u - \mathbf{x}_v\|_1 \ge \frac{1}{16}d(u,v)$ by the Law of Large Numbers with probability 1, and hence if N is large enough, then with high probability

$$\|\mathbf{y}_u - \mathbf{y}_v\|_1 \ge \frac{N}{20}d(u,v) \tag{5}$$

holds for all pairs u, v. So $\mathbf{y}$ is a representation in ℓ_1 with distortion at most $20m$.

To get results for ℓ_p-distance instead of the ℓ_1-distance (in particular, for the the euclidean distance), we first note that the upper bound

$$\|\mathbf{y}_u - \mathbf{y}_v\|_p \le \big\|(d(u,v), \dots, d(u,v))\big\|_p = (Nm)^{1/p}d(u,v)$$

follows similarly from the triangle inequality as the analogous bound for the ℓ_1-distance. The lower bound is similarly easy, using (5):

$$\|\mathbf{y}_u - \mathbf{y}_v\|_p \ge (Nm)^{\frac{1}{p}-1}\|\mathbf{y}_u - \mathbf{y}_v\|_1 \ge \frac{(Nm)^{1/p}}{20m}d(u,v) \tag{6}$$

with high probability for all $u, v \in V$. The ratio of the upper and lower bounds is $20m = O(\log n)$.

We are not yet done, since choosing a large N will result in a representation in a very high dimension. An easy way out is to invoke the Johnson–Lindenstrauss Lemma 2.1 (in other words, to apply a random projection). Alternately, we could estimate the concentration of $(1/N)\|\mathbf{y}_u - \mathbf{y}_v\|_1$ using the Chernoff-Hoeffding Inequality. $\square$

Linial, London and Rabinovitch [10] showed how to construct an embedding satisfying the conditions in Theorem 2.2 algorithmically, and gave the application to be described in Section 2.3. Matoušek [13] showed that if the metric space is the graph distance in an expander graph, then Bourgain's embedding is essentially optimal, in the sense that every embedding in a euclidean space has distortion $\Omega(\log n)$.

2.3 Multicommodity flows and approximate Max-Flow-Min-Cut

We continue with an important algorithmic application of the random subset representation.

A fundamental result in the theory of multicommodity flows is the theorem of Leighton and Rao [9]. Stated in a larger generality, as proved by Linial, London and Rabinovich [10], it says the following. Suppose that we have a multicommodity flow problem on a graph G. This means that we are given k pairs of nodes (s_i, t_i) $(i = 1, \dots, k)$, and for each such pair, we are given a *demand* $d_i \geq 0$. Every edge e of the graph has a *capacity* $c_e \geq 0$. We would like to design a flow f^i from s_i to t_i of value d_i for every $1 \leq i \leq k$, so that for every edge the capacity constraint is satisfied.

We state the problem more precisely. We need a reference orientation of G; let $\overrightarrow{E}$ denote the set of these oriented edges, and for each set $S \subseteq V$, let $\nabla^+(S)$ and $\nabla^-(S)$ denote the set of edges leaving and entering S, respectively. An (s,t)-*flow* is a function $f : \overrightarrow{E}(G) \to \mathbb{R}$ that conserves material:

$$\sum_{e \in \nabla^+(v)} f(e) = \sum_{e \in \nabla^-(v)} f(e) \quad \text{for all } v \in V(G) \setminus \{s,t\}. \tag{7}$$

Since the orientation of the edges serves only as a reference, we allow positive or negative flow values; if an edge is reversed, the flow value changes sign. The number

$$\mathrm{val}(f) = \sum_{e \in \nabla^+(s)} f(e) - \sum_{e \in \nabla^-(s)} f(e) = \sum_{e \in \nabla^-(t)} f(e) - \sum_{e \in \nabla^+(t)} f(e) \tag{8}$$

is the *value* of the flow.

So we want an (s_i, t_i)-flow f^i for every $1 \leq i \leq k$, with prescribed values d_i, so that the total flow through the edge does not exceed the capacity of the edge:

$$\sum_{i=1}^{k} |f^i(e)| \leq c_e \quad \text{for all } e \in \overrightarrow{E}(G) \tag{9}$$

(this clearly does not depend on which orientation of the edge we consider).

Let us hasten to point out that the solvability of a multicommodity flow problem is just the feasibility of a linear program (we treat the values $f(e)$ $(e \in \overrightarrow{E}(G))$ as variables). So the multicommodity flow problem is polynomial time solvable. We can also apply the Farkas Lemma, and derive a necessary and sufficient condition for solvability. If you work out the dual, very likely you will get a condition that is not transparent at all; however, a very nice form was discovered by Iri [7] and by Shahrokhi and Matula [15], which fits particularly well into the topic of this book.

Consider a semimetric D on $V(G)G$. Let us describe an informal (physical) derivation of the conditions. Think of an edge $e = uv$ as a pipe with cross section c_e and length $D(e) = D(u,v)$. Then the total volume of the system is $\sum_e c_d D(e)$. If the multicommodity problem is feasible, then every flow f^i occupies a volume of $\mathrm{val}(f^i)D(s_i,t_i)$, and hence

$$\sum_i d_i D(s_i,t_i) \le \sum_e c_d D(e). \tag{10}$$

These conditions are also sufficient:

Theorem 2.4 *Let $G=(V,E)$ be a graph, let $s_i,t_i \in V$, $d_i \in \mathbb{R}_+$ $(i=1,\ldots,k)$, and $c_e \in \mathbb{R}_+$ $(e \in E)$. Then there exist (s_i,t_i)-flows $(f^i:\ i=1,\ldots,k)$ satisfying the the demand conditions* $\mathrm{val}(f^i)=d_i$ *and the capacity constraints* (9) *if and only if* (10) *holds for every semimetric D in V.*

We leave the exact derivation of the necessity of the condition, as well as the proof of the converse, to the reader.

For the case $k=1$, the Max-Flow-Min-Cut Theorem of Ford and Fulkerson gives a simpler condition for the existence of a flow with given value, in terms of cuts. Obvious cut-conditions provide a system of necessary conditions for the problem to be feasible in the general case as well. If the multicommodity flows exist, then for every $S \subseteq V(G)$, we must have

$$\sum_{e\in\nabla^+(S)} c_e \ge \sum_{i:\ S\cap\{s_i,t_i\}=1} d_i. \tag{11}$$

In the case of one or two commodities ($k \le 2$), these conditions are necessary and sufficient; this is the content of the Max-Flow-Min-Cut Theorem ($k=1$) and the Gomory–Hu Theorem ($k=2$). However, for $k \ge 3$, the cut conditions are not sufficient any more for the existence of multicommodity flows. The theorem of Leighton and Rao asserts that *if the cut-conditions are satisfied, then relaxing the capacities by a factor of $O(\log n)$, the problem becomes feasible.* The relaxation factor was improved by Linial, London and Rabinovich to $O(\log k)$; we state the result in this tighter form, but for simplicity of presentation prove the original (weaker) form.

Theorem 2.5 *Suppose that for a multicommodity flow problem, the cut conditions* (11) *are satisfied. Then replacing every edge capacity c_e by $10c_e\log k$, the problem becomes feasible.*

Proof. Using Theorem 2.4, it suffices to prove that for every semimetric D on $V(G)$,

$$\sum_i d_i D(s_i,t_i) \le 10(\log n)\sum_e c_d D(e). \tag{12}$$

The cut conditions imply the validity of semimetric conditions (10) for 2-partition semimetrics, which in turn imply their validity for semimetrics that are nonnegative linear combinations of 2-partition semimetrics. We have seen that these are exactly the ℓ_1-semimetrics. By Bourgain's Theorem 2.2, there is an ℓ_1-metric D' such that

$$D'(u,v) \le D(u,v) \le 20(\log n)D'(u,v).$$

We know that D' satisfies the semimetric condition (10), and hence D satisfies the relaxed semimetric conditions (12). $\square$

3 Bandwidth and respecting the volume

A very interesting extension of the notion of small distortion was formulated by Feige [4]. We want an embedding of a metric space in a euclidean space such that is contractive, and at the same time volume respecting up to size s, which means that every set of at most s nodes spans a simplex whose volume is almost as large as possible.

Obviously, the last condition needs an explanation, and for this, we will have to define a certain "volume" of a metric space.

3.1 Tree-volume

Let (V, d) be a finite metric space with $|V| = n$. For a spanning tree T on $V(T) = V$, we define $\Pi(T) = \prod_{uv \in E(T)} d(u, v)$, and we define the *tree-volume* of (V, d) by

$$\mathsf{tvol}(V) = \mathsf{tvol}(V, d) = \frac{1}{k!} \min_T \Pi(T)$$

Since we only consider one metric, denoted by d, in this section, we will not indicate it in $\mathsf{tvol}(V, d)$ and similar quantities to be defined below. We will, however, change the underlying set, so we keep V in the notation. Note that the tree T that gives here the minimum is also minimizing the total length of edges (this follows from the fact that it can be found by the Greedy Algorithm).

The following lemma relates the volume of a simplex in a vector labeling $\mathbf{x} : V \to \mathbb{R}^n$ with the tree-volume. For any subset $S \subseteq V$, let $\mathrm{vol}_{\mathbf{x}}(S)$ denote the $(|S|-1)$-dimensional volume of $\mathrm{conv}(\mathbf{x}(S))$.

Lemma 3.1 *For every contractive map $\mathbf{x} : V \to \mathbb{R}^n$ of a metric space (V, d), $\mathrm{vol}_{\mathbf{x}}(V) \le \mathsf{tvol}(V)$.*

Proof. We prove by induction on $n = |V|$ that for any tree T on V,

$$\mathrm{vol}_{\mathbf{x}}(V) \le \frac{1}{(n-1)!} \Pi(T).$$

For $n = 2$ this is obvious. Let $n > 2$, and let $i \in V$ be a node of degree 1 in T, let j be the neighbor of i in T, and let h be the distance of i from the hyperplane containing $\mathbf{x}(V \setminus \{i\})$. Then by induction,

$$\begin{aligned}\mathrm{vol}_{\mathbf{x}}(V) &= \frac{h}{n-1} \mathrm{vol}_{\mathbf{x}}(V \setminus \{i\}) \le \frac{d(i,j)}{n-1} \mathrm{vol}_{\mathbf{x}}(V \setminus \{i\}) \\ &\le \frac{d(i,j)}{n-1} \frac{1}{(n-2)!} \Pi(T - i) = \frac{1}{(n-1)!} \Pi(T).\end{aligned}$$

This completes the proof. $\square$

The main result about tree-volume (Feige [4]) is that the upper bound given in Lemma 3.1 can be attained, up to a polylogarithmic factor, for all small subsets simultaneously. To be more precise, let (V,d) be a finite metric space, let $k \geq 2$, and let $\mathbf{x}: V \to \mathbb{R}^n$ be any contractive vector labeling. We define the *volume-distortion for k-sets* of $\mathbf{x}$ to be

$$\sup_{\substack{S\subseteq V\\|S|=k}} \left(\frac{\mathsf{tvol}(S)}{\mathrm{vol}_{\mathbf{x}}(S)}\right)^{\frac{1}{k-1}}.$$

(The exponent $1/(k-1)$ is a natural normalization, since $\mathsf{tvol}(S)$ is the product of $k-1$ distances.) Volume-distortion for 2-sets is just the distortion defined before.

Theorem 3.2 *Every finite metric space (V,d) with n elements has a contractive map $\mathbf{x}: V \to \mathbb{R}^d$ with $d = O((\log n)^5)$, with volume-distortion for k-sets $O(\sqrt{k}\log n)$ for every $k \leq \log n$.*

The proof is long and difficult, and the interested reader is referred to the paper of Feige [4].

3.2 Bandwidth and density

Our goal is to use volume-respecting embeddings to design an approximation algorithm for the bandwidth of a graph, but first we we have to discuss some basic properties of bandwidth.

The *bandwidth* $\mathsf{bw}(G)$ of a graph $G = (V,E)$ is defined as the smallest integer b such that the nodes of G can be labeled by $1,\dots,n$ so that $|i-j| \leq b$ for every edge ij. The number refers to the fact that for this labeling of the nodes, all 1's in the adjacency matrix will be contained in a band of width $2b+1$ around the main diagonal.

The bandwidth of a graph is NP-hard to compute, or even to approximate within a constant factor [3]. As an application of volume-respecting embeddings, we describe an algorithm that finds a polylogarithmic approximation of the bandwidth of a graph in polynomial time. The ordering of the nodes which approximates the bandwidth will obtained through a random projection of the representation to the line, in a fashion similar to the Goemans–Williamson algorithm.

We can generalize the notion of bandwidth to any finite matric space (V,d): it is the smallest real number b such that there is a bijection $f: V \to [n]$ such that $|f(i)-f(j)| \leq bd(i,j)$ for all $i,j \in V$. It takes a minute to realize that this notion does generalize graph bandwidth: the definition of bandwidth requires this condition only for the case when ij is an edge, but this already implies that it holds for every pair of nodes due to the definition of graph distance. (We do loose the motivation for the word "bandwidth".)

We need some preliminary observations about the bandwidth. It is clear that if there is a node of degree D, then $\mathsf{bw}(G) \geq D/2$. More generally, if there are k nodes on the graph at distance at most t from a node v (not counting v), then $\mathsf{bw}(G) \ge k/(2t)$. This observation generalizes to metric spaces. We define the *local density* of a metric space (V,d) by

$$\mathsf{d}_{\rm loc}(V) = \max_{v,t} \frac{|B(v,t)|-1}{t}.$$

Then

$$\mathsf{bw}(V) \ge \frac{1}{2}\mathsf{d}_{\rm loc}(V). \tag{13}$$

It is not hard to see that equality does not hold here (see Exercise 4.8), and the ratio $\mathsf{bw}(G)/\mathsf{d}_{\rm loc}(G)$ can be as large as $\log n$ for appropriate graphs.

We need the following related quantity, which we call the *harmonic density*:

$$\mathsf{d}_{\rm harm}(V) = \max_{v\in V} \sum_{u\in V\setminus\{v\}} \frac{1}{d(u,v)}. \tag{14}$$

This quantity differs from the local density by at most a logarithmic factor:

Lemma 3.3 *We have*

$$\mathsf{d}_{\rm loc}(V) \le \mathsf{d}_{\rm harm}(V) \le (1+\ln n)\mathsf{d}_{\rm loc}(V).$$

Proof. For an appropriate $t > 0$ and $v \in V$,

$$\mathsf{d}_{\rm loc}(V) = \frac{|B(v,t)|-1}{t} \le \sum_{u\in B(v,t)\setminus\{v\}} \frac{1}{d(u,v)} \le \sum_{u\in V\setminus\{v\}} \frac{1}{d(u,v)} = \mathsf{d}_{\rm harm}(V).$$

On the other hand, $\mathsf{d}_{\rm harm}(V)$ is the sum of $n-1$ positive numbers, among which the k-th largest is at most $\mathsf{d}_{\rm loc}(V)/k$ by the definition of the local density. Hence

$$\mathsf{d}_{\rm harm}(V) \le \sum_{k=1}^{n-1} \frac{\mathsf{d}_{\rm loc}(V)}{k} \le (1+\ln n)\mathsf{d}_{\rm loc}(V).$$

□

The harmonic density is related to the tree-volume:

Lemma 3.4

$$\sum_{S\in\binom{V}{k}} \frac{1}{\mathsf{tvol}(S)} \le n(k-1)!(4\mathsf{d}_{\rm harm}(V))^{k-1}.$$

Proof. Let H be the complete graph on V with edgeweights $\beta_{ij} = 1/d(i,j)$, then

$$\sum_{S\in\binom{V}{k}} \frac{1}{\mathsf{tvol}(S)} \le \sum_{S\in\binom{V}{k}} \sum_{\substack{T \text{ tree} \\ V(T)=S}} \frac{(k-1)!}{\Pi(T)} \le \sum_{T} (k-1)!\,\hom(T,H), \tag{15}$$

where the summation runs over isomorphism types of trees on k nodes. To estimate homomorphism numbers, we use the fact that for every edge-weighting β of the complete graph K_n in which the edgeweights $\beta_{ij} \geq 0$ satisfy $\sum_j \beta_{ij} \leq D$ for every node i, we have

$$\hom(T, H) \leq nD^{k-1}. \tag{16}$$

For integral edgeweights, we can view the left side as counting homomorphisms into a multigraph, and then the inequality follows by simple computation: fixing any node of T as its root, it can be mapped in n ways, and then fixing any search order of the nodes, each of the other nodes can be mapped in at most D ways. This implies the bound for rational edgeweights by scaling, and for real edgeweights, by approximating them by rationals.

By (16), each term on the right side of (15) is at most $n(k-1)!\mathsf{d}_{\rm harm}(V)^{k-1}$, and it is well known that the number of isomorphism types of trees on k nodes is bounded by 4^{k-1} (see e.g. [11], Problem 4.18), which proves the Lemma. □

3.3 Approximating the bandwidth

Now we come to the main theorem describing the algorithm to get an approximation of the bandwidth.

Theorem 3.5 *Let $G = (V, E)$ be a graph on n nodes, let $k = \lceil \ln n \rceil$, and let $\mathbf{x} :\ V \to \mathbb{R}^d$ be a contractive representation such that for each set S with $|S| = k$, the volume of the simplex spanned by $\mathbf{x}(S)$ is at least $\mathsf{tvol}(S)/\eta^{k-1}$ for some $\eta \geq 1$. Project the representing points onto a random line L through the origin, and label the nodes by $1, \dots, n$ according to the order of their projections on the line. Then with high probability, we have $|i-j| \leq 200k\eta\mathsf{d}_{\rm harm}(G)$ for every edge ij.*

Proof. Let $\mathbf{y}_i$ be the projection of $\mathbf{x}_i$ in L. Let ij be an edge; since $\mathbf{x}$ is contractive we have $|\mathbf{x}_i - \mathbf{x}_j| \leq 1$, and hence, with high probability, $|\mathbf{y}_i - \mathbf{y}_j| \leq 2|\mathbf{x}_i - \mathbf{x}_j|/\sqrt{d} \leq 2/\sqrt{d}$ for all edges ij, and we will assume below that this is indeed the case.

We call a set $S \subseteq V$ *bad*, if $\operatorname{diam}(\mathbf{y}(S)) \leq 2/\sqrt{d}$. So every edge is bad, but we are interested in bad k-sets. Let $T = \{i, i+1, \dots, j\}$ and $|T| = B$. Then all the points $\mathbf{y}_u$, $u \in T$, are between $\mathbf{y}_i$ and $\mathbf{y}_j$, and hence $\operatorname{diam}(\mathbf{y}(T)) \leq 2/\sqrt{d}$, so T is bad.

By Lemma 4.2 (with d in place of n and $k-1$ in place of d),

$$\begin{aligned}\mathsf{P}(S \text{ is bad}) &\leq \Big(\frac{2e}{\sqrt{d}}\sqrt{\frac{d}{k-1}}\Big)^{k-1} \frac{\pi_{k-1}}{\mathrm{vol}_{\mathbf{x}}(S)}\big(1 + \ln\ \operatorname{diam}(\mathbf{x}(S))\big) \\ &\leq \Big(\frac{2e}{\sqrt{k-1}}\Big)^{k-1} \frac{\pi_{k-1}}{\mathrm{vol}_{\mathbf{x}}(S)}(1 + \ln\ n) \leq \Big(\frac{2e}{\sqrt{k-1}}\Big)^{k-1} \frac{\pi_{k-1}}{\mathrm{vol}_{\mathbf{x}}(S)} n.\end{aligned}$$

Using the condition that $\mathbf{x}$ is volume-respecting, we get

$$\mathsf{P}(S \text{ is bad}) \leq \Big(\frac{2e\eta}{\sqrt{k-1}}\Big)^{k-1} \frac{\pi_{k-1} n}{\mathsf{tvol}(S)}.$$

Using Corollary 3.4, the expected number of bad sets is

$$\Big(\frac{2e\eta}{\sqrt{k-1}}\Big)^{k-1}\sum_{S\in\binom{V}{k}}\frac{\pi_{k-1}n}{\mathsf{tvol}(S)}\leq\Big(\frac{2e\eta}{\sqrt{k-1}}\Big)^{k-1}\pi_{k-1}(k-1)!n^2(4\mathsf{d}_{\mathrm{harm}}(V))^{k-1}.$$

On the other hand, we have seen that the number of bad sets is at least $\binom{B+1}{k}$, so

$$\binom{B+1}{k}\leq\Big(\frac{8e\eta}{\sqrt{k-1}}\Big)^{k-1}\pi_{k-1}(k-1)!n^2\mathsf{d}_{\mathrm{harm}}(V)^{k-1}.$$

Using the trivial bound $\binom{B+1}{k}\geq(B-k)^{k-1}/(k-1)!$, we get

$$\begin{aligned}B&\leq k+(k-1)!^{2/(k-1)}\frac{8e\eta}{\sqrt{k-1}}\pi_{k-1}^{1/(k-1)}n^{2/(k-1)}\mathsf{d}_{\mathrm{harm}}(V)\\&\leq 200k\eta\mathsf{d}_{\mathrm{harm}}(V).\end{aligned}$$

This completes the proof. □

Corollary 3.6 *For every finite metric space (V,d) with n points, its bandwidth and local density satisfy*

$$\frac{1}{2}\mathsf{d}_{\mathrm{loc}}(V)\leq\mathsf{bw}(V)\leq O((\log n)^4)\mathsf{d}_{\mathrm{loc}}(V).$$

4 Geometric lemmas

The proofs above, as well as many other proofs on the boundary of combinatorics and geometry, depend on elementary inequalities involving volumes of spheres, caps, and their projections, which are sometimes surprisingly nontrivial. In this section we collected some of these facts.

Let $\mathbf{x}'$ denote the projection of a vector $\mathbf{x}\in\mathbb{R}^n$ onto the first d coordinates. It is easy to see that if $\mathbf{x}$ is chosen uniformly at random from the unit sphere S^{n-1}, then

$$\mathsf{E}(|\mathbf{x}'|^2)=\frac{d}{n}, \tag{17}$$

The length of $\mathbf{x}'$ is highly concentrated around $\sqrt{d/n}$, shown by the following estimates.

Lemma 4.1 *If $1\leq d\leq n$, $\varepsilon>0$, and $\mathbf{x}\in S^{n-1}$ is chosen uniformly at random, then*

$$\frac{1}{1+\varepsilon}\sqrt{\frac{d}{n}}\leq|\mathbf{x}'|\leq(1+\varepsilon)\sqrt{\frac{d}{n}}$$

with probability at least $1-4e^{-\varepsilon^2d/4}$.

Proof. We can generate a random unit vector by generating n independent standard Gaussian variables $X_1,\dots,X_n$, and normalizing:

$$\mathbf{x} = \frac{1}{\sqrt{X_1^2+\dots+X_n^2}}(X_1,\dots,X_n).$$

Here $X_1^2+\dots+X_n^2$ is from a chi-squared distribution with parameter n, and many estimates are known for its distribution, of which we use (see e.g. Massart [12])

$$\mathsf{P}\big(\big|X_1^2+\dots+X_n^2-n\big| > \varepsilon n\big) \le 2e^{-\varepsilon^2 n/8}. \tag{18}$$

We have

$$|\mathbf{x}'|^2 = \frac{X_1^2+\dots+X_d^2}{X_1^2+\dots+X_n^2}.$$

Applying (18) to the numerator and denominator, we get the bounds in the Lemma. □

We also need the following "tail bound", an estimate on the probability that $\mathbf{x}'$ is very short.

Lemma 4.2 *If $1 \le d \le n-2$, and $\mathbf{x} \in S^{n-1}$ is chosen uniformly at random, then*

$$\mathsf{P}\Big(|\mathbf{x}| < \varepsilon\sqrt{\frac{d}{n}}\Big) < (2\varepsilon)^d$$

for every $\varepsilon > 0$.

Proof. Let $\mathbf{x}''$ denote the projection of a vector $\mathbf{x}$ onto the last $n-d$ coordinates. Let $A = \big\{\mathbf{x} \in S^{n-1} : |\mathbf{x}'| \le c\big\}$ and $\phi(\mathbf{x}) = \mathbf{x}' + \mathbf{x}''/|\mathbf{x}''|$. Then ϕ maps A onto $A' = (cB^d)\times S^{n-d-1}$ bijectively. We claim that

$$\mathrm{vol}_{n-1}(A) \le \mathrm{vol}_{n-1}(A') = c^d \pi_d \pi_{n-d}(n-d). \tag{19}$$

Indeed, it is not hard to compute that the Jacobian of ϕ^{-1} (as a map $A' \to A$) is $(1-|\mathbf{x}'|^2)^{(n-d-2)/2}$, which is at most 1. □

We also need bounds on the length of 1-dimensional projections of convex sets.

Lemma 4.3 (a) *Let $K \subset \mathbb{R}^n$ be a convex body, let $\mathbf{v} \in S^{n-1}$ be chosen uniformly and randomly, and let $I_\mathbf{v}$ be the orthogonal projection of K onto the line $e_\mathbf{v}$ containing $\mathbf{v}$. Then*

$$\mathsf{P}(\lambda(I_\mathbf{v}) \le s) \le \frac{\pi_n s^n}{2^n \mathrm{vol}(K)}.$$

(b) *Let $K \subset \mathbb{R}^d$ be a convex body, let $\mathbf{v} \in S^{n-1}$ be chosen uniformly and randomly, and let $I_\mathbf{v}$ be the orthogonal projection of K onto the line $e_\mathbf{v}$ containing $\mathbf{v}$ (we consider $\mathbb{R}^d$ as a subspace of $\mathbb{R}^n$). Then*

$$\mathsf{P}\big(\lambda(I_\mathbf{v}) \le s\big) \le \Big(es\sqrt{\frac{n}{d}}\Big)^d \frac{\pi_d}{\mathrm{vol}_d(K)}\big(1+\ln\ \mathrm{diam}(K)\big).$$

For the proof of Lemma 4.3, we need a simple probabilistic inequality.

Lemma 4.4 *Let X and Y be random variables, such that $a \le X \le a'$ and $b \le Y \le b'$ (where $-\infty$ and ∞ are permitted values). Let $s \le a'+b'$, and define $\alpha = \max(a, s-b')$ and $\beta = \min(a', s-b)$). Then*

$$\mathsf{P}(X+Y \le s) \le \int\limits_{\alpha}^{\beta+1} \mathsf{P}\big(X \le t,\ Y \le s+1-t\big)\, dt. \tag{20}$$

Proof. Starting with the right side,

$$\begin{aligned}\int\limits_{\alpha}^{\beta+1} \mathsf{P}\big(X \le t,\ Y \le s+1-t\big)\, dt &= \int\limits_{\alpha}^{\beta+1} \mathsf{E}_{X,Y}\, \mathbb{1}(X \le t \le s+1-Y)\, dt \\ &= \mathsf{E}_{X,Y} \int\limits_{\alpha}^{\beta+1} \mathbb{1}(X \le t \le s+1-Y)\, dt = \mathsf{E}_{X,Y}\big(\min(s+1-Y, \beta+1) - \max(X,\alpha)\big) \\ &\ge \mathsf{E}_{X,Y}\, \mathbb{1}(X+Y \le s) = \mathsf{P}(X+Y \le s)\end{aligned}$$

(one needs to check that if $X+Y \le s$, then $\min(s+1-Y, \beta+1) - \max(X,\alpha) \ge 1$). $\square$

Proof of Lemma 4.3. (a) It is easy to check that

$$\lambda(I_{\mathbf{v}}) = \frac{2}{\lambda((K-K)^* \cap e_{\mathbf{v}})}, \tag{21}$$

Hence by rather simple integration,

$$\mathrm{vol}((K-K)^*) = \pi_n \mathsf{E}\big(\lambda(I_{\mathbf{v}})^{-n}\big). \tag{22}$$

By the Blaschke–Santaló Inequality,

$$\mathrm{vol}(K-K)\mathrm{vol}((K-K)^*) \le \pi_n^2, \tag{23}$$

and by the Brunn-Minkowski Inequality

$$\mathrm{vol}(K-K) \ge 2^n \mathrm{vol}(K). \tag{24}$$

Hence

$$\mathsf{E}\big(\lambda(I_{\mathbf{v}})^{-n}\big) = \frac{\mathrm{vol}((K-K)^*)}{\pi_n} \le \frac{\pi_n}{\mathrm{vol}(K-K)} \le \frac{\pi_n}{2^n \mathrm{vol}(K)}. \tag{25}$$

Markov's Inequality implies part (a) of the lemma.

(b) Let $\mathbf{v} \in S^{n-1}$ be chosen randomly from the uniform distribution. Let $\mathbf{u}$ be the projection of $\mathbf{v}$ onto $\mathbb{R}^d$. Then $\lambda(I_{\mathbf{v}}) = |\mathbf{u}|\lambda(I_{\mathbf{u}})$. We apply Lemma 4.4 to the random variables $X = \ln|\mathbf{u}|$ and $Y = \ln\lambda(I_{\mathbf{u}})$ (with probability 1, both $|\mathbf{u}|$ and $\lambda(I_{\mathbf{u}})$ are positive). For the bounds of integration we use the bounds $|\mathbf{u}| \le 1$ and $\lambda(I_{\mathbf{u}}) \le \mathrm{diam}(K)$. We get

$$\mathsf{P}\Big(\lambda(I_{\mathbf{v}}) \le s\Big) \le \int\limits_{s/\mathrm{diam}(K)}^{es} \frac{1}{t}\mathsf{P}\Big(|\mathbf{u}| \le t,\ \lambda(I_{\mathbf{u}}) \le \frac{es}{t}\Big)\,dt. \tag{26}$$

(The factor $1/t$ comes in because of the change of the variable.) Note that the length $|\mathbf{u}|$ and the direction of $\mathbf{u}$ are independent as random variables, and hence so are $|\mathbf{u}|$ and $\lambda(I_{\mathbf{u}})$. By Lemma 4.2 and part (a),

$$\mathsf{P}\big(|\mathbf{u}| \le t\big)\mathsf{P}\Big(\lambda(I_{\mathbf{u}}) \le \frac{es}{t}\Big) \le \Big(2t\sqrt{\frac{n}{d}}\Big)^d \frac{\pi_d}{2^d \mathrm{vol}_d(K)}\Big(\frac{es}{t}\Big)^d = \Big(es\sqrt{\frac{n}{d}}\Big)^d \frac{\pi_d}{\mathrm{vol}_d(K)}.$$

Substituting this bound in (26), we get

$$\begin{aligned}\mathsf{P}\Big(\lambda(I_{\mathbf{v}}) \le s\Big) &\le \Big(es\sqrt{\frac{n}{d}}\Big)^d \frac{\pi_d}{\mathrm{vol}_d(K)} \int\limits_{s/\mathrm{diam}(K)}^{es} \frac{1}{t}\,dt \\ &= \Big(es\sqrt{\frac{n}{d}}\Big)^d \frac{\pi_d}{\mathrm{vol}_d(K)}\big(1+\ln\ \mathrm{diam}(K)\big). \end{aligned} \tag{27}$$

□

Exercise 4.5 The graph $K_{2,3}$ does not embed isometrically into ℓ_1.

Exercise 4.6 (a) Prove that every finite ℓ_2-space is isometric with an ℓ_1-space. (b) Show by an example that the converse is not valid.

Exercise 4.7 Prove that every finite metric space (V, d) has a vector-labeling with volume distortion at most 2 for V.

Exercise 4.8 Show that there exists a graph G on n nodes for which $\mathsf{bw}(G)/\mathsf{d}_{\mathrm{loc}}(G) \ge \log n$.

References

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

[2] S.D.A. Dasgupta and A.K. Gupta: An Elementary Proof of a Theorem of Johnson and Lindenstrauss, *Random Structures and Algorithms* **22** (2002), 60–65.

[3] C. Dubey, U. Feige, W. Unger: Hardness results for approximating the bandwidth, *Journal of Computer and System Sciences* **77** (2010), 62–90.

[4] U. Feige: Approximating the bandwidth via volume respecting embeddings, Proc. 30th Ann. ACM Symp. on Theory of Computing, ACM New York, NY (1998), 90-99.

[5] P. Frankl, H. Maehara: The Johnson-Lindenstrauss lemma and the sphericity of some graphs, *J. Combin. Theory Ser. B* **44** (1988), 355–362.

[6] P. Indyk and J. Matoušek: Low-distortion embeddings of finite metric spaces, in: *Handbook of Discrete and Computational Geometry* (ed. J.E. Goodman, J. O'Rourke), Second Edition, Chapman and Hall (2004), 177196.

[7] M. Iri: On an extension of the maximum-flow minimum cut theorem to multicommodity flows, *J. Oper. Res. Soc. Japan* **5** (1967), 697–703.

[8] W.B. Johnson, J. Lindenstrauss: Extensions of Lipschitz mappings into a Hilbert space. *Conference in modern analysis and probability* (New Haven, Conn., 1982), 189–206, Contemp. Math., 26, Amer. Math. Soc., Providence, R.I., 1984.

[9] F.T. Leighton and S. Rao, An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms, *Proc. 29th Annual Symp. on Found. of Computer Science,* IEEE Computer Soc., (1988), 422-431.

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

[11] L. Lovász: Combinatorial Problems and Exercises, Second Edition, Akadémiai Kiadó - North Holland, Budapest (1991).

[12] P. Massart: *Concentration Inequalities and Model Selection*, Ecole dEté de Probabilités de Saint-Flour XXXIII, Springer 2003.

[13] J. Matoušek: On embedding expanders into l^p spaces. *Israel J. Math.* **102** (1997), 189–197.

[14] J. Matoušek: *Lectures on discrete geometry*, Graduate Texts in Mathematics bf 212, Springer–Verlag, New York (2002), Chapter 15.

[15] F. Shahroki and D.W. Matula: The maximum concurrent flow problem, J. ACM **37** (1990), 318–334.