SIAM J. COMPUT.
Vol. 0, No. 0, pp. 000–000
© 2012 Society for Industrial and Applied Mathematics

LOCAL VERSUS GLOBAL PROPERTIES OF METRIC SPACES*

SANJEEV ARORA†, LÁSZLÓ LOVÁSZ‡, ILAN NEWMAN§, YUVAL RABANI¶, YURI RABINOVICH§, AND SANTOSH VEMPALA‖

Abstract. Motivated by applications in combinatorial optimization, we study the extent to which the global properties of a metric space, and especially its embeddability into ℓ_1 with low distortion, are determined by the properties of its small subspaces. We establish both upper and lower bounds on the distortion of embedding locally constrained metrics into various target spaces. Other aspects of locally constrained metrics are studied as well, in particular, how far are those metrics from general metrics.

Key words. sparsification, dimension-reduction

AMS subject classifications. 68W40, 68T05, 05E45, 52C45, 60B05

DOI. 10.1137/090780304

1. Introduction.

Suppose that we are given a finite metric space (X, d), and we are told that the induced metric on every small subset of X embeds isometrically into ℓ_1. What can we say about the distortion of embedding the entire metric into ℓ_1? In this paper we study this and related questions.

One reason to study such problems is the intimate relation between certain questions concerning low-distortion embeddings and problems in combinatorial optimization. In particular, since finite ℓ_1 metrics correspond exactly to combinations of cuts (see, e.g., [12]), good approximation algorithms for embedding metric spaces into ℓ_1 yield good approximation algorithms for NP-hard cut problems such as the *sparsest cut*. Indeed, the best known algorithms for the sparsest cut problem used this relation; see, e.g., [24, 6, 5, 9, 4]. The first observation is that the exact solution to this problem can be obtained by optimizing a linear function over the convex cone of all ℓ_1 metrics on n points. However, expectedly, even deciding the membership in this cone is NP-Complete. It is thus a natural strategy to relax this cone to a larger, more tractable metric cone, optimize over it, and then find an ℓ_1 metric d that approximates the optimal non-ℓ_1 metric d'. The price in the objective function, in moving from d' to the approximate d, turns out to be precisely the multiplicative distortion between d' and d (see section 2 for exact definitions). In the earlier papers from the list above, the relaxed cone is the cone of all metrics on n points. In the later papers the relaxed cone is the cone of all negative type metrics on n points. Finding the approximating ℓ_1-metric d is the crux of this approach in both cases, and is highly nontrivial.

*Received by the editors December 16, 2009; accepted for publication (in revised form) November 6, 2011; published electronically DATE. An extended abstract of this paper has appeared in SODA'06.
http://www.siam.org/journals/sicomp/x-x/78030.html

†Department of Computer Science, Princeton University, Princeton, NJ 08544 (arora@cs.princeton.edu). This author's work was supported by a David and Lucile Packard Fellowship and NSF grant CCR 0205594.

‡Microsoft Research, Redmond, WA 98052, and Department of Computer Science, Loránd University, Budapest, Hungary (lovasz@microsoft.com).

§Computer Science Department, University of Haifa, Haifa 31905, Israel (ilan@cs.haifa.ac.il, yuri@cs.haifa.ac.il). The fifth author's work was supported in part by grant ISF-247-02-10.5.

¶Computer Science Department, Technion, Haifa 32000, Israel (yrabani@cs.huji.ac.il).

‖Mathematics Department, MIT, Cambridge, MA 02139 (vempala@math.mit.edu). This author's work was supported in part by NSF CCR-0312339 and a Sloan Foundation Fellowship.

In this paper we examine the possibility of using other relaxed metric cones, or classes of metrics containing the ℓ_1 metrics. Making some natural assumptions about them (e.g., being closed under taking submetrics, being closed under additions, being closed under the permutations of the points of the underlying space, etc.) we arrive at a notion of *baseline* classes of metrics. While the bulk of this paper is dedicated to the study of special baseline classes of metrics defined in terms of their local properties, one section deals with issues in the theory of general baseline metric classes.

We start with a detailed study of some classes of metrics whose *local structure* is close to that of ℓ_1 metrics. For example, for the metrics on a large set of points with the property that for every subset of points of size k (for small enough k), the induced metric embeds isometrically into ℓ_1. This metric class is interesting in its own right, it allows fast membership testing, and it supports fast LP-based algorithms as long as k is not too large. Such metric classes arise naturally when applying k rounds of a lift-and-project procedure such as Lovász–Schrijver, Sherali–Adams, or Lasserre to cones containing the ℓ_1 metrics. These procedures start with an initial LP or SDP relaxation and then systematically derive and plug in, round after round, all inequalities valid for the current integral hull. It turns out that the better (as a function of k) such metric classes approximate the original ℓ_1 cone, the faster should the related lift-and-project procedure converge to the optimum. We address the interested reader to [3, 16].

Other metric classes of interest whose local structure resembles that of ℓ_1 include the k-gonal metrics, and the metrics for which all induced k-size submetrics embed into ℓ_1 with distortion c. See section 3.2 for precise definitions, discussion, and some results about such metrics. In general, the hope is that by imposing a simple (and useful) local structure on metrics, e.g., that all submetrics that are induced on subsets of k points are close to ℓ_1 in sense of low multiplicative distortion, one should get metrics with a simple global structure; e.g., the entire metric is also close to ℓ_1. Namely, we seek local-global phenomena in the context of finite metric spaces.

Our main results can be described as follows. First, we show (Theorem 3.1) that if every subset of size $\frac{n}{c}$ of an n-point metric space embeds isometrically into ℓ_1, then the entire space embeds into ℓ_1 with distortion $O(c^2)$. A similar result holds if the isometric embedding of subsets is replaced by low-distortion embedding. Using the reduction of the sparsest cut problem to ℓ_1 embeddability [24, 6], this yields a $2^{o(n)}$-time $O(c^2)$-approximation algorithm for sparsest cut for any $c = c(n) \to \infty$. We note that recent reductions [10] show that such an approximation is hard (namely, they cannot be done in $poly(n)$), assuming the unique games conjecture [22].

Next, in Theorem 3.7, we show a negative result by constructing metric spaces where every k-size subset embeds isometrically into ℓ_1 (or even in ℓ_2), yet the entire space requires distortion $(\log n)^{\Omega(\frac{1}{k})}$ when embedded into ℓ_1. Thus, these metrics require a superconstant distortion as long as $k = o(\log\log n)$. This construction also implies a polylogarithmic separation between ℓ_1 and k-gonal metrics (namely, there are k-gonal metrics for $k = \theta(1)$ that require $(\log n)^{\theta(1)}$-distortion when embedded into ℓ_1). This answers in negative an open question by Deza whether any 5-gonal metric embeds into ℓ_1 with a constant distortion [26].

In a next result, we consider the setting where the small subsets are embeddable into ℓ_1 with constant distortion, rather than being *isometrically* embeddable into ℓ_1. We construct (in Theorem 3.9) n-point metrics that require $\Omega(\log n)$ distortion to embed into ℓ_1, but every subset of size $n^{1-\epsilon}$ embeds into ℓ_1 with distortion $O(1/\epsilon^3)$. This result indicates that the local structure of a metric (in the context of distortion of embedding into ℓ_1) has a very limited effect on its global structure, making our

algorithmic approach much less promising. A significant contribution of our work is a new insight into shortest path metrics derived from random graphs of bounded degree. These metrics are known to be extremal for many metric-theoretic properties, e.g., embeddability into ℓ_1. Surprisingly, their local structure turns out to be possibly rather simple even when the size of the submetrics is as large as $n^{1-\epsilon}$ (see section 3.3). In addition, some new building bocks for obtaining the upper bound (on the distortion of small submetrics) are constructed for the proof of this result. We believe that these new methods for approximate embedding into ℓ_1 are of independent interest.

Having established these results, we proceed to study the local-global phenomena in other settings. First, we show that a result similar to Theorem 3.9 holds when instead of measuring the proximity to ℓ_1 one measures the proximity to ℓ_2, the Euclidean spaces. Namely, there exist metric spaces of size n such that all their subspaces of size $n^{1-\epsilon}$ are almost Euclidean, while the entire space requires $\sqrt{\Omega(\log n)}$ distortion when embedded into a Euclidean metric (of any dimension).

We also study the local-global phenomena in the context of ultrametrics, i.e., metrics satisfying the following (local) condition: $\forall x, y, z,\ d(x,z) \leq \max\{d(x,y), d(z,y)\}$. These metrics are a class of tree metrics used, e.g., in hierarchical clustering and metric Ramsey theory (see [15, 14, 7] and the references therein). In particular, they form a (very restricted) subset of ℓ_2 metrics. By definition, if every subset of size three is an ultrametric, then so is the whole space. On the other hand, we show (Theorem 5.3) that the situation changes dramatically if we relax the requirement that small subsets be isometrically embeddable into an ultrametric to that they be embeddable with small distortion. For every c and k we construct metric spaces on n points such that every subset of cardinality $(n-1)^{1/k}$ embeds into an ultrametric with distortion bounded by c, yet the entire metric space does not embed into an ultrametric with distortion less than c^k. We show that this bound is essentially tight by establishing a nearly matching upper bound on the distortion.

The last issue addressed in this paper is motivated by the negative result of Theorem 3.9, namely that locally constrained metrics can be as far from ℓ_1 (require large multiplicative distortion) as the general metrics. Could it be that for any fixed k, the class of the general metrics is equivalent, up to a constant multiplicative distortion, to the class of k-locally constrained metrics? A positive answer would yield a very good understanding of such metrics. We show, however, that the answer is negative. Furthermore it applies not only to the k-locally constrained metrics, but also for any nontrivial class $\mathcal{C}$ of baseline metrics. A similar conclusion can also be deduced from a theorem of Matousek from [25]. Yet, unlike the result in [25], our lower bound is of a quantitative nature. We show that a nontrivial baseline class of metrics over the entire R^n is $\Omega(n^\alpha)$ far from ℓ_∞^n for a suitable α.[1]

Local-global phenomena play an important role in many areas such as the construction of probabilistically checkable proofs, program checking, property testing, etc. In these settings one usually has to infer a global property from the knowledge that the local property holds for many (but not necessarily all) small subsets. Although we are mostly interested here in the situation where the local conditions are required for all small subsets, we establish at least one positive result for the case that the local constraints hold only for "many" small subsets (see Lemma 3.4).

Let us remark that our results are related to the study of Ramsey phenomena in metric spaces, a line of work motivated both by some lower bound techniques in online computation, and by deep questions about the local theory of metric spaces (see [7]

[1]Note that here n is the dimension rather than the cardinality of the subspace.

and the references therein). For example, the main lower bound of section 3.3 provides a (weak) partial answer to an open question from [7]. Ramsey theory, in general, shows that in the midst of global "disorder" there is always a significant subset exhibiting "order." In this phrasing, our upper bounds suggest that global "disorder" implies the existence of large "disordered" subsets. For example, Theorem 3.1 implies that if a metric space on n points does not embed into ℓ_1 with distortion α, there must be a subset of the points (in fact, many subsets) of cardinality $O(n\sqrt{\beta/\alpha})$ for which the induced metric does not embed into ℓ_1 with distortion less than β.

Subsequent work. Since the appearance of the preliminary version of this paper in SODA'06, some of its results have been improved in two main directions. Charikar, Makarychev, and Makarychev [8] start where this paper ends and, using finer yet related methods, improve the results of sections 3.1 and 3.3. In particular, the upper bound of Theorem 3.1 is improved from $O(c^2)$ to the essentially optimal $\tilde{O}(\log c)$. The lower bound of Theorem 3.9 about the existence of n-point metric spaces where every subset of size $k = n^{1-\epsilon}$ embeds into ℓ_1 with a constant distortion c_ϵ while the entire space requires $\Omega(\log n)$-distortion to embed into ℓ_1 is shown to (almost) carry on even if all the small subsets embed *isometrically* into ℓ_1. This is a huge improvement over our Theorem 3.9, and it implies that imposing k-local restrictions on the metric cone cannot significantly help, e.g., in resolving the sparsest cut problem.

An important paper of Mendel and Naor [27], motivated by ideas from Banach spaces and the present paper, contains a proof of a conjecture from the conference version of this paper, concerning the universality of nontrivial baseline metric classes. In fact, the only property of the metric class they use is that it is closed under taking submetrics.

We think that some results of the present paper, as well as some of the methods used in their proofs, are of independent value. A number of questions and research directions raised here remain unexploited. We delay the discussion of some of them to the concluding remarks.

2. Preliminaries. A *distance space* $A = (X, d)$ is a pair where X is a set and $d : X^2 \mapsto R^+$ is a nonnegative symmetric function (that is, $d(x, y) = d(y, x)$), for which $d(x, x) = 0$ for every $x \in X$. Let $A = (X, d)$ and $B = (X', d')$ be two distance spaces. An embedding of A into B is just a map $\phi : X \mapsto X'$. The *multiplicative distortion*, or simply the *distortion* of embedding A into B, is defined as[2]

$$\mathrm{dist}(d, d') \;=\; \min_{\phi:X\to X'} \max_{x,y\in X} \frac{d'(\phi(x), \phi(y))}{d(x, y)} \;\cdot\; \max_{x,y\in X} \frac{d(x, y)}{d'(\phi(x), \phi(y))} \;.$$

We call the first term in the above expression the *stretch* of d' w.r.t. d, and the second term the *contraction* of d' w.r.t. d. Sometimes we refer to the expression above as the distortion between d and d' (rather than A and B), when the exact point set plays no particular importance. Similarly, we sometime refer to d instead of (X, d) when X is understood from the context. For two distance functions d, d' and a real $\gamma \geq 1$, we say that d is γ-close to d' if d can be embedded into d' with distortion upper bounded by γ. Distance functions are Real-functions; thus we can add distance functions, multiply by a constant, etc. We say that (X, d) dominates (X, d') (or just "d dominates d'") if $d(x, y) \geq d'(x, y)$ for every two points $x, y \in X$.

For a class $\mathcal{C}$ of distance functions, we use $\mathrm{dist}(d \hookrightarrow \mathcal{C})$ to denote the minimum distortion between d and $d' \in \mathcal{C}$. If $\mathrm{dist}(d \hookrightarrow \mathcal{C}) \leq \gamma$, we say that d is γ-close to $\mathcal{C}$.

[2] If $d'(x, y) = 0$ for every pair of points x, y while $d(x, y) > 0$ for some pair, we define $\mathrm{dist}(d, d') = \infty$.

A *metric* space is a distance space that in addition satisfies the triangle inequality (sometimes also called "semimetric," as we allow 0-distance between two distinct points). We denote by MET the set of all metric spaces. For some background on metric spaces, see [12].

Let d be a distance function (on an underlying point set X), and let $f : R \to R$ be a monotonically nondecreasing 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: $\mathrm{dist}(d^c, (d')^c) = (\mathrm{dist}(d, d'))^c$.

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.

We use the terms Euclidean metrics and ℓ_2-metrics interchangeably. The *negative type* metrics, NEG $\cap$ MET, mentioned in the introduction, are the squares of Euclidean metrics that satisfy the triangle inequality.

A rather large collection of metric spaces that we define below is the *baseline* class of metrics. It can be easily verified that the classes ℓ_1, NEG $\cap$ MET, HYP, and $\mathcal{M}_k$ to be discussed later (see section 3.3) are all baseline.

DEFINITION 2.1 (baseline sets of metrics). *A nontrivial class of metrics $\mathcal{C}$ (that is, it contains a metric that is not everywhere zero) is called baseline if it satisfies the conditions below:*

1. *It is invariant under permutation of points; namely, for every $(X, d) \in \mathcal{C}$, any metric d' derived from d by permuting the members of X is also in $\mathcal{C}$.*
2. *It is a closed cone. That is, for every $(X, d), (X, d') \in \mathcal{C}$, and every two nonnegative Reals, $a, a' \geq 0$, $a \cdot d + a' \cdot d' \in \mathcal{C}$.*
3. *It is hereditary. Namely, for every $(X, d) \in \mathcal{C}$ and a subset $Q \subseteq X$, $(Q, d^Q) \in \mathcal{C}$.*
4. *For every $(X, d) \in \mathcal{C}$, consider any metric d', obtained from d by performing the following* cloning *operation: Pick any point $p \in X$, add a "clone" $q \notin X$, and set $d'(p, q) = 0$ and $d'(q, x) = d(p, x)$ for all points x. Then, $(X \cup \{q\}, d') \in \mathcal{C}$.*

Observe, by item 4 above, that every baseline set of metrics includes all cut metrics,[3] and therefore all metrics that embed isometrically into ℓ_1. Further notice that if $\mathcal{C}$ is a baseline 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 baseline.

3. Baseline sets of metrics. Here we study the local versus global properties of general metrics w.r.t. embeddability into a baseline set of metrics. Namely, our assumption will be that the metric restricted to each small subset embeds well (either isometrically, or with small distortion) into a baseline class $\mathcal{C}$, and we study to what extent this guarantees that the whole metric embeds well into $\mathcal{C}$. We prove both upper bounds on the distortion of the whole metric, along with some algorithmic applications, and then lower bounds for two different assumptions on the local constraints.

3.1. Upper bounds—Local low distortion implies global low distortion. The main result of this section is the following upper bound on the distortion, implied

[3]For a subset $S \subset X$, the cut metric δ_S on X is defined as $\delta_S(x, y) = 1$ if $x \in S$, $y \notin S$, and 0 otherwise. The fact that ℓ_1 metrics are nonnegative linear combinations of cut metrics is basic in the theory of metric spaces; see, e.g., [12].

by the local properties of the metric.

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

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

We use the following definition. Let (X, d) be a metric space. A *tree-like extension* of (X, d) is a metric space obtained from (X, d) by repeatedly performing the following *attachment* operation: Pick a point $p \in X$ and a weight $w \geq 0$, and "attach" to p a new point $q \notin X$ by an edge of weight w, i.e., set $d'(q, x) = d'(p, x) + w$ for all points $x \in X$, and augment X by q.

LEMMA 3.2. *Let $\mathcal{C}$ be a baseline 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, $d' \in \mathcal{C}$. ∎

Proof of Theorem 3.1. Let d be a metric on a finite set of points $X = \{p_1, p_2, \ldots, p_n\}$. It will be assumed w.l.o.g. that for any $p_i \in X$, the distances between p_i and the other points in X are all distinct. (Otherwise just perturb the distances slightly keeping the triangle inequality. This will only introduce a very small distortion.) Let $\sigma \in S_n$ be a permutation on $\{1, 2, \ldots, n\}$. The metric d^σ is defined as follows. We start with restriction of d to $\{p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\}$. Then, for $i = m+1, \ldots, n$, the new point $p_{\sigma(i)}$ is attached to $p_{\sigma(i^*)}$ at distance w_i^σ, where $i^* \in \{1, 2, \ldots, i-1\}$ minimizes $d(p_{\sigma(i)}, p_{\sigma(i*)})$, and $w_i = d(p_{\sigma(i)}, p_{\sigma(i^*)})$. Finally, we define d^* on X as the average of all d^σ's: For every $p, q \in X$,

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

We will next show that $d^* \in \mathcal{C}$ and that $\mathrm{dist}(d, d^*)$ is small, which will complete the proof. In the following m is fixed, while n and d vary. Let $T_{n,m}$, $n \geq m$, denote the supremum of $\mathrm{dist}(d, d^*)$, over all n-point metrics, d. Clearly, $T_{m,m} = 1$ since $d^* = d$ for this case. Notice that for every $p, q \in X$, $d^\sigma(p, q) \geq d(p, q)$, and therefore d^* dominates d. To bound the stretch, observe that

$$\begin{aligned}(1)\ d^*(p, q) &= \mathrm{E}_\sigma[d^\sigma(p, q)] \\ &= \frac{2}{n} \cdot \mathrm{E}_\sigma\left[d^\sigma(p, q) \,|\, \sigma(n) \in \{p, q\}\right] + \left(1 - \frac{2}{n}\right) \cdot \mathrm{E}_\sigma\left[d^\sigma(p, q) \,|\, \sigma(n) \notin \{p, q\}\right].\end{aligned}$$

Notice that $\mathrm{E}_\sigma\left[d^\sigma(p, q) \,|\, \sigma(n) \notin \{p, q\}\right] \leq T_{n-1,m} \cdot d(p, q)$.

For the case $\sigma(n) = p$ let $p^* \in X$ be the point in X that is closest to p. By our assumptions, p^* is unique and hence it will be the point to which p will be attached. As $d(p, p^*) \leq d(p, q)$, the triangle inequality implies that $d(p^*, q) \leq 2d(p, q)$ and thus,

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

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

$$T_{n,m} \leq \left(1 - \frac{2}{n}\right) \cdot T_{n-1,m} + \frac{2}{n} \cdot (2T_{n-1,m} + 1) = \left(1 + \frac{2}{n}\right) \cdot T_{n-1,m} + \frac{2}{n}. \tag{3}$$

Solving the recurrence, we get that $T_{n,m} = O\big(\big(\frac{n}{m}\big)^2\big)$.

Next, recall that $\mathcal{C}_\gamma$ is a baseline set of metrics. By the conditions of the theorem, for every m-point subset Q, $d^Q \in \mathcal{C}_\gamma$. Therefore, Lemma 3.2 implies that $d^\sigma \in \mathcal{C}_\gamma$ for every permutation σ. As $\mathcal{C}_\gamma$ is a closed cone, this further implies that $d^* \in \mathcal{C}_\gamma$. Finally, since $\mathrm{dist}(d, d^*) = O(\big(\frac{n}{m}\big)^2)$, the theorem follows. ∎

THEOREM 3.3. *A metric $\tilde{d} \in \mathcal{C}$, which is an embedding of d satisfying the statement of Theorem* 3.1*, can be computed in randomized polynomial time.*

Proof. We use the same terminology as in the proof of Theorem 3.1. The construction of $\tilde{d}$ is based on the construction of d^*. The difficulty with the explicit construction of d^* is that it is written as an average of exponentially many metrics. We show that this average can be estimated using a much smaller space of metrics. Let $K \subseteq S_n$ be a subset of permutations, with $|K|/|S_n| = \kappa \leq 1$. Extending the definition of d^*, let $d^*_K = \mathrm{E}_\sigma[d^\sigma \,|\sigma \in K]$. Observe that

$$d \leq d^*_K \leq \kappa^{-1} d^*. \tag{4}$$

The first inequality holds since each d^σ dominates d and the second since $d^* = d^*_K \cdot \kappa + d^*_{\bar{K}} \cdot (1 - \kappa)$.

Let $G = \{\sigma \,|\, d^\sigma \leq 10n^2 T_{n,m} \cdot d\} \subseteq S_n$. By Theorem 3.1, the expected stretch of d^* with respect to any pair of points in the space is $\leq T_{n,m}$. Therefore, using Markov's inequality and the union bound on all pairs we get that $|G| \geq 0.9|S_n|$. Equation (4) and the fact that d^*_G dominates d imply then that

$$d \leq d^*_G \leq 1.1 d^* < 1.1 T_{n,m}. \tag{5}$$

Next, let $\tilde{G}$ be a random sample from G of size N and let $d^*_{\tilde{G}} = \frac{1}{N} \sum_{\sigma \in \tilde{G}} d^\sigma$ be the average of the sampled metrics. Thinking of d^σ as a random variable (defined by the uniform distribution on G), for two fixed points, $x, y \in X$, d^σ takes values in the interval $[d(x,y), 10n^2 T_{n,m} d(x,y)]$ and with expectation $d^*_G(x,y)$.

By Hoeffding's large deviation bound [19] (see also[4] exercise 4.7, page 98 in [29]), for any pair of points x, y in the space,

$$\begin{aligned}\Pr\left[d^*_{\tilde{G}}(x,y) > 3.3 T_{n,m} d(x,y)\right] &\leq \Pr\left[d^*_{\tilde{G}}(x,y) - d^*_G(x,y) > 2.2 T_{n,m} d(x,y)\right] \\ &\leq e^{-\frac{2(2.2)^2 T^2_{n,m} \cdot N}{(10n^2 T_{n,m} - 1)^2}} \leq e^{-\frac{8N}{100n^4}},\end{aligned}$$

where the first inequality is by (5). Thus, choosing $N = 50n^4 \log_2 n$, this probability is smaller than $1/n^4$. Using the the union bound on all pairs of points from X, we conclude that $d^*_{\tilde{G}} \leq 3.3 d^*$ with probability close to 1. Finally, to create a random sample $\tilde{G}$, we sample permutations σ from S_n, construct d^σ, and discard d^σ if it stretches an edge by more than $10 T_{n,m} \log_2 n$. Since a 0.9 fraction of the permutations in S_n are in G, this gives a polynomial time randomized algorithm. ∎

The proof of Theorem 3.3 has the following interesting structural implication.

[4] We use the standard statement for Hoeffding's bound. Note that in [29] the random variable is assumed to take values in $[0, 1]$.

LEMMA 3.4. *The assumption of Theorem* 3.1 *that* all *size-m subspaces are γ-close to $\mathcal{C}$ can be replaced by a weaker assumption that only a κ-fraction of the size-m subspaces have this property, at the cost of an additional multiplicative factor of κ^{-1} in the upper bound.*

Proof. For $\sigma \in S_n$ let its m-prefix denote the set of points $S_\sigma = \{p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(m)}\}$. We say that $\sigma \in S_n$ is *good* with respect to d if the restriction of d to S_σ, d^{S_σ}, is γ-close to $\mathcal{C}$.

By the assumption of the lemma, the good permutations constitute a κ-fraction of all permutations in S_n. Thus (4) applies for K, the set of all good permutations. The proof then follows along the lines of the proof of Theorem 3.1. ∎

We now show that Theorems 3.1 and 3.3 imply a subexponential time algorithm for approximating sparsest cut to within any superconstant factor.

THEOREM 3.5. *Sparsest cut can be approximated to within a factor of $O(c^2)$ in time* $\exp\left(\frac{n \log c}{c}\right)$*, where n is the number of nodes in the input graph.*

Proof. Let (G, w, T, h) be an instance of sparsest cut. Here $G = (V, E)$ is an undirected graph with $|V| = n$, $w : E \to N$ is a weight function on the edges of G, $T = \{(s_1, t_1), (s_2, t_2), \ldots, (s_k, t_k)\}$ is a set of pairs of nodes of G (called terminals), and $h : T \to N$ is the demand function. Let D be the set of semimetrics d on V, such that for every $U \subset V$ with $|U| \le \frac{1}{c}|V|$, the restriction of d to U embeds isometrically in ℓ_1. Let

$$z^* = \min \left\{ \frac{\sum_{e \in E} w(e) d(e)}{\sum_{(s,t) \in T} h(s,t) d(s,t)} : \ d \in D \right\}. \tag{6}$$

It is known that if we replace $d \in D$ with $d \in \ell_1$ in (6), we get the value of the sparsest cut (see, e.g., [12, 24]). Hence, z^* is a lower bound on the value of the sparsest cut.

On the other hand, since z^* is achieved by a metric $d \in D$, Theorem 3.1 implies that d is $O(c^2)$-close to an ℓ_1 metric d'. By the results of [24, 6], given an ℓ_1 metric d', one can find, in time polynomial in the representation of d', a cut $(S, V \setminus S)$ in G such that

$$\frac{\sum_{e \in E:\ |e \cap S| = 1} w(e)}{\sum_{(s,t) \in T:\ |\{s,t\} \cap S| = 1} h(s,t)} \le \frac{\sum_{e \in E} w(e) d'(e)}{\sum_{(s,t) \in T} h(s,t) d'(s,t)} = l(d). \tag{7}$$

Thus, as $l(d) = O(c^2 z^*)$ (since as shown above d is $O(c^2)$-close to d'), the cut for which (7) holds approximates the sparsest cut as claimed.

To compute z^* we make use of the following (quite standard) way of optimizing over ℓ_1 metrics using exponentially large linear programs: First note that by the form in (6), defining z^* is invariant to a scaling of d; hence one can impose, say, the extra condition that $\sum_{(s,t) \in T} d(s,t) h(s,t) = 1$. With this, z^* can be expressed as minimization of the linear function $\sum_{e \in E} w(e) d(e)$, for $d \in D$ and $\sum_{(s,t) \in T} d(s,t) h(s,t) = 1$. Thus it is enough to show that $d \in D$ can be expressed as a set of linear constraints.

Indeed, for d given by a set of variables $d(x, y)$, $x, y \in V$, consider the following collection of linear constraints.[5]

First, for every $x, y, z \in V$ we take the triangle inequality constraint $0 \le d(x, z) \le d(x, y) + d(y, z)$. This asserts that d is a (semi-) metric. In addition, for each set $U \subset V$ with $|U| \le \frac{1}{c}|V|$, the following set of variables is added: $\{\alpha_C^U |\ C \subseteq U\}$, and the following constraints: (a) $\alpha_C^U \ge 0$ for every C, and (b) for every $x, y \in U$,

[5]Here we consider unordered pairs; hence d is symmetric by definition.

$d(x,y) = \sum_{C \subseteq U} \alpha_C^U \delta_C(x,y)$, where $\delta_C(x,y) = 1$ if $|\{x,y\} \cap C| = 1$ and 0 otherwise. Namely, δ_C is the cut metric induced by C on U.

These constraints, for specific U, assert that the restriction of d to U is in ℓ_1 (as it is given by a positive combination of cut metrics). Thus the set of solutions to the above constraints assures that $d \in D$.

Note that there are $2^{O(n/c)} \cdot \binom{n}{\frac{n}{c}} = 2^{O(n \log c/c)}$ many variables and a similar number of constraints. Hence z^* and the corresponding d that achieves it can be computed by the linear program in time polynomial in its size, which is $2^{O(n \log c/c)}$.

Once z^* is found, to find the cut we need to compute the low-distortion embedding of d into ℓ_1, and then, as mentioned above, the cut can be found in time polynomial in the representation of d. By Theorem 3.3, it is sufficient to compute the (isometric) embedding of poly($|V|$) tree-like extensions of subsets U of size n/c (at the cost of an extra $O(1)$ factor in the approximation guarantee). The embedding of a tree-like extension of a subset U is trivial to compute, given the embedding of U, and this can be read from the solution to the linear program above (this is just the corresponding sum of cut metrics, as given by the solution to the variables α_C^U). ∎

3.2. Lower bounds—When the local structure is *isometric* to ℓ_1 or is k-gonal. We first address the situation when the subspaces embed *isometrically* into an interesting class of metrics, rather than with small distortion. The latter situation is considered in section 3.3. We need the following definitions (see, e.g., [12]).

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 $\mathcal{M}_k$ to denote the class of all k-gonal distance functions. k-gonal distance functions were studied extensively as such; see Deza and Maehara [13] and an extensive survey and reference for the following facts by Deza and Laurent [12]. Clearly, $\mathcal{M}_3$ is simply MET, the class of all metrics. Also, for every $k \in N$, $k \geq 2$, $\mathcal{M}_{k+2} \subset \mathcal{M}_k$ and $\mathcal{M}_{2k-1} \subset \mathcal{M}_{2k}$. On the other hand, for every $k \in N$, $k \geq 1$, distance functions in $\mathcal{M}_{2k}$ are not necessarily metrics; that is, they do not necessarily satisfy the triangle inequality. The class of all *negative type* distance functions is

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

Schoenberg showed that $d \in$ NEG iff $\sqrt{d}$ embeds isometrically into ℓ_2 [31]. 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.

The main theorem of this section establishes a *poly*log(n) separation between r-gonal metrics, or even locally Euclidean metrics, from NEG $\cap$ MET. In particular,

by the discussion above this also shows similar separation between locally Euclidean metrics and ℓ_1 metrics. We need the following theorem implicit in [13].

THEOREM 3.6 (see [13]). *For any metric d on k points, the following hold:*

1. $d^{\log_2(1+1/(\lceil k/2\rceil-1))} \in \mathcal{M}_k$.
2. $d^{\log_2(1+1/(k-1))}$ *is a hypermetric.*
3. $d^{\frac{1}{2}\cdot\log_2(1+1/(k-1))}$ *is Euclidean.*

The main result of this section is as follows.

THEOREM 3.7. *For every integer $n \geq 2$ and for every $k \in N$, $k \leq n$, the following statements are true:*

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

Proof. Consider, e.g., the third statement; the other two are proved in the same manner using the corresponding item in Theorem 3.6. Let D be the metric of a unit-weighted constant-degree expander on n points. It is well known that it incurs an $\Omega(\log n)$ distortion in embedding into $\textsc{neg} \cap \textsc{met}$ [24, 6]. Observe also that all the distances in D are between 1 and $\Theta(\log n)$. Fix an integer $k \geq 2$ and let $d = D^{\frac{1}{2}\cdot\log_2(1+1/(k-1))}$. We claim that d has the properties as claimed in item 3 in Theorem 3.7. Indeed d restricted to any k points is Euclidean, as asserted by the third item of Theorem 3.6. Since the maximum distance in D is $O(\log n)$, and D dominates d, we conclude that

$$\mathrm{dist}(d, D) \;=\; O\left((\log n)^{1-\frac{1}{2}\log_2(1+1/(k-1))}\right) \;.$$

On the other hand, for any $d' \in \textsc{neg}\cap\textsc{met}$, $\mathrm{dist}(D, d') = \Omega(\log n)$. Since $\mathrm{dist}(d, d')\cdot \mathrm{dist}(d, D) \geq \mathrm{dist}(D, d')$, we conclude that $\mathrm{dist}(d, d') = \Omega\big((\log n)^{\frac{1}{2}\log_2(1+1/(k-1))}\big)$, as claimed. ∎

COROLLARY 3.8. *For $k = o(\log\log n)$, there exist an n-point metric d such that every k-point subspace is in ℓ_2, yet* $\mathrm{dist}(d \hookrightarrow (\textsc{neg} \cap \textsc{met})) = \omega(1)$. *(Recall that $\ell_2 \subset \ell_1 \subset \textsc{neg}$ [12].)*

3.3. Lower bounds—When a small distortion is allowed. The main result in this section is the construction of a class of metrics on n points for any large enough n, for which the metric induced on any $n^{1-\delta}$ points embeds well into ℓ_1, while any embedding of the whole metric into ℓ_1 has distortion of $\Omega(\log n)$.

THEOREM 3.9. *For every $\delta > 0$ and for large enough n (with respect to $1/\delta$), the following statements hold:*

1. *There is an n-point metric d such that for every $n^{1-\delta}$-point subspace Q,* $\mathrm{dist}(d^Q \hookrightarrow \ell_1) = O(1/\delta^3)$, *yet* $\mathrm{dist}(d \hookrightarrow \ell_1) = \Omega(\log n)$.[6]
2. *There is an n-point metric d such that for every $n^{1-\delta}$-point subspace Q,* $\mathrm{dist}(d^Q \hookrightarrow \ell_2) = O(1/\delta^{1.5})$, *yet* $\mathrm{dist}(d \hookrightarrow \ell_2) = \Omega(\sqrt{\log n})$.

The metric we construct for the first item of Theorem 3.9 is essentially the shortest path metrics of a unit weighted, constant degree expander with some additional requirements. For the second item we just take the square root of this metric. Thus, the lower bounds claimed in the two items will automatically follow from the (by now

[6] The lower bound holds even for embedding into $\textsc{neg} \cap \textsc{met}$, the class of negative type metrics.

standard) fact that any embedding of the shortest path metric of a constant degree expander on n points into ℓ_1 incurs $\Omega(\log n)$ distortion.

The main technical contribution of this section is the corresponding upper bounds. We show that a family of expanders with certain parameters, although, as explained above, induces metrics that are hard to embed into ℓ_1, the restricted metrics on any subset of $n^{1-\delta}$ embed into ℓ_1 with low distortion. In particular, we present a new general embedding lemma for locally sparse graphs, related but much more powerful than the results of [18] on the ℓ_1 embeddability of graphs with small Euler number. We then need to construct the family of expanders with the suitable parameters, which is done using a rather standard probabilistic argument. Formally, the two parts are asserted by the following two lemmas. First we state them, then show how they imply Theorem 3.9, and then present the proofs of the lemmas.

LEMMA 3.10. *There is a universal constant c such that for any $\epsilon < 1$ and n large enough (w.r.t. $1/\epsilon$), there exists a graph G_n on n^*, with $n/2 \le n^* \le n$ vertices, of maximum degree $100c$, such that the following hold:*

1. *The diameter of G_n is at most $c \log n$.*
2. *The subgraph $G[S]$ induced by any subset of vertices $S \subseteq V(G)$ of size at most $n^{1-\epsilon}$ has at most $(1+\frac{b}{\epsilon \log n})(|S|-1)$ edges for some constant b (e.g., $b = 30$).*
3. *G is a good edge expander; namely, for every set $S \subseteq V(G)$ with $|S| \le n/2$,*

$$e(S, \bar{S}) := |\{(u,v) \in E(G)|\ u \in S,\ v \in \bar{S}\}| \ \ge\ |S|.$$

LEMMA 3.11. *Let d be the shortest path metric of an unweighted graph $H = (V, E)$ with diameter D, and with a property that the subgraph induced by any subset $S \subset V$ has at most $(|S|-1)(1+1/p)$ edges. Then, d can be embedded into ℓ_1 (in fact, into a distribution over dominating tree-metrics) with distortion $O((p+D)/p)$.*

Proof of Theorem 3.9. We start with the first item. Let $\epsilon = \delta/2$ and let $G = (V, E)$ be the graph whose existence is asserted in Lemma 3.10 for ϵ and $|V| = n$. Let d be its shortest path metric. The metric d, being a shortest path of a constant degree expander, requires distortion $\Omega(\log n)$ to embed into ℓ_1 (see, e.g., [24]).

Our goal is to show that d^S, the restriction of d to a set S of size at most $n^{1-2\epsilon}$, is embeddable into ℓ_1 with $O(1/\epsilon^3)$ distortion for any such S.

Observe that the diameter of G, $diam(G)$, and hence the largest distance in d^S, is $O(\log n)$. Applying Lemma 3.11 to $G[S]$, with $p = \frac{\epsilon \log n}{b}$, we conclude that its shortest path metric embeds into ℓ_1 with distortion $O(1/\epsilon)$. However, the shortest path metric of $G[S]$ is not d^S, and a finer argument is needed.

Consider a complete graph K on a vertex set S such that the weight of the edge (i,j) is $d(i,j)$. Clearly, the shortest path metric of K is d^S. We use the following basic result about graph spanners [2]. Let $\alpha < 1$ be any positive constant and let $H = (X, F)$ be a positively weighted graph. The result in [2] asserts the existence of a weighted subgraph $H_1 = (X, F_1)$ of H, with $|F_1| = |X|^{1+\alpha}$, such that the shortest path metric of H_1 distorts that of H by at most $O(1/\alpha)$. Moreover, the shortest path of H_1 dominates that of H, and the weights on F_1 are the same as in H. Applying this result for H being K, with $\alpha = \epsilon$, we conclude that there exists a subgraph K_1 with at most $|S|^{1+\epsilon} \le n^{1-\epsilon-2\epsilon^2}$ edges whose shortest path metric distorts d^S by at most $1/\epsilon$. Next we embed K_1 into G in the following manner. Replace each edge (u,v) in K_1 with the shortest path between (u,v) in G. Define graph G_1 (a subgraph of G) as the union of these paths. Clearly, the restriction of the shortest path metric of G_1 on S dominates d^S, while it is dominated by d_{K_1}. Hence, d_{G_1} distorts d^S by $O(1/\epsilon)$. Since G is unit-weighted, the size of $V(G_1)$ is at most $|E(K_1)| \cdot diam(G) \ \le\ n^{1-\epsilon}$.

The diameter of G_1 is at most $O(1/\epsilon)$ times the diameter of d^S. Also, being a subgraph of G of size less than $n^{1-\epsilon}$, property 2 of Lemma 3.10 implies that the graph G_1 satisfies the local density condition of Lemma 3.11 with $p = \frac{\epsilon \log n}{b}$. Therefore, by Lemma 3.11, the shortest path metric of G_1, d_{G_1}, is embeddable into ℓ_1 with distortion $O(1/\epsilon^2)$. Consider the restriction of d_{G_1} to S, and keeping in mind that it distorts d^S by at most $1/\epsilon$, we conclude that d^S is embeddable into ℓ_1 with $O(1/\epsilon^3)$.

For the second claim of the theorem, take $d_2 = \sqrt{d}$, where d is as above. Keeping in mind that the square root of an ℓ_1-metric is an ℓ_2-metric (since ℓ_1 is a subclass of NEG, the class of all metrics whose square root is an ℓ_2-metric), the statement follows. ∎

Proof of Lemma 3.10. The method is routine and similar to that of [3]. Observe first that the bound on diameter is implied by the third property and the bound on the degree. Hence it suffices to show a construction that satisfies the last two properties. Choose a random graph randomly from the Erdös–Rényi distribution on random graphs, $\mathcal{G}(n, c/n)$. Namely, choose one with independent and identically distributed (i.i.d.) edge probabilities c/n for some constant c (e.g., $c = 200000$). It is quite standard to show that such a graph will be a good expander, and, in particular, will have logarithmic diameter. It will also be locally very sparse, as needed. However, it might have high degree vertices. To correct this we delete all vertices of high degree. While this ensures the desired bound on the degree, the expansion may deteriorate and must be taken care of. We present here the complete argument, delaying the routine technical calculations to the appendix.

For a graph $G = (V, E)$ and subsets of vertices $X, Y \subseteq V$, denote $e_G(X, Y) = |\{(u, v) \in E|\ u \in X,\ v \in Y\}|$.

Let $G \in \mathcal{G}(n, c/n)$ be a random graph. Let $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{C}', \mathcal{D}$ be the following events:

- $\mathcal{A}$ the event that G is edge expanding on large sets; namely, for every subset $S \subseteq V(G)$ with $n/4 \le |S| \le n/2$, $e_G(S, \bar{S}) \ge c|S|/8$.
- $\mathcal{B}$ the event that for every subset $S \subseteq V(G)$ with $|S| \ge 99n/100$, $e_G(S, S) \ge 0.99 \cdot \frac{c}{n} \cdot \binom{|S|}{2}$.
- $\mathcal{C}$ the event that G is extremely locally sparse; namely, for every subset $S \subseteq V(G)$ with $\frac{\epsilon \log n}{b} \le |S| \le n^{1-\epsilon}$, $e(S, S) \le (|S| - 1) \cdot (1 + \frac{b}{\epsilon \log n})$.
- $\mathcal{C}'$ the event that there are only few small nonsparse sets; namely, the number of sets, $S \subset V$, for which $|S| < \frac{\epsilon \log n}{b}$, and $e(S, S) > (|S| - 1) \cdot (1 + \frac{b}{\epsilon \log n})$, is $o(n/\log^2 n)$.
- $\mathcal{D}$ the event that G has at most $n/100$ vertices of degree larger than $100c$.

Claim 3.12. *For $c = 200000$, $b = 30$, and G chosen as above, events $\mathcal{A}, B, C, C', D$ hold simultaneously with probability* 0.99.

The proof of Claim 3.12 is in the appendix.

Assume that the events $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{C}', \mathcal{D}$ hold for G and $c = 200000$. We now show how to deterministically construct our final graph G^* from G. This, together with Claim 3.12, implies that such a G^* exists.

First we want to ensure that every small subset $S \subset V$ induces a sparse graph as needed. This is quite easy: by event C', there are only $o(n/\log^2 n)$ sets, $S \subset V$, for which $e(S, S) > (|S| - 1) \cdot (1 + \frac{b}{\epsilon \log n})$. Moreover, each of these sets is of size bounded by $\epsilon \log n/b$. For every such S, we remove all edges in $G[S]$, namely at most $\binom{|S|}{2} = o(\log^2 n)$ edges. This results in a graph G', for which the sparsity condition (item 2 in Lemma 3.10) holds. Moreover, G' is obtained from G by removing $o(n)$ edges.

Let L be the set of vertices of degree at most $100c$. Since event $\mathcal{D}$ holds, $|L| \ge$

$\frac{99}{100}\cdot n$. Since event $\mathcal{B}$ holds, the set L satisfies $e_G(L,L) \geq 0.99\cdot\frac{c}{n}\cdot\binom{|L|}{2} \geq 0.99^3\cdot cn/2 \geq 0.97cn/2$. Since $|E(G)|$ is expected to be at least $c(n-1)/2$, we conclude (by any Chernoff like tail bound) that almost surely it is bounded by $1.01cn/2$. In this case, by deleting $\bar{L}$ from G' we have deleted additional $e_1 \leq (1.01-0.97)cn/2 = 0.02cn$ edges. Thus, altogether, $G_1 := G'[L]$ has all edges of G except for $0.02cn - o(n)$ edges. Recall that G_1 has $|V(G_1)| \geq \frac{99}{100}n$ vertices.

It remains to take care of expansion. Recall that the event $\mathcal{A}$ ensures that originally G had a good edge expansion for large sets. Thus, after deleting at most $0.02cn - o(n)$ edges, every subset $S \subseteq V(G_1)$ with $0.25n \leq |S| \leq 0.5n$ has $e_{G_1}(S,\bar{S}) \geq |S|\cdot c/8 - 0.02cn - o(n) \geq c|S|/100 \geq 2000|S|$.

We further delete now, one by one, (small) subsets that are not expanding enough. Namely, assume that there is a subset $A_1 \subseteq V(G_1)$ such that $e_{G_1}(A_1,\bar{A}_1) < |A_1|$, and $|A_1| \leq |\bar{A}_1|$. We let $G_2 = G_1[V(G_1)\setminus A_1]$. We now define similarly $G_3,\ldots$, and, in general, G_{i+1}, by deleting an arbitrary $A_i \subseteq V(G_i)$ for which $e_{G_i}(A_i,\bar{A}_i) < |A_i|$, and $|A_i| \leq |\bar{A}_i|$. Let $A_1, A_2, \ldots$ be the sequence of the deleted subsets, and let $G_1, G_2 = G_1 - A_1, \ldots$ be the sequence of the resulting graphs. Let $e_i = e_{G_i}(A_i,\bar{A}_i)$ be the size of the cut defined by A_i w.r.t. G_i. Then, by definition $e_i < |A_i|$.

Set $X_i = \cup_1^i A_i$ and note that $e_{G_1}(X_i,\bar{X}_i) < |X_i|$. This is since $e_j \geq e_{G_1}(A_j,\bar{X}_i)$ for every $j \leq i$. Hence, by assumption on G_1, for every i, $|X_i|$ is either smaller than $n/4$ or larger than $n/2$. Assume that for some i, $|X_i| > n/2$, and let i be the smallest index with this property. Thus $|X_{i-1}| < n/4$. It follows that $i \geq 2$, $|A_i| > n/4$ but less than $n/2$ (by definition, since $|A_i| \leq |\bar{A}_i|$), and $e_i < |A_i|$. However, $e_i \geq e_{G_1}(A_i,\bar{A}_i) - \sum_{j<i} e_j \geq 2000|A_i| - |X_{i-1}| \geq 2000\cdot|A_i| - \frac{n}{4} \geq 1999|A_i|$, contradicting the lower bound on e_i.

We conclude that for every i, $|X_i| < n/4$, and since the sequence must be finite, the last graph obtained $G^* = G_\ell$ is on $n^* \geq |V(G_1)| - n/4 \geq \frac{74n}{100}$, and for which every subset S of size $|S| \leq n^*/2$ has edge expansion $e_{G^*}(S,\bar{S}) \geq |S|$. ∎

We now address Lemma 3.11. It will follow from the following two claims.

For $t \in R^+$ and a metric μ, the t-truncated metric $\mu^{(t)}$ is defined by $\mu^{(t)}(x,y) = \min\{\mu(x,y),t\}$.

Claim 3.13. *Let μ be a tree metric and let $t \geq 0$ be a real number; then $\mu^{(t)}$ can be embedded into ℓ_1 with constant distortion.*[7]

Proof. Let T be a (possibly weighted) tree and μ its corresponding shortest path metrics. Let $t \geq 0$. We construct a weighted graph T_t by adding to T a new vertex u that is connected to every vertex of T with an edge of length $t/2$. Observe that the shortest path metric of T_t restricted to $V(T)$ is precisely $\mu^{(t)}$, the t-truncation of μ, and that T_t is 2-outerplanar. It is shown in [11] that the shortest path metric of any k-outerplanar graph can be embedded into ℓ_1 (in fact, into a distribution over dominating tree-metrics), with $f(k)$-distortion. This implies that $\mu^{(t)}$ can be embedded into ℓ_1 with constant distortion. ∎

Claim 3.14. *Let H, p be as in the statement of Lemma 3.11. Then, there is a probability distribution on spanning trees of H, such that each edge of H occurs with probability at least $p/(p+1)$.*

Proof. Let $H = (V,F)$ as in Claim 3.14.

Let $\mathcal{T}$ be the set of all spanning trees of H. Consider the following LP:

$$\max \lambda, \tag{8}$$

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

$$\text{s.t.} \quad \forall e \in F, \quad \sum_{T \in \mathcal{T}, e \in T} x_T \geq \lambda, \tag{9}$$

$$\sum_{T \in \mathcal{T}} x_T \leq 1, \tag{10}$$

$$\lambda, x_T \geq 0 \quad \forall T \in \mathcal{T}. \tag{11}$$

Any optimal solution $(x_T,\ T \in \mathcal{T})$ can be viewed as a probability distribution in which each edge is covered with probability at least λ. Thus we want to prove that the optimum is $\lambda \geq \alpha = p/(p+1)$.

Let the dual program be

$$\min z, \tag{12}$$

$$\text{s.t.} \quad \forall T \in \mathcal{T}, \quad \sum_{e \in T} y_e \leq z, \tag{13}$$

$$\sum_{e \in F} y_e \geq 1, \tag{14}$$

$$z, y_e, \geq 0 \quad \forall e \in F. \tag{15}$$

Thus by duality we need only show that the optimal z in the dual is at least α. Let $y_e,\ e \in F$, be taken as weights on F. Then, (13) implies that z is the maximum cost spanning tree with respect to the weights y_e. Thus, we need only show that under the assumption of Claim 3.14, for any nonnegative weights on F that sum to 1, the maximum cost spanning tree is at least α.

Indeed, let $y_e,\ e \in F$ be any such weighting, and suppose that the maximum spanning tree w.r.t. the weights y_e is achieved by a tree $T = (V, E)$. Let the girth of H be g. By the local density condition of Lemma 3.11, it must hold that $(g-1)(1+1/p) \geq g$, implying $g > p$.

Consider now the following bipartite graph. The left-hand side of the bipartition has a point corresponding to each edge of T. The right-hand 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 $y_e \geq y_f$. Also, by the girth lower bound above, it follows that the degree of every vertex in the right-hand side is at least p.

We claim that this bipartite graph has a p-matching of the vertices in the right-hand side, namely, a subgraph with degree at most 1 for points on the left and degree exactly p for points on the right. Suppose not; then by the Hall Marriage Theorem there is some minimal subset X on the right-hand 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 the assumptions of Lemma 3.11.

The existence of the p-matching implies that the edges of T that are matched (that is, of degree 1 in the p-matching) 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,

$$cost(T) = \sum_{e \in E} y_e \geq p \cdot \sum_{e \notin E} y_e.$$

Recalling that $\sum_{e \in F} y_e = 1$, this implies that the cost of T is at least $\alpha = p/(p+1)$ as claimed. ∎

We now return to the proof of Lemma 3.11.

Proof of Lemma 3.11. By Claim 3.14, there is a probability distribution on spanning trees $\{T_i\}$ of H such that each edge of H occurs with probability at least $\alpha = p/(p+1)$. Recall that D is the diameter of H. For each T_i in the distribution, consider the corresponding metric $\mu_i = \min\{D, d_{T_i}\}$, namely $\mu_i = d_{T_i}^{(D)}$, the D-truncated tree metric. Set $\mu = \sum w_i \mu_i$, where w_i is the weight of T_i in the distribution. Since for every i, μ_i dominates d, it follows that μ dominates d. Hence, to upper bound $\mathrm{dist}(\mu, d)$, it is enough to bound the stretch of each edge of H.

Indeed, for any edge, its μ-length is at most

$$1 \cdot \frac{p}{p+1} + D \cdot \left(1 - \frac{p}{p+1}\right) = \frac{p+D}{1+p}.$$

Finally, by Claim 3.13, every μ_i, and hence μ, can be embedded into ℓ_1 with constant distortion. ∎

4. Separating a baseline metric class from ℓ_∞. We say that a class of metric spaces $\mathcal{C}$ is *universal* if any metric space is embeddable into it with some constant distortion, $c = c_{\mathcal{C}}$, that depends only on $\mathcal{C}$. It is known (and simple) that the normed space ℓ_∞ is universal. Moreover, ℓ_∞^{n-1} is universal for finite metrics on n points (see, e.g., [12]). Namely, every metric on n points is embeddable isometrically into ℓ_∞^{n-1}.

As noted in section 3, a baseline class of metrics includes all ℓ_1 metrics, and should be thought of as "large," or potentially allowing small distortion as a host space. Thus it is natural to study their limitations, namely, how universal can a baseline class of metric $\mathcal{C}$ be if $\mathcal{C}$ does not contain the set of all metrics.

In the preliminary conference version of this work, we have conjectured that the following strong separation holds.

Conjecture 1. *Let $\mathcal{C}$ be nonuniversal baseline class of metrics. Then, for any $n \in N$, there exists an n-point metric d_n such that* $\mathrm{dist}(d_n \hookrightarrow \mathcal{C}) \geq \Omega(\log^\alpha n)$ *for some constant $\alpha > 0$.*

This conjecture was subsequently proved in [27] even under weaker assumptions: it is enough to demand that $\mathcal{C}$ be a nonuniversal class of metrics closed undertaking submetrics.

Here, we examine the separation question between $\mathcal{C}$ and all universal metrics from another perspective. Unlike in the rest of the paper, we shall discuss here metrics whose underlying space is infinite: the entire R^n or Z^n, and, in particular, $\{\ell_1^n\}$ and $\{\ell_\infty^n\}$. The main result of this section is a separation between $\mathcal{C}$ and ℓ_∞ in terms of the *dimension* of the underlying host space.

Theorem 4.1. *Let $\mathcal{C}$ be a baseline 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*

R^n,

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

Observe that the upper bound on the worst distortion cannot exceed $\sqrt{n}$. Indeed, notice that any ℓ_2 metric on n points is in $\mathcal{C}$, since $\mathcal{C}$ contains ℓ_1 metrics (see the remark right after Definition 2.1), and any ℓ_2 metric on n points is isometrically embeddable into ℓ_1 [28]. On the other hand, the distortion between (R^n, ℓ_∞^n) and (R^n, ℓ_2^n) is $O(\sqrt{n})$ as shown by the identity mapping.

The proof of Theorem 4.1 uses the following lemma.

LEMMA 4.2. *For any $d \in \mathcal{C}$ on R^n, there exists a* norm $\| * \| \in \mathcal{C}$ *on R^n such that*

$$\mathrm{dist}(\ell_\infty^n, \| * \|) \;\le\; \mathrm{dist}(\ell_\infty^n, d). \tag{16}$$

The proof of this lemma is delayed to the end of this section.

Next, we need the following quantitative version of a theorem by James [20], communicated to us, together with an outline of its proof, by W.B. Johnson and G. Schechtman.

THEOREM 4.3. *Let $\gamma = (1+\delta)^{2^r}$ and let $n \ge k^{2^r}$, where $r, k \in N$, and $0 \le \delta < 1$. Assume that an n-dimensional norm $\| * \|$ is embeddable into ℓ_∞^n with distortion at most γ. Then there exists a subspace L of R^n of dimension* $\dim(L) \;=\; k$, *with an ℓ_∞ norm d_∞ on L, such that for the restriction of $\|*\|$ to L, $\|*\|_L$,* $\mathrm{dist}(\|*\|_L, d_\infty) \le \frac{1+\delta}{1-\delta}$.

Proof. The following lemma from [28, pp. 74–75] states the following.

LEMMA 4.4 (see [28]). *Under the conditions of the above theorem, there exists a subspace L of dimension k such that for every $\{v_i\}_{i=1}^k \subset L$ the following holds. Denote by $\| * \|_L$ the restriction of $\| * \|$ to L. Then, for any $v = \sum_i \alpha_i v_i$, $\| \sum_i \alpha_i v_i \|_L \;\le\; (1+\delta) \cdot \max_i \|\alpha_i v_i\|_L$.*

Observe that it also implies that for any $v = \sum_i \alpha_i v_i$, $\| \sum_i \alpha_i v_i \|_L \;\ge\; (1-\delta) \cdot \max_i |\alpha_i| \cdot \|v_i\|_L$. Indeed, were it not the case for some $v = \sum_i \alpha_i v_i$, one would get a contradiction as follows: Assume, w.l.o.g., that $|\alpha_1| \cdot \|v_1\|_L$ is the maximal term, and, moreover, it is 1. Let $u = \alpha_1 v_1 \;-\; \sum_{i=2}^k \alpha_i v_i$. Then,

$$1 \;=\; \|\alpha_1 v_1\|_L \;=\; \|0.5v + 0.5u\|_L \;<\; 0.5(1-\delta) + 0.5(1+\delta) \;=\; 1\,.$$

To conclude the argument, choose a basis for L, $\{v_i\}_{i=1}^k$, where $\|v_i\|_L = 1$ for all i's, and consider the mapping ϕ from L equipped to R^n equipped with ℓ_∞ norm, where each v_i is mapped to the corresponding unit vector e_i. In view of Lemma 4.4, the contraction of ϕ is at most $(1+\delta)$, and in view of the consequence of this lemma, the stretch of ϕ is at most $\frac{1}{1-\delta}$. The desired bound on the distortion follows. ∎

Proof of Theorem 4.1. Assume for simplicity that n is of the form $n = k^{2^r}$. The metric $\mu_k \notin \mathcal{C}$, being a metric on k points, isometrically embeds into ℓ_∞^k. We conclude by Theorem 4.3 that for any $\mathcal{C}$-norm $\| * \|$ on R^n it holds that

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

The same estimate holds, by Lemma 4.2, for any metric $d \in \mathcal{C}$ on R^n. Thus, for such n, the theorem holds with constant $\alpha = \log_k(1 + \frac{\beta-1}{\beta+1})$. If n is not of the form k^{2^r}, take the largest such power $\le n$, at the cost of paying an extra factor $1/2$ in the above α. This concludes the proof of Theorem 4.1. ∎

Proof of Lemma 4.2. In what follows we restrict our attention, w.l.o.g., to d's dominating ℓ_∞^n. Hence, we may replace distortion with the (supremum) stretch incurred by d. It will be convenient to bring the discussion back to the realm of discrete metric spaces. Instead of proving (16) for R^n, we shall prove it for Z^n. Clearly, this is a fully equivalent statement (by scaling and taking limits).

Observe that a norm on Z^n is just a translation-invariant scalable metric.

First, we construct a translation-invariant metric $d_* \in \mathcal{C}$ on 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 $p \in Z^n$, define a metric d^{+p} on 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 Z^n equipped with the ℓ_∞^n metric and has the same stretch as d.

For an integer i let $[-i,i] = \{-i,\ldots,i\}$. Let $[-i,i]^n \subseteq Z^n$ denote the corresponding discrete cube. For a point $x \in Z^n$ let $[-i,i]^n - x = \{y - x \mid y \in [-i,i]^n\}$ denote the shifted cube. 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 belongs to $\mathcal{C}$, 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 $x, y \in Z^n$ we have

$$\begin{aligned}
&\lim_{i\to\infty} |d_i(x,y) - d_i(0, y-x)| \\
&= \lim_{i\to\infty} \left| \frac{1}{(2i+1)^n} \sum_{p\in[-i,i]^n} d(x+p, y+p) - \frac{1}{(2i+1)^n} \sum_{p\in[-i,i]^n} d(p, y-x+p) \right| \\
&\le \lim_{i\to\infty} \frac{1}{(2i+1)^n} \sum_{p\in\, [-i,i]^n \,\triangle\, ([-i,i]^n - x)} d(x+p, y+p) \\
&\le \lim_{i\to\infty} \frac{1}{(2i+1)^n} \cdot 2n \cdot \|x\|_\infty \cdot (2i+1)^{n-1} \;\cdot \mathrm{dist}(\ell_\infty^n, d) \|x-y\|_\infty \;=\; 0. \qquad (17)
\end{aligned}$$

Next, we employ the following standard procedure. Order all vectors of 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 subsubsequence, and continue in the same manner ad infinitum. Finally, for each $x, y \in Z^n$, 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 nonincreasing with r, since for any $a \in N^+$, and for $a = 2$ in particular, $d_*(ax, ay) \leq ad_*(x, y)$ 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. For example, 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 $a \in N^+$, and hence with respect to all $a \in Q^+$, as required. ∎

5. Ultrametrics. The set of ultrametrics is the set of metrics $\text{ULT} = \{d : \ d(p, q) \leq \max\{d(p, r), d(q, r)\} \ \forall p, q, r\}$. Ultrametrics form an important class of simple metrics studied previously in algorithmic contexts, e.g., in [15, 14, 7]. It is well known that ultrametrics embeded isometrically into ℓ_2 are a special case of tree metrics, and they are not closed under addition (see, e.g., [12]). Our aim here is to study embeddability into ultrametrics from the local-global view. Namely, we consider the set of ultrametrics as the host space and see what local embeddability can imply for global embeddability. Note that ULT is not baseline, as it is not a cone (recall the definition of baseline in section 2). Hence, the results from the previous section do not apply to this class.

For a metric d and two points x, y, an xy-path P is a sequence of distinct points $(x = p_0, p_1, p_2, \ldots, p_m = y)$ of arbitrary length. We say that $pq \in P$ 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 set

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

The following basic result characterizes the ultrametric closest to a given metric d.

THEOREM 5.1 (see [15]). *Let d be a metric; then the distance function u is an ultrametric which is dominated by d (i.e., $u(x, y) \leq d(x, y)$ 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 5.2. *Let d be a metric and let $c \leq 1$ be the maximum value such that for every $x, y \in X$, every xy-path P contains $pq \in P$ such that $d(p, q) \geq c \cdot d(x, y)$. Then,* $\text{dist}(d \hookrightarrow \text{ULT}) = c^{-1}$.

Using this criterion we establish the following theorem.

THEOREM 5.3. *Fix an integer m, and assume that (X, d) is a metric on $n \geq m$ points such that for every m-subset Q of X,* $\text{dist}(d^Q \hookrightarrow \text{ULT}) \leq \gamma$. *If $n \leq (m-1)^k + 1$, then* $\text{dist}(d \hookrightarrow \text{ULT}) \leq \gamma^k$.

Proof. The proof is by induction on k. For $k = 1$ the claim is obvious. Let $n \leq (m-1)^k + 1$, $k \geq 2$. To bound the distortion using Corollary 5.2, it suffices to show that for every $x, y \in X$, every xy-path P contains $pq \in P$ with $d(p, q) \geq d(x, y)/\gamma^k$. Let $P = (x = v_1, v_2, \ldots, v_r = y)$, $r \leq n$, be such a path. Consider the xy-path $P' = (v_1, v_m, v_{2m-1}, \ldots v_{im-(i-1)}, \ldots, v_r)$. By our assumption on n, P' contains at most $(m-1)^{k-1} + 1$ points. Moreover, as a submetric of X it satisfies the premises of the theorem with $k-1$. Hence, by induction, there exists an i such that $v_{im-(i-1)}v_{(i+1)m-i} \in P'$ and such that $d(v_{im-(i-1)}, v_{(i+1)m-i}) \geq \frac{d(x,y)}{\gamma^{k-1}}$. Now consider the segment P'' of P from $v_{im-(i-1)}$ to $v_{(i+1)m-i}$, which is a $v_{im-(i-1)}v_{(i+1)m-i}$-path containing at most m points. By the base case of the induction, there exists $pq \in P''$ such that $d(p, q) \geq d(v_{im-(i-1)}, v_{(i+1)m-i})/\gamma \geq d(x, y)/\gamma^k$. ∎

The next theorem implies that Theorem 5.3 is asymptotically best possible.

In the following let d be the metric induced by the unit weighted path on $n+1$ vertices.

THEOREM 5.4. *For any constants* $\epsilon, c \in [0,1]$, $\mathrm{dist}(d^c \hookrightarrow \mathrm{ULT}) = \Omega(n^c)$, *whereas for every subset of* n^ϵ *points,* S, *the restriction* d^c_S *of* d^c *to* S *has* $\mathrm{dist}(d^c_S \hookrightarrow \mathrm{ULT}) = O(n^{\epsilon c})$.

Proof. Corollary 5.2 implies that $\mathrm{dist}(d^c \hookrightarrow \mathrm{ULT}) = \Omega(n^c)$, as for the two endpoints, x, y, of the path $d^c(x,y) = n^c$, while for the path itself, as a xy-path, for every two adjacent vertices $d^c(p,q) = 1$.

On the other hand, for any set S of points of size n^ϵ and any $u, v \in S$, the monotone path going from u to v contains at most n^ϵ points. Hence there exists an adjacent pair p, q on the path whose length is at least $d(p,q) \geq d(u,v)/n^\epsilon$. This implies that $d^c(p,q) \geq n^{-\epsilon c} d^c(u,v)$, which by Corollary 5.2 implies the upper bound on the distortion. ∎

Remark 5.1. The same result is essentially true for d' being the metric of the unit cycle of length n instead of the metric d of the unit path. We note, however, that the metrics $(d')^c$ are $\Omega(n^c)$ far from the more general set of tree metrics (by the argument from [30, Corollary 5.3]). Hence, the lower bounds hold for tree metrics as well.

6. Concluding remarks. As already explained in the introduction, the results of this paper strengthened by the subsequent paper of Charikar, Makarychev, and Markarychev [8] make it unlikely that the k-local restrictions may help in dealing with the sparsest cut problem. Yet, the structural local-global results proved here seem to be of an independent value for the theory of finite metric spaces. The most interesting open problem in this direction is tightening the result of Theorem 3.9 about metrics with (almost) Euclidean local structure.

Approximating general (or special) metrics by metrics from some nontrivial baseline class $\mathcal{C}$ may have interesting structural and algorithmic applications. For example, do planar metric embed into $\mathcal{M}_5$ with a constant distortion? This is closely related to the famous question about ℓ_1-embeddability of planar metrics (see, e.g., [18]). Gupta [17] has shown in his Ph.D. thesis that planar metrics embed with constant distortion into NEG (see also [23]). Hence the same holds for any $\mathcal{M}_{2k}$.

We also find it intriguing to understand the structure of metrics of maximum distortion w.r.t., e.g., the class $\mathcal{M}_5$. Similar question w.r.t. the classes ℓ_1 and NEG lead to the extremely important notions of edge expansion and spectral gap. The results of Mendel and Naor [27], proving the conjecture that is discussed in section 4, show that for any nontrivial baseline class of metric, there is a metric whose distortion when embedding into this class is $\Omega(\log^\alpha n)$ for some positive $\alpha \leq 1$. The value of the right α remains open. Is it strictly smaller than 1 for any baseline class of metrics? The methods of this paper imply that the shortest path metrics of constant degree expanders (extremal both for ℓ_1 and NEG) can yield only $\alpha \geq \log_2 1.333$.

Although Theorem 3.9 was improved in [8], we think that the methods we developed to establish the upper bounds in this theorem (Claim 3.13 and Lemma 3.11) are of independent interest. In particular, Lemma 3.11 is a generalization of an older result, [11], on embedding metrics that are the shortest path metrics of sparse graphs into ℓ_1. It is open to interpretation how much further such generalizations can go. In particular, we do not have an example that attests to whether the parameters used in Lemma 3.11 are best possible.

Finally, the findings of this paper and of [8] 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 in

various fields.

Appendix. We prove here the claims needed for Lemma 3.10. We do not try to optimize the constants; hence we make the crudest computations.

Proof (that events $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{C}'$, and $\mathcal{D}$ in the proof of Lemma 3.10 hold with high probability). In the following we set $c = 200000$. For any unordered pair $e = (u, v)$, let X_e be the indicator function of e, namely 1, if $e \in G$ and 0 otherwise. We use the following Chernoff bound (see, e.g., [1, Theorem A.1.13]): let $\{X_i\}_1^m$ be boolean independent random variables, each with $\text{Prob}(X_i = 1) = p$; then $\text{Prob}(\sum_1^m X_i \leq pm - a) \leq e^{-a^2/2pm}$.

Let $S \subset V$ with $|S| = \alpha \cdot n$, $0.25 \leq \alpha \leq 0.5$. Then, using the bounds on α, $\mu := \mathit{Expect}(e(S, \bar{S})) = \sum_{e=(u,v), u \in S, v \notin S} \mathit{Expect}[X_e] = c \cdot \alpha \cdot (1 - \alpha)n \geq \alpha cn/2$. Hence, $\text{Prob}(e(S, \bar{S}) \leq \alpha cn/8) \leq \text{Prob}(\mu - e(S, \bar{S}) \geq 3\alpha cn/8) \leq e^{-9cn/256}$. For $c > 200000$ this is $o(2^{-n})$. Since there are at most 2^n such subsets, the union bound implies that with probability $1 - o(1)$ for all subsets $S \subseteq V(G)$, with $n/4 \leq |S| \leq n/2$, $e(S, S) > |S|/8$. This proves the claim for event $\mathcal{A}$.

Similarly for event $\mathcal{B}$, let $S \subseteq V$ with $|S| \geq 99n/100$. Then $\mu_B := \mathit{Expect}[e(S, S)] = \binom{|S|}{2} \cdot \frac{c}{n} \geq 98cn/200$. By the same Chernoff bound above, $\text{Prob}(e(S, S) < 0.99\mu_B) < e^{-10^{-4}\mu_B/2} \leq e^{-10^{-6}49cn} = o(2^{-n})$ for our choice of c. Again, as there are at most 2^n such subsets, with probability $1 - o(1)$ event $\mathcal{B}$ holds.

To prove the bound on the local density we follow essentially the same computation as in [3, Lemma 3]. Let $\beta = \frac{b}{\epsilon \log n}$ and let $1/\beta \leq \ell \leq n^{1-\epsilon}$. The probability that there exists a subset of size ℓ that spans more than $(1 + \beta)(|S| - 1)$ edges is given by

$$\begin{aligned} p_\ell = \text{Prob}(\exists S, |S| = \ell, e(S, S) \geq (1 + \beta)(\ell - 1) + 1) \\ &\leq \binom{n}{\ell} \cdot \binom{\binom{\ell}{2}}{(1 + \beta)(\ell - 1) + 1} \cdot \left(\frac{c}{n}\right)^{(1+\beta)(\ell-1)+1} \\ &\leq \left(\frac{ne}{\ell}\right)^{\ell} \cdot \left(\frac{\ell e}{2}\right)^{\ell+\beta\ell-\beta} \cdot \left(\frac{c}{n}\right)^{\ell+\beta\ell-\beta} \\ &\leq \left[\left(\frac{\ell}{n}\right)^{\beta} \cdot K\right]^{\ell} \cdot \left(\frac{n}{\ell}\right)\beta \end{aligned}$$

for a constant $K = e^{(2+\beta)} \cdot c^{(1+\beta)}$.

Hence for $\beta = \Omega(1/\epsilon \log n)$, $p_\ell < 2^{-\ell}$ (one can check that for $\beta = 30/(\epsilon \log n)$, namely for $b = 30$, this is sufficient).

This implies, using the union bound, that with high probability, for every subset $S \subset V$ with $\frac{\epsilon \log n}{b} \leq |S| \leq n^{1-\epsilon}$, S spans at most $(|S| - 1)(1 + \beta)$ edges as needed.

To prove that $\mathcal{C}'$ holds with high probability, note that for a subset $S \subset V$, for which $|S| < 1/\beta$, $(|S| - 1) < (|S| - 1) \cdot (1 + \beta) \leq |S|$. Hence, the condition that $e(S, S) > (|S| - 1)(1 + \beta)$ becomes $e(S, S) \geq |S|$. Thus the expected number of such subsets μ is given by the expression

$$\mu \leq \sum_{\ell=3}^{1/\beta} \binom{n}{\ell} \cdot \binom{\binom{\ell}{2}}{\ell} \cdot \left(\frac{c}{n}\right)^{\ell} \leq \sum_{\ell}^{1/\beta} (ec)^{\ell}.$$

For $\beta = 30/\epsilon \log n$ this is $o(n/\log^2 n)$; hence, Markov's inequality implies that $\mathcal{C}'$ holds with high probability.

Finally, the fact that $\mathcal{D}$ holds with probability at least 0.99 directly follows from Markov's inequality. This is since the probability for each vertex v separately, that $deg_v > 100c$ is at most $\binom{n}{100c} \cdot (\frac{c}{n})^{100c}$, is extremely small. $\square$

Acknowledgments. We are deeply grateful to G. Schechtman and W.B. Johnson for their explanation of Theorem 4.3. We also thank J. Matousek for pointing out important references. Finally we thank the referees, who made a significant impact on the preliminary version of this paper and helped to considerably improve it.

REFERENCES

[1] N. Alon and J. H. Spenser, *The Probabilistic Method*, 3rd ed., John Wiley and Sons, Hoboken, NJ, 2008.

[2] I. Althoefer, G. Das, D. Dobkin, D. Joseph, and J. Soares, *On generating sparse spanners for weighted graphs*, Discrete Comput. Geom., 9 (1993), pp. 81–100.

[3] S. Arora, B. Bollobás, L. Lovász, and I. Tourlakis, *Proving integrality gaps without knowing the linear program*, Theory Comput., 2 (2006), pp. 19–51.

[4] S. Arora, J. R. Lee, and A. Naor, *Euclidean distortion and sparsest cut*, J. Amer. Math Soc., 21 (2008), pp. 1–21.

[5] S. Arora, S. Rao, and U. V. Vazirani, *Expander flows, geometric embeddings and graph partitioning*, J. ACM, 56 (2009), article 5.

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

[7] Y. Bartal, N. Linial, M. Mendel, and A. Naor, *On metric Ramsey-type phenomena*, Ann. of Math. (2), 162 (2005), pp. 643–709.

[8] M. Charikar, K. Makarychev, and Y. Makarychev, *Local global tradeoffs in metric embeddings*, SIAM J. Comput., 39 (2010), pp. 2487–2512.

[9] S. Chawla, A. Gupta, and H. Räcke, *Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut*, ACM Trans. Algorithms, 4 (2008), article 22.

[10] S. Chawla, R. Kumar, R. Krauthgamer, Y. Rabani, and D. Sivakumar, *On the hardness of approximating multicut and sparsest cut*, Comput. Complexity, 15 (2006), pp. 94–114.

[11] C. Chekuri, A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair, *Embedding k-outerplanar graphs into l_1*, SIAM J. Discrete Math., 20 (2006), pp. 119–136.

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

[13] M.M. Deza and H. Maehara, *Metric transform and Euclidean embeddings*, Trans. Amer. Math. Soc., 317 (1990), pp. 661–671.

[14] J. Fakcharoenphol, S. Rao, and K. Talwar, *A tight bound on approximating arbitrary metrics by tree metrics*, in Proceedings of the 35th Annual ACM Symposium on Theory of Computing, ACM, New York, 2004, pp. 60–70.

[15] M. Farach, S. Kannan, and T. Warnow, *A robust model for finding optimal evolutionary trees*, Algorithmica, 13 (1995), pp. 155–179.

[16] K. Georgiou, A. Magen, and I. Tourlakis, *Vertex cover resists SDPs tightened by local hypermetric inequalities*, in Integer Programming and Combinatorial Optimization, Lecture Notes in Comput. Sci. 5035, Springer, Berlin, 2008, pp. 140–153.

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

[18] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair, *Cuts, trees and ℓ_1-embeddings of graphs*, Combinatorica, 24 (2004), pp. 233–269.

[19] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc., 58 (1963), pp. 13–30.

[20] R.C. James, *Uniformly non-square Banach spaces*, Ann. of Math. (2), 80 (1964), pp. 542–550.

[21] W.B. Johnson and G. Schechtman, *private communication*, 2003.

[22] S. Khot, *On the power of unique 2-prover 1-round games*, in Proceedings of the 34th Annual ACM Symposium on Theory of Computing, ACM, New York, 2002, pp. 767–775.

[23] J. Lee, M. Mendel, and A. Naor, *Metric structures in L_1: Dimension, snowflakes, and average distortion*, European J. Combin., 26 (2005), pp. 1180–1190.

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

[25] J. Matousek, *Ramsey-like properties for bi-Lipschitz mappings of finite metric spaces*, Comment. Math. Univ. Carolin., 33 (1992), pp. 451–463.

[26] J. Matousek, *Open Problems on Embeddings of Finite Metric Spaces*, http://kam.mff.cuni.cz/∼matousek/metrop.ps.

[27] M. Mendel and A. Naor, *Metric cotype*, Ann. of Math. (2), 168 (2008), pp. 247–298.

[28] V. Milman and G. Schechtman, *Asymptotic Theory of Finite-Dimensional Spaces*, Lecture Notes in Math. 1200, Springer-Verlag, Berlin, 1986.

[29] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, Cambridge, UK, 1995.

[30] Y. Rabinovich and R. Raz, *Lower bounds on distortion of embedding finite metric spaces in graphs*, Discrete Comput. Geom., 19 (1998), pp. 79–94.

[31] I.J. Schoenberg, *Remarks to Maurice Frechet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert"*, Ann. of Math. (2), 36 (1935), pp. 724–732.