Local versus Global Properties of Metric Spaces
(Extended Abstract)

Sanjeev Arora* László Lovász† Ilan Newman‡ Yuval Rabani§
Yuri Rabinovich¶ Santosh Vempala‖

Abstract

Motivated by applications in combinatorial optimization, we initiate a study of the extent to which the global properties of a metric space (especially, embeddability in ℓ_1 with low distortion) are determined by properties of "small" sets. We prove both positive and negative results.

1 Introduction

The study of metric spaces has come to occupy a central place in theoretical computer science. Finite ℓ_1 metrics have attracted special attention, since in computational settings they seem more complex and interesting than say ℓ_2. One reason is that ℓ_1 metrics correspond exactly to metrics that lie in the *cut cone*, and cuts are important graph-theoretic objects (see the book [10]). Approximation algorithms for NP-hard cut problems such as SPARSEST CUT are derived by embedding general metric spaces into ℓ_1 [17, 5] (and more recently, negative type metrics into ℓ_1 [4, 8, 3]). Furthermore, ℓ_1 metrics also seem to possess a richer structure than ℓ_2. For instance, deciding whether a finite metric space embeds isometrically into ℓ_2 is decidable in polynomial time whereas it is NP-hard for ℓ_1. Similarly, dimension of ℓ_2 spaces can be reduced to $O(\log n/\epsilon^2)$ while distorting distances by at most $1+\epsilon$, whereas such dimension reduction is impossible in ℓ_1 without significant distortion [7].

When a class of metrics exhibits such complex behavior, one may try to understand how does this complex behavior arise. Specifically, this paper studies questions of the following type: To what extent can the properties of the finite metric space be inferred from looking at the induced metric on *all* "small" subsets? To give one concrete example, can the embeddability of a finite metric space into ℓ_1 be inferred by checking the embeddability of all "small" subsets? (We list many other such questions later.)

A main reason to consider such questions is design of algorithms for cut problems. SDP relaxations for these problems involve finding points in $\mathbb{R}^n$ such that square of the interpoint distances satisfy constraints such as triangle inequality, or more generally, k-gonal inequalities. The latter correspond to requiring that locally, the submetrics "look" like ℓ_1 metrics on subsets of size k. One could conceive of more complicated SDPs that impose constraints on larger and larger subsets. Could this lead to progressively tighter relaxations? A more ambitious way to tighten the relaxation

*Supported by David and Lucile Packard Fellowship and NSF grant CCR 0205594. arora@cs.princeton.edu
†Microsoft Research and Eötvös Loránd University. lovasz@microsoft.com
‡Computer Science Department, University of Haifa, Haifa 31905, Israel. Email: ilan@cs.haifa.ac.il.
§Computer Science Department, Technion, Haifa 32000, Israel. Email: rabani@cs.technion.ac.il.
¶Computer Science Department, University of Haifa, Haifa 31905, Israel. Email: yuri@cs.haifa.ac.il.
‖Mathematics Department, MIT. vempala@math.mit.edu

would be to do k rounds of a lift-and-project method such as Lovasz-Schrijver or Sherali-Adams. These yield metrics in which the induced metric on *all* subsets of size k is isometrically embeddable in ℓ_1. Relaxations obtained by k rounds of lift-and-project are solvable in $n^{O(k)}$ time, so if the relaxation becomes very tight for say $k = \log n$ then one would obtain a $n^{O(\log n)}$ time approximation algorithm. (Such observations were made in a recent paper [2], where it was observed that studying such questions leads to study of "local versus global" structure. That paper restricted attention to vertex cover, however.) Note that to prove the tightness of such a relaxation, it would suffice to show that if subsets of size k embed isometrically in ℓ_1, then the entire metric embeds in ℓ_1 with low distortion. One of the results in this paper is that this is false for general metrics for $k = \omega(1)$ (see Section 3.2 for exact results, including related results about negative type metrics, a generalization of ℓ_1), while it is indeed true when $k = \Theta(n)$ (see Section 3.1). Presumably the second statement can not be improved significantly. We give some evidence to support this, by showing that the property "embeddable in ℓ_1 with $O(1)$ distortion" cannot be locally inferred. We describe an n-point metric which requires distortion $\Omega(\log n)$ into ℓ_1 but whose every subset of size $n^{1-\epsilon}$ embeds into ℓ_1 (or even weaker subclasses of ℓ_1) with distortion $O(1/\epsilon^2)$ (Section 3.2). We note that such constructions are nontrivial precisely because we need a fairly strong property to hold for *every* subset.

General comments. "Local versus global" is an old theme in mathematics. The subject initiated here is sometimes similar in spirit to past work, and sometimes not. The field of Helly-type theorems (e.g., in a family of bounded convex sets in $\mathbb{R}^n$, if every subset of $n+1$ has a nonempty intersection, then so do all) shows that sometimes local properties do determine global properties. Our positive results fit in this tradition: assuming local order exists in *all* subsets of a certain size, we try to infer some global order. Ramsey theory is another field that seems related: its main lesson is that in midst of global "disorder," there is always a significant subset exhibiting "order." (Ramsey phenomena exist even for metric spaces; see [6].) In this phrasing, our positive results (phrased contrapositively) can be seen as assuming global "disorder," and giving a lowerbound on the size of the smallest "disordered" subset. (For ℓ_1 embeddability this lowerbound matches some results in [6].) Our negative results on the other hand give examples where *every* small subset has "order" and nevertheless globally there is "disorder."

We would like to point out that in areas such as PCPs, program checking, property testing, etc, the "local versus global" questions also play an important role, but in those settings the local property only has to hold for *many* local neighborhoods, not all.

In context of metric spaces, a classical result of local versus global nature is Menger's theorem [19], which shows that if every subset of size $n+3$ of a metric space embed isometrically into ℓ_2^n, then the entire space embeds isometrically into ℓ_2^n. For ℓ_1^n it is not even known if $n+3$ can be replaced by any finite size. Another well known result of this kind is that a metric is a tree-metric iff every submetric of size 4 is a tree-metric.

A significant aspect of our work is a new insight into shortest-path metrics derived from random graphs of bounded degree, which are used in most of our negative results. These metrics were shown to be extremal for many metric-theoretic properties in the past. Surprisingly, their local structure turns out to be rather simple, even when the size of the submetrics is as large as $m = n^{1-\epsilon}$. Based on this, we conclude that there exist metrics whose distortion in ℓ_1 is the largest possible one for n, while all their submetrics of size $m = n^{1-\epsilon}$ embed into ℓ_1 with a constant distortion.

It should be stressed that the most important open problem arising from our work is that we failed to construct metrics that require large distortion in ℓ_1, but where every subset of size at most say n^ϵ (or even $\Omega(\log n)$) embeds *isometrically* (as opposed to embedding with low distortion) into ℓ_1. Thus the possibility remains of improving the recent $\sqrt{\log n}$-approximations for many cut problems via the lift-and-project approach outlined above.

To emphasize the distinction between isometric and *almost* isometric embeddings, we show that for the class of ultrametrics the former assumption is much more powerful than the latter. Ultrametrics are metrics satisfying $\forall x, y, z\ d(x,z) = \max\{d(x,y), d(z,y)\}$; thus if every subset of size three is an ultrametric, then so is the whole metric. On the contrary, we show for every c and ϵ that if every subset of cardinality n^ϵ embeds into an ultrametric with distortion bounded by c, then the whole metric on n points may still require distortion $c^{1/\epsilon}$ for embedding into an ultrametric. (On the optimistic side, we show that this lower bound is tight by a establishing a matching upper bound; see Section 4.)

Our final contribution has to do with a notion of a *normal* class of metrics, obtained by postulating some properties shared by many metric classes that are extensively used in approximation algorithms. First, we establish a general positive local vs. global result for any normal class of metrics. Second, we separate any nontrivial normal class of metrics from general metrics. This opens a door to a potential use of various normal classes of metrics in approximation algorithms.

2 Preliminaries

We use $\mathrm{dist}(d, d')$ to denote the *distortion* between two distance functions d and d' on the same set of points. For a class $\mathcal{C}$ of distance functions, we use $\mathrm{dist}(d \hookrightarrow \mathcal{C})$ to denote the minimum distortion between d and $d' \in \mathcal{C}$. (This assumes, of course, that $\mathcal{C}$ contains distance functions on the same set of points as d.)

Let d be a distance function (on an underlying point set P), and let $f : \mathbb{R} \to \mathbb{R}$ be a monotonically non-decreasing function with $f(0) = 0$. We denote by $f(d)$ the distance function where $\forall p, q \in P,\ f(d)(p,q) = f(d(p,q))$. Notice that if d is a metric and f is concave, then $f(d)$ is a metric. The *power scale* $f(x) = x^c$, $c \in [0,1]$ plays an important role in this paper. It is worth noting the following simple fact:

$$\mathrm{dist}(d^c, (d')^c) = \left(\mathrm{dist}(d, d')\right)^c.$$

Let d be a distance function, and let Q be a subset of the points on which d is defined. We use d^Q to denote the restriction of d to the pairs of points in Q.

A set of metrics $\mathcal{C}$ is called *normal* if it has the following properties, shared e.g., by the classes NEG, HYP, and $\mathcal{M}_k$ to be discussed later (see Section 3.2):

1. It is symmetric, i.e., for every $d \in \mathcal{C}$, any metric d' derived from d by permuting the underlying set of points is also in $\mathcal{C}$.

2. It is a closed cone, i.e., for every $d, d' \in \mathcal{C}$ on the same set of points, for every $a, a' \geq 0$, also $a \cdot d + a' \cdot d' \in \mathcal{C}$.

3. It is hereditary, i.e., for every $d \in \mathcal{C}$, for every subset of points Q on which d is defined, also $d^Q \in \mathcal{C}$.

4. For every $d \in \mathcal{C}$, consider a metric d', obtained from d by performing the following *cloning* operation: Pick a point p, add a "clone" q, and set $d'(q,x) = d(p,x)$ for all points x. Then, $d' \in \mathcal{C}$.

Observe that every normal set of metrics includes all cut metrics, and therefore all metrics that embed isometrically in ℓ_1. Further notice that if $\mathcal{C}$ is a normal set of metrics, then for every $\gamma \geq 1$, the set of metrics

$$\mathcal{C}_\gamma = \{d : \ \mathrm{dist}(d \hookrightarrow \mathcal{C}) \leq \gamma\}$$

is also normal.

3 Normal Sets of Metrics

3.1 Upper Bounds

This section is devoted to the proof of the following theorem.

Theorem 3.1. *Let $m, n \in \mathbb{N}$, $m \leq n$, let $c \geq 1$, and let $\mathcal{C}$ be a normal set of metrics. Let d be a metric on n points such that every m-point subspace Q has* $\mathrm{dist}(d^Q \hookrightarrow \mathcal{C}) \leq \gamma$. *Then,*

$$\mathrm{dist}(d \hookrightarrow \mathcal{C}) = O\left(\gamma \cdot \left(\frac{n}{m}\right)^2\right).$$

We require a definition. Let d be a distance function. A *tree-like extension* of d is a distance function d' which is obtained from d by repeatedly performing the following *attachment* operation: Pick a point p and a weight $w \geq 0$, "attach" to p a new point q, and set $d'(q, x) = d'(p, x) + w$ for all points x.

Lemma 3.2. *Let $\mathcal{C}$ be a normal set of metrics, let $d \in \mathcal{C}$, and let d' be a tree-like extension of d. Then $d' \in \mathcal{C}$.*

Proof. Clearly, it suffices to prove this for a single attachment operation. Let d_p be the metric obtained from d by adding a clone q of a point p. Let δ be the cut metric defined by $\delta(x, y) = 1$ if exactly one of the points x, y is q, and $\delta(x, y) = 0$ otherwise. Both d_p and δ are in $\mathcal{C}$ (the former by definition, the latter because $\mathcal{C}$ must contain all cut metrics). Attaching q to p at distance w gives the metric $d' = d_p + w \cdot \delta$. As $\mathcal{C}$ is a closed cone, $d' \in \mathcal{C}$. □

Next, we introduce the construction that will be used in the proof of Theorem 3.1. Let d be a metric on the finite set of points $P = \{p_1, p_2, \ldots, p_n\}$. Let $m \in \{1, \ldots, n\}$, and let $\sigma \in S_n$ be a permutation on $\{1, 2, \ldots, n\}$. We define the metric d_m^σ as follows. Let $P_m^\sigma = \{p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\}$. The metric d_m^σ is a tree-like extension of $d^{P_m^\sigma}$, generated by attaching, for $i = m+1, \ldots, n$, the point $p_{\sigma(i)}$ to p_{j_i} at distance w_i, where $j_i \in \{1, 2, \ldots, i-1\}$ minimizes $d(p_{\sigma(i)}, p_{\sigma(j)})$, and $w_i = d(p_{\sigma(i)}, p_{\sigma(j_i)})$. We now average over σ. For every $p, q \in P$ put

$$d_m^*(p, q) = \frac{1}{n!} \cdot \sum_{\sigma \in S_n} d_m^\sigma(p, q).$$

The proof of Theorem 3.1 now follows.
Proof of Theorem 3.1. Let $T_{n,m}$ denote the supremum over all n-point metrics d of $\mathrm{dist}(d, d_m^*)$. Clearly, $T_{m,m} = 1$. Notice that for every $p, q \in P$, $d_m^\sigma(p, q) \geq d(p, q)$. On the other hand,

$$d_m^*(p, q) = \mathrm{E}_\sigma[d_m^\sigma(p, q)] = \frac{2}{n} \cdot \mathrm{E}_\sigma\left[d_m^\sigma(p, q) : \ \sigma(n) \in \{p, q\}\right] + \left(1 - \frac{2}{n}\right) \cdot \mathrm{E}_\sigma\left[d_m^\sigma(p, q) : \ \sigma(n) \notin \{p, q\}\right].$$

Notice that

$$\mathrm{E}_\sigma\left[d_m^\sigma(p, q) : \ \sigma(n) \notin \{p, q\}\right] = T_{n-1,m} \cdot d(p, q).$$

Consider the case that $\sigma(n) = p$. As $d(p, t_n) \leq d(p, q)$, so $d(t_n, q) \leq 2d(p, q)$, we have that

$$\mathrm{E}_\sigma[d_m^\sigma(p, q) : \ \sigma(n) = p] = d(p, t_n) + \mathrm{E}_\sigma[d_m^*(t_n, q) : \ \sigma(n) = p] \leq d(p, q) + T_{n-1,m} \cdot 2d(p, q).$$

The case that $\sigma(n) = q$ is analogous. Therefore,

$$T_{n,m} \leq \left(1 + \frac{2}{n}\right) \cdot T_{n-1,m} + \frac{2}{n}.$$

Solving the recurrence, we get that

$$T_{n,m} = O\left(\left(\frac{n}{m}\right)^2\right).$$

Notice that $\mathcal{C}_\gamma$ is a normal set of metrics. By the conditions of the theorem, for every m-point subset Q, $d^Q \in \mathcal{C}_\gamma$. Therefore, by Lemma 3.2, for every permutation σ, $d_m^\sigma \in \mathcal{C}_\gamma$. As $\mathcal{C}_\gamma$ is a closed cone, also $d_m* \in \mathcal{C}_\gamma$. As $\mathrm{dist}(d, d_m^*) \leq \left(\frac{n}{m}\right)^2$, the theorem follows. ☐

3.2 Lower Bounds

The main result in this section is the following theorem which implies a nearly tight counterpart to some of the upper bounds from the previous section.

Theorem 3.3. *Let d be the shortest path metric of a random n-node 3-regular graph $G = (V, E)$, and let $0 \leq \varepsilon \leq 1$. Then, with high probability, for every $S \subseteq V$ with $|S| \leq n^{1-\varepsilon}$, the corresponding d^S can be embedded in ℓ_1 (in fact, even into a distribution over dominating tree-metrics) with distortion $O(1/\varepsilon^2)$.*

Corollary 3.4. *For every $\varepsilon > 0$ and for every integer $n \geq 2$, the following statements hold.*

1. *There is an n-point metric d such that for every $n^{1-\varepsilon}$-point subspace Q, $\mathrm{dist}(d^Q \hookrightarrow \ell_1) = O(1/\varepsilon^2)$ yet $\mathrm{dist}(d \hookrightarrow \ell_1) = \Omega(\log n)$.* [1]

2. *There is an n-point metric d such that for every $n^{1-\varepsilon}$-point subspace Q, $\mathrm{dist}(d^Q \hookrightarrow \ell_2) = O(1/\varepsilon)$, yet $\mathrm{dist}(d \hookrightarrow \ell_2) = \Omega(\sqrt{\log n})$.*

Proof. It is well-known that the distortion of embedding the shortest path metric of an n-node bounded degree expander into ℓ_1 is $\Omega(\log n)$, so the metric d from Theorem 3.3 satisfies the first statement. To see the second statement, use the metric $\sqrt{d}$, keeping in mind that the square root of an ℓ_1 metric is an ℓ_2 metric. ☐

To prove the theorem we will use three properties of random graphs. The first two are standard and well known: (1.) The girth is $\Omega(\log n)$. (N.B. More correctly, the high girth property holds only after deleting $o(n)$ edges, but this will not affect the other properties, and in what follows we shall gloss over this point.) (2.) The diameter is $O(\log n)$. The 3rd property is about sparsity of subgraphs of a random graph, and it is not hard as well (see [2]).

Lemma 3.5. **[2]** *With high probability, the subgraph induced by any subset S of size at most $n^{1-\varepsilon}$ has at most $(1 + \frac{c}{\varepsilon \log n})(|S| - 1)$ edges where c is an absolute constant.*

Before starting with the proof of Theorem 3.3, let us establish some preparatory lemmas.

Let $H = (U, F)$ be a subgraph of G of with $O(n^{1-\varepsilon'})$ vertices. We define two polytopes in $\mathbb{R}^{|F|}$. The first polytope, P, will be the *spanning tree polytope* of H, i.e., the set of all vectors that are

[1] In fact, the lower bound holds even for embedding into NEG, the class of negative type metrics (for definition — see below).

convex combinations of incidence vectors of spanning trees of H. The second polytope, B, will be the following axis-parallel box with one corner being the vector of all 1's.

$$B_\alpha = \{v \in \mathbb{R}^{|F|} | \forall e : \alpha \leq v_e \leq 1\}. \tag{1}$$

Keeping in mind Lemma 3.5, define $p = \Theta(\epsilon' \log n)$ so that for every $X \subseteq U$, the subgraph of H induced by X has at most $|X|(1 + 1/p)$ edges. In addition, assume w.l.o.g. that the girth of G exceeds p.

The following Lemma is key:

Lemma 3.6. *Let $\alpha \leq p/(p+1)$. Then, $P \cap B_\alpha \neq \emptyset$.*

Proof. By Farkas' Lemma, it suffices to show that for any $w \in \mathbb{R}^{|F|}$, there exists a vector $v \in B$ such that

$$\max_{x \in P} w \cdot x \geq w \cdot v. \tag{2}$$

Note that since the extreme points of P are spanning trees of H, the LHS is always maximized by the incidence vector of some spanning tree. We consider two extreme cases:

1. $w \leq 0$: In this case we set $v_{ij} = 1$ for every edge (i, j). The inequality follows.

2. $w \geq 0$: In this case we set $v_{ij} = \alpha$ for every edge. Suppose the LHS of (2) is maximized by the spanning tree T. We will prove that the total weight of all the edges in H is only slightly larger than the weight of T. For this consider the following bipartite graph. The left side of the bipartition has a point corresponding to each edge of T. The right side has a point for each edge of H that is not in T. There is an edge (e, f) if $e \in T$ belongs to the fundamental cycle of $f \notin T$. Note that the optimality of T implies that $w_e \geq w_f$. Let the girth of G be g. Recall $g \geq p$. Thus the degree of each vertex in $\overline{T}$ is at least p. We claim that this bipartite graph has a a p-matching: a subgraph with degree 1 for points on the left and degree p for points on the right. Suppose not. Then there is some minimal subset X on the right side whose neighborhood $N(X)$ has size $|N(X)| < |X|p$. Now consider the subtree of T induced by $N(X)$ (if the edges corresponding to $N(X)$ do not form a connected component, then X is not minimal). This subtree has $|N(X)| + 1$ vertices and the subgraph of H induced by these vertices has at least $|N(X)| + |X| > |N(X)|(1 + 1/p)$ edges. But this contradicts Lemma 3.5.

   The existence of the p-matching implies that the edges of T can be partitioned into p subsets such that the weight of each subset is more than the weight of all the edges not in T. Thus,

   $$\sum_{ij} w_{ij} \leq \left(1 + \frac{1}{p}\right) \sum_{ij \in T} w_{ij}.$$

   This implies that inequality (2) holds for any $\alpha \leq p/p + 1$.

3. For the general case, take an arbitrary vector w and set $c_{ij} = 1$ for $w_{ij} \leq 0$ and $c_{ij} = \alpha$ if $w_{ij} > 0$. Consider the connected components induced by the nonnegative edges. For each component the inequality is implied separately by the second case above. Now shrink all the components to single vertices. The inequality on the induced graph follows from the first case. Summing up, (2) is proved.

□

The second preparatory result is about truncated tree-metrics. It is well-known that metrics induced by trees are isometrically embeddable in ℓ_1. We show that *truncated* tree metrics are embeddable with a constant distortion.

Claim 3.7. *Given a tree metric t and a number $M \geq 0$, let $t'_{ij} = \min\{t_{ij}, M\}$. Then t' can be embedded into ℓ_1* [2] *with constant distortion.*

We give two alternative proofs. The first embeds into ℓ_1 with distortion at most 2 and the second into a distribution of dominating tree metrics with distortion $O(1)$.

First proof: We describe a probabilistic embedding ϕ from the vertices of the tree into ℓ_1 with expected distortion at most 2. This will prove the result since a convex combination of ℓ_1 metrics is itself an ℓ_1 metric.

The embedding will have D coordinates. Let $u^1, u^2, \ldots, u^D$ be the D axis vectors. Label each edge of the underlying tree with one of the D axis vectors, chosen at random.

Next, fix a root vertex and embed it in the origin. Every other vertex i is mapped to the point $\sum_k b_k u^k$ where $b_k \in \{0,1\}$ is 1 iff u^k occurs an odd number of times as a label of the edges on the path from the root to i.

Clearly, $|\phi(i) - \phi(j)| \leq d'_{ij}$ always.

For the lower bound, we consider the expected distance between two vertices i and j. We are choosing d_{ij} vectors independently from D possibilities at random (with replacement). The expected value of $|\phi(i) - \phi(j)|$ is the expected the number of vectors that are chosen an odd number of times. This is at least $\frac{1}{2}\min\{d_{ij}, D\}$. □

Second proof: Let T be the tree corresponding to t. Build a (weighted) graph T' by introducing a new vertex u, and connecting it to every vertex of T by an edge of length $M/2$. Observe that the shortest-path metric of T' restricted to $V(T)$ is precisely t, and that T' happens to be 2-outerplanar. By [9], this implies that t', and hence t, can be embedded into a distribution of dominating tree metrics (and thus into ℓ_1) with constant distortion. □

We are now ready to prove Theorem 3.3.

Proof of Theorem 3.3: Fix a subset S with at most $n^{1-\varepsilon}$ vertices, and consider the corresponding induced metric d^S. Recall the definition of a *spanner* of (S, d^S): it is a graph $Y = (S, L)$ on the vertex set S, such that the weight of an edge $(i,j) \in L$ is $d^S(i,j)$. By [1], there exists a *spanner* Y of (S, d^S) such that $|L(Y)| \leq n^{1-\varepsilon} \cdot n^{\varepsilon/2} = n^{1-\varepsilon/2}$, and the shortest-path metric of Y approximates d^S up to a factor $O(1/\varepsilon)$.

Let $H = (U, F)$ be the subgraph obtained by including all the edges on the shortest paths between pairs of vertices $i, j \in S$, such that $(i,j) \in L(Y)$. Since the diameter of G is $O(\log n)$, H has $O(n^{1-\varepsilon/2}\log n)$ vertices.

We will show that the shortest-path metric d^H induced by H can be embedded into a distribution of dominating tree metrics with distortion $O(1/\varepsilon)$. Hence, d^S embeds into such a distribution with distortion $1/\varepsilon \cdot 1/\varepsilon$, and the theorem follows.

Indeed, applying Lemma 3.6 to H, we conclude that there is a probability distribution on spanning trees $\{T_i\}$ of H such that each edge of H occurs with probability at least $\alpha = p/(p+1)$, where $p = \Theta(\varepsilon \log n)$. Let $D = O(\log n)$ be the diameter of H.

For each T_i in the distribution, consider the corresponding metric $t_i = \min\{D, d_{T_i}\}$. Define a metric $t = \sum w_i t_i$, where w_i is the weight of T_i in the distribution. Clearly, t dominates d_H. To upper-bound $\mathrm{dist}(t, d^H)$, consider an edge of H. It's $t-$length is at most

$$1 \cdot \alpha + D \cdot (1-\alpha) \;\leq\; \frac{1+D}{1+p} \;=\; \frac{1+O(\log n)}{\Theta(\varepsilon \log n)} \;=\; O(1/\varepsilon)\,.$$

[2]In fact, ℓ_1 can be replaced by a distribution of dominating tree-metrics, a more restricted class of metrics.

Thus, $\mathrm{dist}(t, d^H) = O(1/\varepsilon)$.

To conclude the proof, recall that by Claim 3.7, every t_i, and hence t can be embedded in a distribution of H-dominating tree metrics with a constant distortion. $\square$

We now turn our attention to the case where subspaces embed isometrically into an interesting class of metrics. Our lower bounds in this case are much weaker. In order to state our results, we need a few definitions.

A distance function d is *k-gonal* iff for every two sequences of points $p_1, p_2, \ldots, p_{\lfloor k/2 \rfloor}$ and $q_1, q_2, \ldots, q_{\lceil k/2 \rceil}$ (where points are allowed to appear multiple times in each sequence) the following inequality holds:

$$\sum_{i=1}^{\lfloor k/2 \rfloor} \sum_{j=1}^{\lceil k/2 \rceil} d(p_i, q_j) \geq \sum_{i=1}^{\lfloor k/2 \rfloor} \sum_{i'=1}^{\lfloor k/2 \rfloor} d(p_i, p_{i'}) + \sum_{j=1}^{\lceil k/2 \rceil} \sum_{j'=1}^{\lceil k/2 \rceil} d(q_j, q_{j'}).$$

We use $\mathcal{M}_k$ denote the class of all k-gonal distance functions. Clearly, $\mathcal{M}_3$ is simply all metrics. Also, for every $k \in \mathbb{N}$, $k \geq 2$, $\mathcal{M}_{k+2} \subset \mathcal{M}_k$ and $\mathcal{M}^n_{2k-1} \subset \mathcal{M}^n_{2k}$. On the other hand, for every $k \in \mathbb{N}$, $k \geq 1$, distance functions in $\mathcal{M}^n_{2k}$ are not necessarily metrics. The class of all *negative type* distance functions is

$$\text{NEG} = \bigcap_{k=2}^{\infty} \mathcal{M}_{2k}.$$

Schoenberg showed that $d \in \text{NEG}$ iff $\sqrt{d}$ embeds isometrically into ℓ_2. The class of all hypermetrics is

$$\text{HYP} = \bigcap_{k=2}^{\infty} \mathcal{M}_{2k-1}.$$

Thus, all hypermetrics are negative type metrics. It is known that all ℓ_1 metrics are hypermetrics. All classes of metrics discussed above (except for ℓ_2 metrics) are normal.

A theorem in [11], combined with an argument similar to the proof of Corollary 3.4 gives the following theorem. The proof is omitted from this extended abstract.

Theorem 3.8. *For every integer $n \geq 2$ and for every $k \in \mathbb{N}$, $k \leq n$, the following statements are true:*

1. *There exists an n-point k-gonal metric d such that* $\mathrm{dist}(d \hookrightarrow \text{NEG}) = \Omega\left((\log n)^{\log_2(1+1/(\lceil k/2 \rceil - 1))}\right)$.
2. *There exists an n-point metric d such that every k-point subspace is hypermetric, yet* $\mathrm{dist}(d \hookrightarrow \text{NEG}) = \Omega\left((\log n)^{\log_2(1+1/(k-1))}\right)$.
3. *There exists an n-point metric d such that every k-point subspace embeds isometrically in ℓ_2, yet* $\mathrm{dist}(d \hookrightarrow \text{NEG}) = \Omega\left((\log n)^{\frac{1}{2}\log_2(1+1/(k-1))}\right)$.

4 Ultrametrics

The set of ultrametrics is the set of metrics

$$\text{ULT} = \{d : \ d(p,q) \leq \max\{d(p,r), d(q,r)\}, \ \forall p, q, r\}.$$

All ultrametrics embed isometrically into ℓ_2. Notice that ULT is not normal, so the results from the previous section do not apply to this set. We use the following basic fact about ultrametrics. Consider a metric d. Given two points x, y, an xy-path P is a sequence of points

$(x = p_0, p_1, p_2, \ldots, p_m = y)$ of arbitrary length. We say that $p, q \in P$ iff there exists $j \in \{1, 2, \ldots, m\}$ such that $p = p_{j-1}$ and $q = p_j$. For every two points x, y put

$$u(x,y) = \min_{xy\text{-paths } P} \{\max\{d(p,q) : \ pq \in P\}\}.$$

Theorem 4.1 (Farach-Colton [12]). *The distance function u is an ultrametric which is dominated by d (i.e., $u(x,y) \le d(x,y)$, for every $x, y \in X$). Moreover, every ultrametric u' that is dominated by d is also dominated by u.*

As an immediate corollary we get the following criterion.

Corollary 4.2. *Let $c \le 1$ be the maximum value such that for every $x, y \in X$, every xy-path P contains $pq \in P$ such that $d(p,q) \ge c \cdot d(x,y)$. Then,* $\mathrm{dist}(d \hookrightarrow \textsc{ult}) = c^{-1}$.

Using this criterion we establish the following theorem.

Theorem 4.3. *Let $c \ge 1$, and let $\epsilon > 0$. Let d be an n-point metric such that for every n^ϵ-point subspace Q,* $\mathrm{dist}(d^Q \hookrightarrow \textsc{ult}) \le c$. *Then,*

$$\mathrm{dist}(d \hookrightarrow \textsc{ult}) = c^{\lceil 1/\epsilon \rceil}.$$

This bound is essentially tight.

Proof. For the upper bound, it suffices to show that for $n = m^k - m^{k-1} + 1$, $k \in \mathbb{N}$, it holds that $\mathrm{dist}(d \hookrightarrow \textsc{ult}) \le c^k$. The proof is by induction on k. For $k = 1$ the theorem is trivially true. For $k > 1$, by Corollary 4.2, it suffices to show that for every $x, y \in X$, any simple xy-path P contains $pq \in P$ with $d(p,q) \ge d(x,y)/c^k$. Let $P = (x = v_1, v_2, \ldots, v_r = y)$, $r \le n$, be such a path. Consider the xy-path $P' = (v_1, v_m, v_{2m-1}, v_{3m-2}, \ldots, v_r)$. As P' has at most $1 + \frac{n-1}{m} = m^{k-1} - m^{k-2} + 1$ points, the induction hypothesis implies that there exists $v_{j-1}v_j \in P'$ with $d(v_{j-1}, v_j) \ge d(x,y)/c^{k-1}$. Consider the segment of P $(v_{j-1}, \ldots, v_j)$ containing at most m points. By the base case of the induction, there exists $pq \in (v_{j-1}, \ldots, v_j)$ such that $d(p,q) \ge d(v_{j-1}, v_j)/c \ge d(x,y)/c^k$.

For the lower bound, consider the metrics d_n^c, where d_n is the shortest path metric of the n-node cycle and $c \in [0,1]$. The reader can verify easily using Corollary 4.2 that $\mathrm{dist}(d_n^c \hookrightarrow \textsc{ult}) = \Omega(n^c)$, whereas for every subspace on n^ϵ points, the restriction d' of d_n^c to this subspace has $\mathrm{dist}(d' \hookrightarrow \textsc{ult}) = O(n^{\epsilon c})$. □

Remark 4.1. *The metrics d_n^c are, in fact, $\Omega(n^c)$ far from the more set of general tree metrics (by the argument from [21, Corollary 5.3]). Hence, the lower bounds hold for tree metrics as well.*

5 Separating a Normal Metric Class from ℓ_∞

Let $\mathcal{C}$ be a normal metric class such that there exists a metric $\mu \notin C$. How well can the metrics from $\mathcal{C}$ approximate general metrics? The following purely existential result of Matousek [18] implies a separation between the class of all metrics and $\mathcal{C}$, i.e., for every $\gamma > 1$ there exists a metric D such that $\mathrm{dist}(D \hookrightarrow \mathcal{C}) \ge \gamma$.

Theorem 5.1. *For every finite metric μ and any constants $\varepsilon > 0, \gamma > 1$, there exists a (larger) finite metric D such that, for any metric M on the same set of points as D, if* $\mathrm{dist}(D, M) \le \gamma$, *then M contains a submetric μ' with* $\mathrm{dist}(\mu, \mu') \le (1 + \varepsilon)$.

We conjecture that a much stronger separation holds.

Conjecture 1. *For any $n \in \mathbb{N}$, there exists an n-point metric d_n such that* $\mathrm{dist}(d_n \hookrightarrow C) \geq \Omega(\log^\alpha n)$ *for some constant $\alpha > 0$.*

In what follows, we produce a supporting evidence for this conjecture by proving its analogue for normed spaces. Unlike in the rest of the paper, we assume here that $\mathcal{C}$ contains not only finite metrics, but also metrics whose underlying space is the entire $\mathbb{R}^n$ or $\mathbb{Z}^n$, and, in particular $\{\ell_1^n\}_{n=1}^\infty \in \mathcal{C}$.

Theorem 5.2. *Let $\mathcal{C}$ be a normal metric class, and assume that there exists a metric μ_k on k points such that* $\mathrm{dist}(\mu_k \hookrightarrow \mathcal{C}) = \beta > 1$. *Then, for any $\mathcal{C}$-metric d on $\mathbb{R}^n$, it holds*

$$\mathrm{dist}(d, \ell_\infty^n) \;=\; \Omega(n^\alpha), \quad \textit{where} \;\; \alpha \;\approx\; \frac{1}{2}\frac{\beta-1}{\beta+1}\frac{1}{\ln k} \;.$$

Observe that the gap between the two may not exceed $\sqrt{n}$, the gap between ℓ_∞^n and $\ell_2^n \subset \mathcal{C}$.
The proof of the theorem uses the following lemma.

Lemma 5.3. *For any $d \in \mathcal{C}$ on $\mathbb{R}^n$, there exists a* norm $\| * \| \in \mathcal{C}$ *on $\mathbb{R}^n$, such that*

$$\mathrm{dist}(\ell_\infty^n, \| * \|) \;\leq\; \mathrm{dist}(\ell_\infty^n, d) \;. \tag{3}$$

The proof of this lemma appears in the Appendix.
Next, we need the following quantitative version of a theorem by James [15], communicated to us, together with an outline of its proof, by W.B. Johnson and G. Schechtman:

Theorem 5.4. *Assume that $\gamma = (1+\delta)^{2^r}$, and $n \geq k^{2^r}$, where $r, k \in \mathbb{N}$, and $0 \leq \delta < 1$. Then, if an n-dimensional norm $\| * \|$ is γ-close (in the sense of metric distortion) to ℓ_∞^n, then there exists a subspace L of $\mathbb{R}^n$ of dimension* $\dim(L) = k$, *such that the restriction of $\| * \|$ to L is $\frac{1+\delta}{1-\delta}$-close to an ℓ_∞ norm on L.*

The theorem as stated follows from a lemma from [20], pp.74-75, which establishes L of dimension k, such that the restriction of $\| * \|$ to L satisfies

$$\| \sum_i \alpha_i v_i \| \;\leq\; (1+\delta) \cdot \max_i \; |\alpha_i| \cdot \|v_i\| \,,$$

and a simple claim [W.B. Johnson and G. Schechtman, private communication] that

$$\| \sum_i \alpha_i v_i \| \;\leq\; (1+\delta) \cdot \max_i |\alpha_i| \cdot \|v_i\| \quad\Rightarrow\quad \| \sum_i \alpha_i v_i \| \;\geq\; (1-\delta) \cdot \max_i |\alpha_i| \cdot \|v_i\| \;.$$

Finally, we directly approach Theorem 5.2. Assume for simplicity that n is of the form $n = k^{2^r}$. Since the metric $\mu_k \notin \mathcal{C}$, being a metric on k points, isometrically embeds into ℓ_∞^k, we conclude by Theorem 5.4 that for any $\mathcal{C}$-norm $\| * \|$ on $\mathbb{R}^n$ it holds

$$\mathrm{dist}(\| * \|, \ell_\infty^n) \;\geq\; \left(1 + \frac{\beta-1}{\beta+1}\right)^{2^r} \;.$$

The same estimate holds, by Lemma 5.3, for any metric $d \in \mathcal{C}$ on $\mathbb{R}^n$. Thus, for such n, the theorem holds with constant $\alpha = \log_k \left(1 + \frac{\beta-1}{\beta+1}\right)$.

If n is not of the form k^{2^r}, take the largest such power $\leq n$, at the cost of paying an extra factor $1/2$ in the above α. This concludes the proof of Theorem 5.2.

6 Concluding remarks

We already mentioned the main open problem in the introduction, namely, to understand to what extent we can understand metrics whose small sets embed isometrically into ℓ_1. Theorem 3.8 provides a good starting point for a further research.

Approximating general (or special) metrics by metrics from some nontrivial normal class $\mathcal{C}$ may have interesting algorithmic applications. We conjecture that for any $n \in \mathbb{N}$, there exists an n-point metric d_n such that $\text{dist}(d_n \hookrightarrow C) \geq \Omega(\log_n^\alpha)$ for some constant $\alpha > 0$ depending on $\mathcal{C}$. A corresponding upper bound with $\alpha < 1$ would be most interesting. Regarding special metrics, it would be interesting to show, e.g., that any planar metric can be approximated by a metric in $\mathcal{M}_6$ with *constant* distortion. This is closely related to the famous question about ℓ_1-embeddability of planar metrics (see, e.g., [14]). Gupta [13] showed that planar metrics embed with constant distortion into NEG, and hence into $\mathcal{M}_{2k}$.

It might be of interest to study the implications of a local property on a different global property. For example, the extremal metrics constructed in the proof of Theorem 4.3, while far from being ultrametrics, are essentially very simple metrics. In particular, they are outerplanar and, up to a factor of $\pi/2$, Euclidean. It makes sense to ask, for example, if metrics that are locally almost ultrametric are globally almost ℓ_1 metrics.

The findings of this paper and other results indicate that the shortest path metrics of random k-regular graphs have a surprisingly simple local structure. Further research leading to a better understanding of this local structure, may prove useful for constructing lower bounds.

Acknowledgments

We are deeply grateful to G. Schechtman and B.W. Johnson for their explanation of Theorem 5.4. We also thank J. Matousek for pointing out important references.

References

[1] I. Althoefer, G. Das,D. Dobkin, D. Joseph and J. Soares, "On generating sparse spanners for weighted graphs", *Discrete and Computational Geometry* 9 (1993), pp.81-100.

[2] S. Arora, B. Bollobas, and L. Lovász. Proving integrality gaps without knowing the linear program. In *Proc. of the 43rd Ann. IEEE Symp. on Foundations of Computer Science*, 2002.

[3] S. Arora, J. Lee, and A. Naor. Euclidean distortion and sparsest cut. *Manuscript, Nov 2004*.

[4] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *36th Annual Symposium on the Theory of Computing*, pages 222–231, 2004.

[5] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. SIAM Journal on Computing, 27(1):291–301, 1998.

[6] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric Ramsey type phenomena. *Annals of Mathematic*, to appear. Preliminary version appeared in *Proc. STOC '03*.

[7] W. Brinkman and M. Charikar. On the impossibility of dimension reduction in ℓ_1. IEEE FOCS 2003.

[8] S. Chawla, A. Gupta, and H. Räcke. An improved approximation to sparsest cut. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, Vancouver, 2005. ACM. To appear.

[9] C. Chekuri, A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Embedding k-outerplanar graphs into l_1. In *Proc. of the ?? Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 527–536, 2003.

[10] M.M. Deza and M. Laurent. *Geometry of Cuts and Metrics.* Springer Verlag, 1997.

[11] M.M. Deza and H. Maehara. Metric transform and Euclidean embeddings. *Transactions of the American Mathematical Society*, 317:661–671, 1990.

[12] M. Farach-Colton. (Nearly) optimal embeddings into additive and ultrametric trees. *DIMACS Workshop on Discrete Metric Embeddings*, Princeton, August 2003.

[13] A. Gupta. *Embeddings of Finite Metrics.* Ph.D. thesis, University of California, Berkeley. August 2000.

[14] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and ℓ_1-embeddings of graphs. In *Proc. of the 40th Ann. IEEE Symp. on Foundations of Computer Science*, pages 399–408, 1999.

[15] R.C. James. Uniformly non-square Banach spaces. *Annals of Mathematics*, 80:542–550, 1964.

[16] W.B. Johnson and G. Schechtman. Private communication, 2003.

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

[18] J. Matousek, Ramsey-like properties for bi-Lipschitz mappings of finite metric spaces, Comment. Math. Univ. Carolinae 33 (1992), 451-463.

[19] K. Menger. Untersuchungen über allgemeine Metrik. *Mathematische Annalen*, 100:75–163, 1928.

[20] V. Milman and G. Schechtman. *Asymptomic Theory of Finite Dimensional Spaces.* LNM 1200, Springer-Verlag 1986.

[21] Y. Rabinovich and R. Raz. Lower bounds on distortion of embedding finite metric spaces in graphs. *Discrete and Computational Geometry*, 19:79–94, 1998.

A Proof of Lemma 5.3:

W.l.o.g., in what follows we restrict our attention to d's dominating ℓ_∞^n, and respectively, speak of (supremum) stretch incurred by d instead of speaking of distortion. It will be convenient to bring the discussion back to the realm of discrete metric spaces. Instead of proving (3) for $\mathbb{R}^n$, we shall prove it for $\mathbb{Z}^n$. Clearly, this is a fully equivalent statement. Observe that a norm on $\mathbb{Z}^n$ is just a translation-invariant scalable metric.

First, we construct a translation-invariant metric $d_* \in \mathcal{C}$ on $\mathbb{Z}^n$, such that the stretch incurred by d_* is no more than that of d. The construction is as follows. Given d and a point $p \in \mathbb{Z}^n$, define a metric d^{+p} on $\mathbb{Z}^n$ by

$$d^{+p}(x,y) \;=\; d(x+p, y+p)\,.$$

Observe that by the symmetry of $\mathcal{C}$, $d^{+p}(x,y) \in \mathcal{C}$. Moreover, it dominates $\mathbb{Z}^n$ equipped with the ℓ_∞^n metric, and has the same stretch as d.

Consider a sequence of metrics $d = d_0, d_1, d_2, \ldots$ defined by:

$$d_i \;=\; \frac{1}{|[-i..i]^n|} \sum_{p \in [-i..i]^n} d^{+p} \,.$$

Clearly, $d_i \in \mathcal{C}$, it dominates the ℓ_∞^n metric, and the stretch incurred by d_i is no more than that incurred by d. Observe also that For every $x, y \in \mathbb{Z}^n$ we have

$$\begin{aligned}
& \lim_{i\to\infty} |d_i(x,y) - d_i(0, y-x)| \\
= \;& \lim_{i\to\infty} \left| \frac{1}{(2i+1)^n} \sum_{p \in \{-i,\ldots,i\}^n} d(x+p, y+p) \;-\; \frac{1}{(2i+1)^n} \sum_{p \in \{-i,\ldots,i\}^n} d(p, y-x+p) \right| \\
\le \;& \lim_{i\to\infty} \frac{1}{(2i+1)^n} \sum_{p \in (\{-i,\ldots,i\}^n \triangle (\{-i,\ldots,i\}^n - x))} d(x+p, y+p) \\
\le \;& \lim_{i\to\infty} \frac{1}{(2i+1)^n} \cdot 2n \cdot \|x\|_\infty \cdot (2i+1)^{n-1} \cdot \mathrm{dist}(\ell_\infty^n, d) \|x-y\|_\infty \\
= \;& 0.
\end{aligned}$$

Next, we employ the following standard procedure. Order all vectors of $\mathbb{Z}^n$ in some order $v_1, v_2, v_3, \ldots$. Consider an infinite subsequence of $\{d_i\}$ such that the value of $d_i(0, v_1)$ converges on it; call this limit $\nu(v_1)$. Do the same with the latter subsequence to obtain $\nu(v_2)$ and a sub-subsequence, and continue in the same manner ad infinitum. Finally, for each $x, y \in \mathbb{Z}^n$, define

$$d_*(x,y) \;=\; \nu(y-x)\,.$$

The above observation implies that d_* is indeed a translation-invariant metric. Clearly, $d_* \in \mathcal{C}$, it is ℓ_∞^n-dominating, and the stretch incurred by it is bounded by the stretch incurred by d.

Second, we use d_* to construct $d_{**} \in \mathcal{C}$ with the same properties, which is not only translation-invariant, but also scalable. The construction is similar to the previous one, but is a bit simpler. Consider a sequence of translation-invariant metrics $d^{(0)}, d^{(1)}, d^{(2)}, \ldots$ defined as follows:

$$d^{(r)}(x,y) \;=\; 2^{-r}\, d_*(2^r \cdot x,\, 2^r \cdot y)\,.$$

Observe that $d^{(r)}$'s are (pointwise) monotone non-increasing with r, since for any $a \in \mathbb{N}^+$, and for $a = 2$ in particular, $d^*(ax, ay) \;\le\; a d_*(x,y)$ due to translation-invariance of d^*.

Taking the limit of $d^{(r)}$'s we obtain the desired d_{**}. It is easy to check that d_{**} has all the required properties. E.g., the scalability holds, since, by the previous observation, the limit $\lim_{r\to\infty} a^{-1} d^{(r)}(ax, ay)$ exists for every natural a. Therefore d_{**} is scalable with respect to all $a \in \mathbb{N}^+$, and hence with respect to all $a \in \mathbb{Q}^+$, as required. □