Metric decomposition, smooth measures, and clustering

James R. Lee
U.C. Berkeley
jrl@cs.berkeley.edu

Assaf Naor
Microsoft Research
anaor@microsoft.com

Abstract

In recent years, randomized decompositions have become a fundamental tool in the design of algorithms on metric spaces. A randomized decomposition is a distribution over partitions of a metric space satisfying certain properties. Two main types of decompositions have emerged: The "padded" decomposition, see e.g. [20, 15, 9, 19], and the "separating" decomposition, e.g. [2, 3, 5, 14]. Here, we show that if a metric space admits the former type of decomposition, then it also admits the latter. This result froms a non-trivial transformation of distributions. Using this technique, we give new approximation algorithms for the 0-extension problem when the input metric (T, d) on terminals admits such a decomposition. For instance, we achieve an $O(1)$ approximation when the metric on terminals is planar or a subset of $\mathbb{R}^d$ under any norm.

We then introduce two additional techniques. The first is based on a powerful theorem about the existence of "well-behaved" probability measures on metric spaces. The second yields improved randomized decompositions for n-point subsets of L_p, $1 < p < 2$, by utilizing probabilistic techniques of Marcus and Pisier based on p-stable random variables. Both yield improved approximation algorithms for important classes of metric spaces.

1 Introduction

Metric decomposition. In recent years, randomized decompositions have become a fundamental tool in the design of algorithms on metric spaces. A randomized decomposition is a distribution over partitions of a metric space satisfying certain properties. Consider a metric space (X, d) and let some value $\Delta > 0$ be given. A typical "low-diameter" randomized decomposition produces a partition P of X such that for every $C \in P$, $\text{diam}(C) \leq \Delta$ (the sets $C \in P$ are often called *clusters*). The decomposition usually satisfies some additional competing property which holds on average or with high probability (for clearly the first property is easily satisfied by trivial decompositions).

Two main types of decomposition have emerged. The first, for which the term "padded" was coined in [15], ensures that the distance from a point to the "boundary" of the partition is large in expectation (see, for example, [20, 15, 9, 19]). The second type, which we called "separating," ensures that the probability of separating two points $x, y \in X$ into different clusters of the partition is small (e.g., [2, 3, 5, 14, 7]).

Each decomposition has its strengths and weaknesses. The separating condition is often natural to use in applications (as in the tree decompositions of Bartal [3]). The padded decomposition, on the other hand, possesses certain properties that the separating decomposition cannot. For instance, in [9], a padded decomposition is produced which has *small support*, i.e. a distribution which is supported on only constantly many partitions (in terms of n, the number of points) and which is uniform over that set. It is not difficult to see that no such distribution can exist for separating decompositions.

In this paper, we show another important property of padded decompositions. Let (X, d) be a metric space and $T \subseteq X$ a subspace. In Section 2.2, we show that a padded decomposition for T can be *extended* to a padded decomposition of a neighborhood of T. For instance, consider a net N in the metric space X (one might think of a grid in $\mathbb{R}^3$, say). If the net is sufficiently fine, then we show that it is possible to extend a padded decomposition of the net to one of the whole space. Again, separating decompositions cannot possess this property (basically because behavior on a net does not yield enough precision to say something about separating points $x, y \in X$ which are very close together).

In Section 2, we show that *every space* (X, d) which admits a padded decomposition with certain parameters also admits a separating decomposition with (almost) the same parameters! It follows that the former notion is strictly stronger than the latter. This is proved by means of an efficient procedure which converts a distribution μ of one type to a distribution ν of the other in a non-trivial way. In particular, a partition $P \in \text{supp}(\nu)$ is produced by gluing together an unbounded (though small in expectation) number of partitions from μ in a highly coordinated manner. The usefulness of this technique is exhibited in Section 3, where it is used to obtain new approximation algorithms for the 0-extension problem.

The 0-extension problem. In the *0-extension problem*, one is given an undirected graph $G = (V, E)$ with a non-negative cost function $c : E \to \mathbb{R}_+$ on the edges, a subset $T \subseteq V$ called the *terminals*, and a metric d on T. The goal is output an assignment $f : V \to T$ of vertices to terminals so that $f|_T$ is the identity on T and which minimizes $\sum_{uv \in E} c(u, v) \cdot d(f(u), f(v))$. This is a generalization of the *multi-way cut problem* and a special case of *metric labelling* (see [4, 6] for a more comprehensive history of the problem). It is clearly a natural formulation of many clustering and classification problem.

In [4], an LP relaxation of this problem was used to obtain an $O(\log k)$ approximation where k is the number of terminals. In [6], this was improved to $O(\log k / \log \log k)$. Both of these papers achieve an approximation algorithm by rounding the optimal solution of the LP (called the *semi-metric relaxation*, which appears in Section 3) to an integer solution and thus also bound the integrality ratio by the same factors. Additionally, in [4], it is shown that if the *cost graph* is planar, then an $O(1)$ approximation can be achieved.

We obtain improved approximations (and corresponding bounds on the integrality ratio of the semi-metric relaxation) for special cases of the problem which occur when restrictions are placed on the *metric on terminals* (T, d). Restrictions on the structure of (T, d) are far more natural and likely to occur in practical situations. In particular, we show that if a metric space (T, d) admits a δ-padded stochastic decomposition (see Section 2 for the definition) at every scale, then there exists an $O(\frac{1}{\delta})$-approximation for the 0-extension problem when (T, d) is the metric on terminals.

As a result, we prove that when (T, d) is a planar metric, there is a constant factor approximation. In addition to such a topological restriction, we address the case when the "volume growth" of T is bounded. Let λ_T be the smallest value such that every ball in T can be covered by λ_T balls of half the radius. If $\lambda_T < \infty$, then T is called a *doubling metric*, and for such metrics we again achieve a constant factor approximation. More generally, we achieve an approximation ratio of $O(\log \lambda_T)$ which, even in extremal cases, matches the approximation guarantee of [4] (since clearly $\lambda_T \leq n$ always). It has been observed recently that doubling metrics occur naturally in many applications like peer-to-peer networks, computational biology, and machine learning (see, e.g. [16]), contexts in which clustering and classification problems seem highly relevant.

The main obstacle to achieving improved approximations for the 0-extension problem using randomized decompositions is that the semi-metric relaxation (see Section 3) produces a metric (V, ρ) with an arbitrarily complex structure, and thus, even if the metric on terminals is decomposable, it is unclear how one can use this to round the LP solution. If (T, d) admits a padded decomposition μ, we proceed by first *extending* μ to a neighborhood of T in (V, ρ) using the partition extension lemma (Lemma 2.9); the goal is to use μ to round those vertices which are close to T. Unfortunately, the padded decomposition is not suited to this purpose, and it is unclear how one might utilize it. Fortunately, we use the fact that μ can be converted to a *separating* decomposition of the neighborhood (using Theorem 2.2). The separating decomposition can then be used quite naturally. Finally, we repeat this process "on all scales," eventually rounding all of V.

Smooth measures. In the algorithm of [4] for the 0-extension problem, the randomized rounding proceeds by first choosing a permutation σ of the terminals uniformly at random. We alter their algorithm in a novel way by instead choosing a permutation from a different distribution, one which is tailored to the metric structure of T. To do this, we use a compelling tool from abstract measure theory which says that every compact doubling metric admits a comparable doubling measure [21]. We believe this is a powerful technique with many potential applications.

As a result, we obtain an $O(\log \lambda_T)$ approximation in a completely different way from that described above. Since their algorithm uses "uniform sampling," it performs badly in many cases when the terminal metric has some "hidden structure." For instance, consider complex subsets of $\mathbb{R}^3$ which, when considered in isolation, seem to lack the regular structure of Euclidean 3-space. Our technique first lays down the "proper" measure on such a point set so that the algorithm of [4] is, in some sense, able to "see all of the ambient space."

Improved decompositions for finite subsets of L_p, $1 < p < 2$. In [5], it is shown that, for $1 \leq p \leq 2$, subsets of ℓ_p^d (i.e. d-dimensional L_p space) admit separating decompositions with parameter $\frac{1}{2}d^{-1/p}$. It follows that any n-point subset $X \subseteq L_2$ admits a separating decomposition with parameter $\approx (\log n)^{-1/2}$, because the Johnson-Lindenstrauss flattening lemma [11] says that any n-point subset of Euclidean space can be embed into $d = O(\log n)$ dimensions with bounded distortion.

Unfortunately, such a result does not follow for $1 < p < 2$, since it is unknown whether dimension reduction exists in these cases (and is, in fact, a major open question). We bypass this obstacle using probabilistic techniques of Marcus and Pisier [18], namely their analysis of p-stable random processes. By first embedding L_p spaces into L_2 on average and then applying [11], we achieve separating decompositions with parameter $\approx (\log n)^{-1/p}$ for n-point subsets of L_p, $1 < p < 2$.

1.1 Preliminaries

We briefly recall some definitions about metric spaces (see, e.g. [10]). Let (X, d) be a metric space and T a subspace of X. For $x \in X$ and $r > 0$, we denote by $B_T(x, r) = \{y \in T : d(x, y) < r\}$ an open ball in T. When it is clear from context, we omit the subscript T. If $A \subseteq X$, we write $d(x, A) = \inf\{d(x, y) : y \in A\}$. By convention, $d(x, \emptyset) = \infty$.

A metric space (X, d) is said to be *doubling* if there exists some constant λ such that every ball in X can be covered by λ balls of half the radius. The smallest λ satisfying this condition is called the *doubling constant* of X, and is denoted λ_X. It is not difficult to see that, when X is finite, $\lambda \leq |X|$, or that every subset of $\mathbb{R}^k$ (under any norm) has doubling constant at most $2^{O(k)}$.

2 Metric decomposition

In this section, we introduce various notions of *stochastic decomposition* on a metric space. Related notions of probabilistic partitioning were used in [17, 2, 20, 9, 7], to name a few.

Let (T, d) be a metric space. Denote by $\mathcal{P}$ the set of all partitions of T. Given a $P \in \mathcal{P}$ and $x \in T$, define

$$\pi_P(x) = \sup_{C \in P} \min \left\{ d(x, T \setminus C), \operatorname{diam}(C) \right\}.$$

For every $t \in T$, let $\alpha_P(t) \in P$ be the unique cluster containing t.

We say that T admits a δ-padded Δ-bounded stochastic decomposition if there exists a probability distribution μ on $\mathcal{P}$ satisfying the following conditions.

(1) For every P with $\mu(P) > 0$, and for every $C \in P$, $\operatorname{diam}(C) \leq \Delta$.

(2) For every $x \in T$,

$$\Pr_{\mu}[\pi_P(x) \geq \delta\Delta] \geq \frac{1}{2}.$$

If condition (2) is replaced by

(2') For every $x, y \in T$

$$\Pr_{\mu}[\alpha_P(x) \neq \alpha_P(y)] \leq \frac{d(x, y)}{\delta\Delta},$$

then we say that T admits a δ-separating Δ-bounded stochastic decomposition. Finally, we say that such a decomposition μ is *efficiently samplable* if there exists an efficient (polynomial time) randomized procedure which outputs partitions according to μ. We end this section with a simple lemma about π_P which will be required in what follows.

Lemma 2.1. *For any partition P of T and any $x, y \in T$, $|\pi_P(x) - \pi_P(y)| \leq d(x, y)$.*

Proof. Let P be a partition of T, and let C_x, C_y be such that $x \in C_x, y \in C_y$. If $\pi_P(x) > d(x, T \setminus C_x)$, then $T = C$ and hence $\pi_P(x) = \pi_P(y)$; the same holds if $\pi_P(y) > d(y, T \setminus C_y)$. In the other case, we have

$$|\pi_P(x) - \pi_P(y)| = |d(x, T \setminus C_x) - d(y, T \setminus C_y)| \leq d(x, y).$$

The final inequality follows directly from the triangle inequality if $C_x = C_y$. On the other hand, when $C_x \neq C_y$, we must have $d(x, T \setminus C_x), d(y, T \setminus C_y) < d(x, y)$ (otherwise both $x, y \in C_x$). $\square$

2.1 From padded to separating

We will now show that it is possible to pass from padded decompositions to separating decompositions with only a small loss in the relevant constants.

Theorem 2.2. *Suppose that a metric space (T, d) admits a stochastic decomposition which is δ-padded and Δ-bounded, then it also admits one which is $\frac{1}{4}\delta$-separating and 2Δ-bounded. Additionally, if the former distribution is efficiently samplable, then so is the latter.*

Proof. Let μ be a δ-padded Δ-bounded stochastic decomposition of (T, d) and consider the following Las Vegas algorithm whose output we will show to be $\frac{1}{4}\delta$-separating and 2Δ-bounded. The algorithm produces a map $f : T \to T$ and the induced partition is $\{f^{-1}(t)\}_{t \in T}$, i.e. each preimage of an element $t \in T$ is a cluster. Let us define a selection function ψ on subsets of T by choosing, for every $C \subseteq T$, an arbitrarily element $t_C \in C$ and setting $\psi(C) = t_C$.

Algorithm PADDED-TO-SEPARATING
1. Start with the initial assignment $f(t) = \mathsf{null}$ for all $t \in T$.
2. **repeat until** all $t \in T$ are assigned:
3. Choose a partition P according to the distribution μ.
4. Choose a value $\gamma \in [\frac{1}{2}, 1]$ uniformly at random.
5. For every $t \in T$ which is not yet assigned,
   if $\pi_P(t) \geq \gamma\delta\Delta$, **then** set $f(t) = \psi(\alpha_P(t))$.

Note that, since μ is δ-padded, the **if** statement in line (5) succeeds with probability at least $\frac{1}{2}$ for any $t \in T$. It follows that the algorithm terminates with probability 1 and that the expected number of iterations is $O(1)$.

Lemma 2.3. *The diameter of each cluster* $f^{-1}(t)$ *is at most* 2Δ.

Proof. Let $C = f^{-1}(t)$, and consider $x \in C$. When x was assigned to t, it was because $\psi(\alpha_P(x)) = t$ so that $x, t \in \alpha_P(x)$. Since P was chosen from a Δ-bounded distribution, we conclude that $d(x, t) \leq \Delta$. Hence for any $x, y \in C$, we have $d(x, y) \leq d(x, t) + d(t, y) \leq 2\Delta$. ☐

Lemma 2.4. *For any* $x, y \in T$, $\Pr[f(x) \neq f(y)] \leq \frac{4\,d(x,y)}{\delta\Delta}$.

Proof. Consider $x, y \in T$ for which $d(x, y) < \frac{1}{2}\delta\Delta$ (since otherwise the claim is trivial). Initially, $f(x) = f(y) = \mathsf{null}$, and eventually (with probability 1), we have $f(x), f(y) \in T$. To ease in the analysis, let $f_i(x)$ be the value of $f(x)$ at the end of iteration i of the loop, and let q_i be the probability that $f_k(x) = f_k(y)$ for all $k < i$ but that $f_k(x) \neq f_k(y)$ for some $k \geq i$. Notice that $\Pr[f(x) \neq f(y)] \leq q_1$, hence it suffices to bound the latter quantity.

First, we see that

$$q_i \leq \Pr[f_i(x) \neq f_i(y) \mid f_{i-1}(x) = f_{i-1}(y) = \mathsf{null}] + q_{i+1} \cdot \Pr[f_i(x) = \mathsf{null}].$$

Let P be the partition drawn from μ in iteration i, and observe that $\Pr[f_i(x) = \mathsf{null}] \leq \Pr[\pi_P(x) < \delta\Delta] \leq \frac{1}{2}$.

Now we consider $\Pr[f_i(x) \neq f_i(y) \mid f_{i-1}(x) = f_{i-1}(y) = \mathsf{null}]$. To this end, suppose that $f_{i-1}(x) = f_{i-1}(y) = \mathsf{null}$, and note that the random variable $z = \gamma\delta\Delta$ ranges uniformly over an interval of size $\frac{1}{2}\delta\Delta$, thus *exactly one* of $f_i(x)$ or $f_i(x)$ is null only if z falls between the values $\pi_P(x)$ and $\pi_P(y)$. Since $|\pi_P(x) - \pi_P(y)| \leq d(x, y)$ by Lemma 2.1, this happens with probability at most $\frac{2\,d(x,y)}{\delta\Delta}$.

On the other hand, if $f_i(x)$ and $f_i(y)$ are both non-null, then $f_i(x) \in T$. It follows that $\pi_P(x) \geq \frac{1}{2}\delta\Delta$, and since $d(x, y) < \frac{1}{2}\delta\Delta$ by assumption, this implies that $\alpha_P(x) = \alpha_P(y)$ because $y \in B(x, d(x, y)) \subseteq \alpha_P(x)$. We conclude that

$$\Pr[f_i(x) \neq f_i(y) \mid f_{i-1}(x) = f_{i-1}(y) = \mathsf{null}] \leq \frac{2\,d(u, v)}{\delta\Delta}.$$

It follows that $q_i \leq \frac{2\,d(u,v)}{\delta\Delta} + \frac{1}{2} q_{i+1}$, from which we conclude that

$$q_1 \leq \frac{2\,d(u,v)}{\delta\Delta} \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i = \frac{4\,d(u,v)}{\delta\Delta}.$$

$\square$

The theorem now follows. $\square$

Remark 2.5. It should be noted that, although we prove it is always possible to obtain a separating decomposition from a padded decomposition, there certainly exist padded decompositions which are not separating.

Although we proved that whenever (T, d) admits a padded decomposition, it also admits a separating decomposition with roughly the same parameters, there are examples in which there exist separating decompositions with parameters that are not achievable by any padded decomposition. In particular, it has been proved in [5] that for every $\Delta > 0$, ℓ_2^d admits a $\frac{1}{2\sqrt{d}}$-separating Δ-bounded stochastic decomposition (and these decompositions partition $\mathbb{R}^d$ into Borel subsets). The following proposition shows that it is impossible to construct similar padded decompositions.

Proposition 2.6. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$ and let μ be a distribution over partitions of $\mathbb{R}^d$ into Borel subsets of diameter at most 1. Then there exists $x \in \mathbb{R}^d$ for which $\mathbb{E}_\mu \pi_P(x) \leq \frac{3}{d}$. In particular, if μ induces a δ-padded 1-bounded decomposition then $\delta \leq \frac{6}{d}$. The diameter and $\pi_P(\cdot)$ are computed above with respect to the norm $\|\cdot\|$.*

It should be remarked here that, under any norm, $\mathbb{R}^d$ admits a $\Omega(1/d)$ padded 1-bounded stochastic decomposition (see [9]).

Proposition 2.6 is an immediate consequence of the following geometric lemma:

Lemma 2.7. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$ with unit ball B. Then for every partition P of $\mathbb{R}^d$ into sets of diameter at most 1,*

$$\frac{1}{\mathrm{vol}\,(d^2 B)} \int_{d^2 B} \pi_P(x) dx \leq \frac{3}{d}.$$

For a set $C \subseteq \mathbb{R}^d$ and $\varepsilon > 0$ denote $C_{-\varepsilon} = \{y \in \mathbb{R}^d : \ d(y, \mathbb{R}^d \setminus C) > \varepsilon\}$ (distances are measured here according to the norm $\|\cdot\|$). The following geometric fact is a simple consequence of the Brunn-Minkowski inequality:

Lemma 2.8. *Let $C \subseteq \mathbb{R}^d$ be a Borel set of diameter at most 1. Then for every $\varepsilon \in (0, 1)$,*

$$\mathrm{vol}(C_{-\varepsilon}) \leq (1-\varepsilon)^d \mathrm{vol}(C).$$

Proof. Denote $\theta = \left(\frac{\mathrm{vol}(B)}{\mathrm{vol}(C)}\right)^{1/d}$. Since $\mathrm{diam}(C) \leq 1$, C is contained in a translate of B, so that $\theta \geq 1$. By definition, $(\theta C)_{-(\theta\varepsilon)} + \theta\varepsilon B \subseteq \theta C$, so the Brunn-Minkowski inequality (see e.g. [1]) implies that

$$\begin{aligned} \theta \mathrm{vol}(C)^{1/d} &= \mathrm{vol}(\theta C)^{1/d} \\ &\geq \mathrm{vol}((\theta C)_{-(\theta\varepsilon)} + \theta\varepsilon B)^{1/d} \\ &\geq \mathrm{vol}((\theta C)_{-(\theta\varepsilon)})^{1/d} + \mathrm{vol}(\theta\varepsilon B)^{1/d} \\ &= \theta \mathrm{vol}(C_{-\varepsilon})^{1/d} + \varepsilon\theta^2 \mathrm{vol}(C)^{1/d}. \end{aligned}$$

Hence, $\mathrm{vol}(C_{-\varepsilon}) \leq (1-\theta\varepsilon)^d \mathrm{vol}(C) \leq (1-\varepsilon)^d \mathrm{vol}(C)$, where we have used the fact that $\theta \geq 1$. $\square$

Proof of Lemma 2.7. Let P be a partition of $\mathbb{R}^d$ into Borel subsets of diameter at most 1 and fix $m > 1$, $\varepsilon > 0$. Observe that

$$\begin{aligned}(mB) \cap \{x \in \mathbb{R}^d : \pi_P(x) > \varepsilon\} &\subseteq \left(\bigcup_{mB \supseteq C \in P} C_{-\varepsilon} \right) \bigcup \left(\bigcup_{mB \not\supseteq C \in P} ((mB) \cap C) \right) \\ &\subseteq \left(\bigcup_{mB \supseteq C \in P} C_{-\varepsilon} \right) \bigcup (mB \setminus (m-1)B),\end{aligned}$$

where we have used the fact that for $C \in P$, $\mathrm{diam}(C) \le 1$. Hence, using Lemma 2.8 we get

$$\begin{aligned}\mathrm{vol}((mB) \cap \{x \in \mathbb{R}^d : \pi_P(x) > \varepsilon\}) &\le \sum_{mB \supseteq C \in P} \mathrm{vol}(C_{-\varepsilon}) + [m^d - (m-1)^d]\mathrm{vol}(B) \\ &\le (1-\varepsilon)^d \sum_{mB \supseteq C \in P} \mathrm{vol}(C) + [m^d - (m-1)^d]\mathrm{vol}(B) \\ &\le (1-\varepsilon)^d \mathrm{vol}(mB) + \left[1 - \left(1 - \frac{1}{m} \right)^d \right] \mathrm{vol}(mB).\end{aligned}$$

Finally,

$$\frac{1}{\mathrm{vol}(mB)} \int_{mB} \pi_P(x) dx = \int_0^1 \frac{\mathrm{vol}((mB) \cap \{x \in \mathbb{R}^d : \pi_P(x) > \varepsilon\})}{\mathrm{vol}(mB)} d\varepsilon \le \frac{1}{d-1} + 1 - \left(1 - \frac{1}{m}\right)^d.$$

Choosing $m = d^2$ yields the required result. □

2.2 Partition extension

Now we show that if (X, d) is a metric space and T is a subspace of X, then a δ-padded Δ-bounded stochastic decomposition of T can be extended to a (roughly) δ-padded Δ-bounded stochastic decomposition of an r-neighborhood of T in X, i.e. to the entire set

$$N_r(T) = \{x \in X : d(x, T) < r\},$$

with $r \approx \delta\Delta$. For this purpose, the padded decomposition is essential as a similar extension result does not hold for separating decompositions. This extension will be key in applications of decompositions to the 0-extension problem in Section 3.

The new distribution is produced by modifying the old distribution one partition at a time and thus the transformation preserves efficient samplability.

Lemma 2.9 (Partition extension). *Let (X, d) be a metric space and T a subspace of X. If T admits a δ-padded Δ-bounded stochastic decomposition, then $N_r(T)$ admits a $\delta/8$-padded $(1 + \frac{\delta}{2})\Delta$-bounded stochastic decomposition, where $r = \frac{\delta}{8}\Delta$.*

Proof. Let μ be an δ-padded Δ-bounded stochastic decomposition of T, and consider some partition $P \in \mathrm{supp}(\mu)$. We extend P to a partition of X as follows. For every point $x \in X$, let $t_x \in T$ be such that $d(x, t_x) = d(x, T)$. Now, for every cluster $C \in P$, create a super-cluster

$$C' = C \cup \{x \in X : B_T(t_x, \delta\Delta/2) \subseteq C \text{ and } d(x, t_x) \le \delta\Delta/4\}.$$

Finally, for any point $x \in X \setminus \bigcup_{C \in P} C'$, place x in a singleton cluster $\{x\}$. This constitutes a partition P' of X. Finally, we arrive at a partition P'' of $N_r(T)$ by restricting P' to $N_r(T)$, i.e. by letting $C'' = C' \cap N_r(T)$ and setting $P'' = \{C''\}_{C' \in P'}$.

Let us now show that the distribution on partitions P'' is $\delta/8$-padded and $(1+\frac{\delta}{2})\Delta$-bounded. The $(1+\frac{\delta}{2})\Delta$-bounded condition is easy; singleton clusters have diameter zero. For points $x, y \in C''$, we have

$$d(x,y) \le d(x,t_x) + d(t_x,t_y) + d(y,t_y) \le \delta\Delta/4 + \Delta + \delta\Delta/4 = \left(1 + \frac{\delta}{2}\right)\Delta.$$

To see that the decomposition is $\delta/8$-padded, fix some $x \in N_r(T)$, i.e. such that $d(x,T) < \delta\Delta/8$. By property (2) of the δ-padded decomposition of T, with probability at least $\frac{1}{2}$, $B_T(t_x, \delta\Delta) \subseteq C$ for some $C \in P$. Our goal will be to show that, in this case, $B_X(x, \delta\Delta/8) \subseteq C'$. It will follow that with probability at least $\frac{1}{2}$, x is $\delta\Delta/8$-padded, yielding the desired result.

We now show that $B_X(x, \delta\Delta/8) \subseteq C'$. Fix some $y \in X$ with $d(x,y) \le \delta\Delta/8$. Observe that

$$d(t_x,t_y) \le d(t_x,x) + d(x,y) + d(y,t_y) \le 2 \cdot (d(t_x,x) + d(x,y)) \le \varepsilon\Delta/2,$$

where we note that $d(y,t_y) = d(y,T) \le d(x,y) + d(x,T)$. It follows that $B_T(t_y, \delta\Delta/2) \subseteq B_T(t_x, \delta\Delta) \subseteq C$. Since we also have that $d(y,t_y) \le d(x,y) + d(x,t_x) \le \delta\Delta/4$, we see that, indeed, $y \in C'$. $\square$

We recall the following two theorems.

Theorem 2.10 ([13], [20], [8]). *If (T,d) excludes a $K_{r,r}$-minor, then there exist a constant $\delta = \Omega(\frac{1}{r^2})$ such that, for every $\Delta > 0$, (T,d) admits a δ-padded Δ-bounded stochastic decomposition.*

Theorem 2.11 ([9]). *For any metric (T,d) and any $\Delta > 0$, (T,d) admits a $\frac{1}{64\log\lambda}$-padded Δ-bounded stochastic decomposition, where λ is the doubling constant of T.*

3 The 0-extension problem

In the *0-extension problem*, one is given an undirected graph $G = (V,E)$ with a non-negative cost function $c : E \to \mathbb{R}_+$ on the edges, a subset $T \subseteq V$ called the *terminals*, and a metric d on T. The goal is output an assignment $f : V \to T$ of vertices to terminals so that $f|_T$ is the identity on T and which minimizes $\sum_{uv \in E} c(u,v) \cdot d(f(u), f(v))$. This is a generalization of the *multi-way cut problem* and a special case of *metric labeling* (see [4, 6] for a more comprehensive history of the problem). It is clearly a formulation of a natural clustering/classification problems.

In this section, we prove that when the input metric (T,d) on terminals admits, for every $k \in \mathbb{Z}$, a δ-padded 2^k-bounded efficiently samplable stochastic decomposition, that there exists an $O(\frac{1}{\delta})$-approximation algorithm. In particular, we obtain a constant factor approximation for planar metrics and an $O(\log\lambda)$-approximation where λ is the doubling constant of T (this follows from Theorems 2.10 and 2.11 of the previous section).

As in [4, 6], we obtain an approximation algorithm by designing a randomized rounding scheme for the following linear program which was first used by [12].

$$\text{Minimize} \qquad \sum_{uv \in E} c(u,v) \cdot \rho(u,v) \quad \text{subject to}$$

$$(V, \rho) \text{ is a semi-metric}$$
$$\rho(u,v) = d(u,v) \qquad \forall\, u,v \in T.$$

Recall that a semi-metric on V is a function $\rho(u,v)$ which is a metric except that we may have $\rho(u,u) = 0$. This is clearly a linear relaxation of the 0-extension problem since an assignment $f : V \to T$ yields a semi-metric $\rho(u,v) = d(f(u), f(v))$. Thus a feasible solution yields a lower bound on the cost of an optimal extension. Our goal will be to compute a map $f : V \to T$ assigning nodes of V to terminals (f will be the identity map when restricted to T). We will then show that, for every $x, y \in V$, $\mathbb{E}\, d(f(x), f(y)) \leq O(\frac{1}{\delta})\, \rho(x,y)$. It will follow that the expected cost of our computed solution is within a $O(\frac{1}{\delta})$ factor of the cost of the LP solution, and hence within the same factor of the optimal.

Before we proceed, let us assume that ρ is not a semi-metric, but an actual metric. This assumption can be made without loss for the following reason. If, for two points $x, y \in V \setminus T$, we have $\rho(x,y) = 0$, then we may delete y and apply the convention that $f(y) = f(x)$. The other exceptional case, when $\rho(x,t) = 0$ and $t \in T$ is some terminal, can be dealt with identically, noting that $f(x) = f(t)$ makes sense since f is the identity map on T.

The rounding algorithm. We now present the rounding algorithm. Let (V, ρ) be the metric obtained from the LP. For each relevant value of $k \in \mathbb{Z}$, let μ_k be an δ-padded 2^k-bounded stochastic decomposition of T. Let $N_k = \{x \in V : \rho(x,T) < \frac{\delta}{8} 2^k\}$, and using partition extension (Lemma 2.9), let μ'_k be a $\frac{\delta}{8}$-padded 2^{k+1}-bounded stochastic decomposition of N_k. Finally, using Theorem 2.2, convert μ'_k to a $\frac{\delta}{32}$-separating 2^{k+1}-bounded decomposition of N_k called μ''_k. Define a selection function ψ on subsets of V as follows. For each $C \subseteq V$, let $\psi(C) \in T$ be such that $d(C,T) = d(\psi(C), C)$.

Algorithm RoundLP

1. Start with the initial assignment $f(t) = t$ for all $t \in T$.
2. Choose a value $\beta \in [1,2]$ uniformly at random, and define, for every $v \in V$,
   $\mathsf{scale}(v) = \min\{k : \frac{8\beta}{\delta} \rho(v,T) \geq 2^k\}$.
3. For every relevant value of $k \in \mathbb{Z}$, let P_k be a partition of N_k chosen according to μ''_k.
4. For $v \in V \setminus T$, let $m = \mathsf{scale}(v)$ and set $f(v) = \psi(\alpha_{P_m}(v))$.

The analysis. As a sanity check, note that there are at most $|V|$ relevant values of $k \in \mathbb{Z}$. Also, notice that line (4) always makes sense, since if $m = \mathsf{scale}(v)$, then $\rho(v,T) < \frac{\delta}{8} 2^m$ which implies that $v \in N_m$. Now, recall that our goal is to show that $\mathbb{E}\, d(f(u), f(v)) \leq O\left(\frac{1}{\delta}\right) \rho(u,v)$. We begin by showing that $f(v)$ is never too far away from v. In what follows, we make no attempt to optimize the constants.

Lemma 3.1. *For every $v \in V$, we have $\rho(v, f(v)) \leq \frac{33}{\delta}\, \rho(v,T)$.*

Proof. Let $m = \mathsf{scale}(v)$ (as assigned by the algorithm), and notice that $2^{m-1} \le \frac{8}{\delta}\,\rho(v,T)$, hence, letting $C = \alpha_{P_m}(v)$ be the cluster which contains v, we have (1) $\mathrm{diam}(C) \le 2^{m+1} \le \frac{32}{\delta}\,\rho(v,T)$ and (2) $\rho(C, f(v)) = \rho(C,T)$. It follows that $\rho(v, f(v)) \le \mathrm{diam}(C) + \rho(v,T) \le \frac{33}{\delta}\,\rho(v,T)$. □

Lemma 3.2. *If $\rho(u,v) \ge \frac{1}{2}\,\rho(\{u,v\},T)$, then $d(f(u),f(v)) \le \frac{166}{\delta}\,\rho(u,v)$.*

Proof. Assume without loss of generality that $\rho(u,v) \ge \frac{1}{2}\,\rho(u,T)$. Note that $\rho(v,T) \le \rho(u,v) + \rho(u,T) \le 3\,\rho(u,v)$. Now, using Lemma 3.1,

$$\begin{aligned} d(f(u),f(v)) = \rho(f(u),f(v)) &\le \rho(f(u),u) + \rho(u,v) + \rho(v,f(v)) \\ &\le \frac{33}{\delta}\,(\rho(u,T) + \rho(v,T)) + \rho(u,v) \\ &\le \frac{166}{\delta}\,\rho(u,v). \end{aligned}$$

□

Thus we are left to consider the case when $\rho(u,v) < \frac{1}{2}\,\rho(\{u,v\},T)$.

Lemma 3.3. *If $\rho(u,v) < \frac{1}{2}\,\rho(\{u,v\},T)$, then $\Pr_\beta[\mathsf{scale}(u) \ne \mathsf{scale}(v)] \le \frac{2\rho(u,v)}{\rho(u,T)}$.*

Proof. Without loss of generality, assume that $\rho(u,T) \le \rho(v,T)$, and note that $|\rho(u,T) - \rho(v,T)| \le \rho(u,v)$. Let $m = \mathsf{scale}(v)$, and notice that $\mathsf{scale}(u) \ne \mathsf{scale}(v)$ only if $\frac{8}{\delta}\,\rho(u,T) \le \frac{2^{m+1}}{\beta} \le \frac{8}{\delta}\,\rho(v,T)$. Since $\frac{2^{m+1}}{\beta}$ ranges over an interval of size $2^m \ge \frac{4}{\delta}\,\rho(u,T)$, the probability of this occurring is at most $\frac{\frac{8}{\delta}\rho(u,v)}{\frac{4}{\delta}\rho(u,T)} = \frac{2\rho(u,v)}{\rho(u,T)}$. □

Lemma 3.4. $\Pr[f(u) \ne f(v) \,|\, \mathsf{scale}(u) = \mathsf{scale}(v)] \le \frac{4\,\rho(u,v)}{\rho(u,T)}$.

Proof. This follows because μ''_m is a $\frac{\delta}{32}$-separating 2^{m+1}-bounded stochastic decomposition of N_m (as before, $m = \mathsf{scale}(v)$). In particular,

$$\Pr[f(u) \ne f(v) \,|\, \mathsf{scale}(u) = \mathsf{scale}(v)] = \Pr[\alpha_P(u) \ne \alpha_P(v)] \le \frac{32}{\delta}\frac{\rho(u,v)}{2^{m+1}} \le \frac{32}{\delta}\frac{\delta}{8}\frac{\rho(u,v)}{\rho(u,T)}.$$

□

We conclude by noticing that, when $\rho(u,v) < \frac{1}{2}\,\rho(\{u,v\},T)$, we have $d(f(u),f(v)) \le \rho(u,f(u)) + \rho(u,v) + \rho(v,f(v)) \le O(\frac{1}{\delta})\,\rho(u,T)$ by Lemma 3.1. Hence in this case,

$$\mathbb{E}\,d(f(u),f(v)) \le \Pr[f(u) \ne f(v)] \cdot O\left(\frac{1}{\delta}\right)\,\rho(u,T) \le 6\,\frac{\rho(u,v)}{\rho(u,T)} \cdot O\left(\frac{1}{\delta}\right)\,\rho(u,T),$$

and the latter quantity is at most $O(\frac{1}{\delta})\,\rho(u,v)$, completing the proof.

In conjunction with Theorems 2.10 and 2.11, we get the following guarantees.

Corollary 3.5. *For any excluded minor space (T,d) (e.g. planar metrics), there exists a constant factor approximation for the 0-extension problem when (T,d) is the metric on terminals.*

Corollary 3.6. *For any metric space (T,d) with doubling constant λ, there exists an $O(\log\lambda)$-approximation for the 0-extension problem when (T,d) is the metric on terminals.*

4 Improved approximations from non-uniform distributions

In this section, we give an alternate proof of the $O(\log \lambda)$-approximation from Section 3 by replacing the uniform distribution in the algorithm of [4] by a probability distribution which reflects the structure of the terminal metric.

A measure μ on a finite metric space (T, d) is called *doubling* if there exists a constant K such that for every $t \in T$ and $r > 0$, we have

$$\mu(B(t, 2r)) \leq K \mu(B(t, r)).$$

The following is a special case of a theorem of [21] (see also, [10]).

Theorem 4.1. *For a finite metric (T, d), there exists a measure μ on T which is doubling with constant $K = O(\lambda^c)$, where λ is the doubling constant of T, and $c > 0$ is some universal constant.*

Notice that if we define, for every subset $S \subseteq T$, $p(S) = \mu(S)/\mu(T)$, then $p(T) = 1$ and p becomes a doubling *probability measure* on T. In this section, we show that replacing the uniform probability measure in the algorithm of [4], we can improve the approximation ratio in that paper from $O(\log |T|)$ to $O(\log \lambda)$ without any further alteration. We believe this is a powerful technique which has many potential applications in theory and in practice.

Let (T, d) be the metric on terminals, $k = |T|$, and (V, ρ) be the semi-metric returned by the LP of Section 3. The main element of the approximation algorithm of [4] is a rounding algorithm which proceeds as follows:

> Set $f(t) = t$ for all terminals $t \in T$.
> Pick a random permutation $\sigma = \langle \sigma_1, \ldots, \sigma_k \rangle$ of the terminals. (*)
> Pick α uniformly at random in $[2, 4)$.
> **for** $j = 1$ to k **do**
>   **for** all unassigned terminals u such that $\rho(u, \sigma_j) \leq \alpha\, \rho(u, T)$, **do**
>   set $f(u) = \sigma_j$

Their main lemma is the following.

Lemma 4.2 ([4]). *For every pair $x, y \in V$,*

$$\Pr[f(x) \neq f(y)] \leq O(\log k)\, \rho(x, y) \left(\frac{1}{\rho(x, T)} + \frac{1}{\rho(y, T)} \right).$$

We now prove that if (*) is replaced by the following line, then their algorithm becomes an $O(\log \lambda)$-approximation. Let $p : T \to \mathbb{R}_+$ be the probability measure induced on T by the doubling measure guaranteed by Theorem 4.1.

> Pick a random permutation $\sigma = \langle \sigma_1, \ldots, \sigma_k \rangle$ where σ is generated as follows.
>   Choose elements $t \in T$ with probability proportional to $p(t)$ (without replacement) until T is exhausted. Let σ_i be the ith chosen element.

In other words, we choose the permutation σ, not uniformly, but relative to a probability measure on T which is tailored to its metric structure. To exhibit the claimed approximation ratio, it is enough to prove the following claim (see [4] for details).

Claim 4.3. *For every pair $x, y \in V$,*

$$\Pr[f(x) \neq f(y)] \leq O(\log \lambda)\, \rho(x,y) \left(\frac{1}{\rho(x,T)} + \frac{1}{\rho(y,T)} \right).$$

Proof. Let $\mathcal{E}_{xy}$ be the event that x is assigned to a terminal *strictly preceding* y in the permutation σ. We will argue that

$$\Pr[\mathcal{E}_{xy}] \leq O(\log \lambda)\, \frac{\rho(x,y)}{\rho(x,T)}.$$

Noting that $\Pr[f(x) \neq f(y)] = \Pr[\mathcal{E}_{xy}] + \Pr[\mathcal{E}_{yx}]$, this will complete the proof.

To this end, fix $x, y \in X$. For each $t \in T$, let $I_t = [\rho(t,x), \rho(t,x) + \rho(x,y)]$, and note that if x is assigned to $t \in T$ before y is assigned, it must be the case that $\rho(x,t) < \rho(y,t)$, and it follows that for such t, $B(t,R)$ separates x and y precisely when $R \in I_t$. We set $R = \alpha\, \rho(x,T)$ and think of R as being chosen uniformly in $[2\,\rho(x,T), 4\,\rho(x,T)]$. Let $B_x = B_T(x, 4\,\rho(x,T))$, and let $\mathcal{E}_t$ be the event that $f(x) = t$, but $y \notin B(t,R)$, then

$$\begin{aligned} \Pr[\mathcal{E}_{xy}] = \sum_{t \in T} \Pr[\mathcal{E}_t] = \sum_{t \in B_x} \Pr[\mathcal{E}_t] \quad &\leq \quad \sum_{t \in B_x} \Pr[R \in I_t] \cdot \Pr[f(x) = t \,|\, R \in I_t] \\ &\leq \quad \frac{\rho(x,y)}{2\,\rho(x,T)} \sum_{t \in B_x} \Pr[f(x) = t \,|\, R \in I_t]. \end{aligned}$$

For the sake of analysis (i.e., *without* altering the algorithm), we make some assumptions without loss of generality. First, we may assume that for every $t \in T$, $p(t) = 1$ by rational approximation of $p(t)$, scaling by common denominators, and then duplication of points (i.e. replacing t by many copies of itself). Although p is no longer a probability measure, only the relative measures of sets will come into the analysis, and these will remain the same. Also, we assume that $\rho(x,t_1) \neq \rho(x,t_2)$ for all $t_1, t_2 \in T$ by applying an arbitrarily small perturbation to T (and copies of points in T). (This is used only for convenience in what follows and is not actually necessary.)

Now, let us consider annuli $\{A_k\}_{k=1}^m$ centered at x, with $A_k = \{t \in T : R_{k-1} \leq \rho(x,t) < R_k\}$, where $R_1 = 2\,\rho(x,T)$, $R_m = 4\,\rho(x,T)$, and the intermediate values $\{R_k\}$ are chosen so that $p(A_1) = p(A_2) = \cdots = p(A_{m-1})$ and $p(A_m) \leq p(A_1)$. Notice that such values R_k can be chosen because of the assumption that no two points of T lie at exactly the same distance from x. Now, we can write

$$\sum_{t \in B_x} \Pr[f(x) = t \,|\, R \in I_t] = \sum_{k=1}^{m} \left(\sum_{t \in A_k} \Pr[f(x) = t \,|\, R \in I_t] \right) \leq 1 + \sum_{k=1}^{m-1} \frac{1}{k} = O(\log m).$$

To see that the penultimate inequality holds, note that when $R \in I_t$ for some $t \in A_k$, x would be assigned to any point in $\cup_{i=1}^{k-1} A_i$ if such a point were chosen before t. Thus the probability that $f(x) = t$ for some $t \in A_k$ is at most

$$\frac{p(A_k)}{p(\cup_{i=1}^{k-1} A_i)} = \frac{p(A_k)}{\sum_{i=1}^{k-1} p(A_i)} \leq \frac{1}{k-1}.$$

Finally, let t_x be such that $d(x,T) = d(x,t_x)$ and note that $p(B_T(x, 4\,\rho(x,T))) = \sum_{k=1}^m p(A_k) \geq (m-1)\, p(A_0) = (m-1)\, p(B(x, 2\,\rho(x,T)))$. On the other hand, since p is doubling with constant

K, we see that

$$\begin{aligned} p\Big(B_T(x, 4\,\rho(x,T))\Big) \le p\Big(B_T(t_x, 5\,\rho(x,T))\Big) &\le K^3\, p\Big(B_T(t_x, \rho(x,T))\Big) \\ &\le K^3\, p\Big(B_T(x, 2\,\rho(x,T))\Big). \end{aligned}$$

It follows that $m = O(K^3)$, hence $\Pr[\mathcal{E}_{xy}] \le O(\log K)\,\frac{\rho(x,y)}{\rho(x,T)}$. Since $O(\log K) = O(\log \lambda)$, this yields the desired result. □

Remark. It is easy to see that the *optimal* doubling measure on a finite metric space, i.e. the one which minimizes K, can be computed in polynomial time.

5 Separating decompositions of finite subsets of L_p, $1 < p \le 2$

Our goal here is to prove the following result.

Theorem 5.1. *For every $1 < p \le 2$ there is a constant $0 < C_p < \infty$ such that for every integer n and $\Delta > 0$, every n-point subset of L_p admits a $C_p(\log n)^{-1/p}$-separating Δ-bounded stochastic decomposition.*

The proof of Theorem 5.1 uses the following decomposition, due to Charikar, Chekuri, Goel, Guha and Plotkin [5].

Theorem 5.2. *For every $\Delta > 0$, ℓ_2^d admits a $\frac{1}{2\sqrt{d}}$-separating Δ-bounded stochastic decomposition.*

A straightforward application of the Johnson-Lindenstrauss dimension reduction lemma [11] shows that when $p = 2$, Theorem 5.1 follows from Theorem 5.2. We pass to arbitrary $1 < p < 2$ using a result of Marcus and Pisier [18].

Proof of Theorem 5.1. Fix $1 < p \le 2$ and let $X \subseteq L_p$ be an n-point subset. In [18] Marcus and Pisier show that there is a probability space (Ω, P) such that for every $\omega \in \Omega$ there is a linear operator $S_\omega : L_p \to L_2$ such that for every $x \in L_p \setminus \{0\}$ the random variable $X = \frac{\|S_\omega(x)\|_2}{\|x\|_p}$ satisfies for every $a \in \mathbb{R}$, $\mathbb{E}e^{-aX^2} = e^{-a^{p/2}}$. A standard application of Markov's inequality shows that there is constant c_p and a subset $A \subset \Omega$ with $P(A) \ge \frac{1}{2}$ such that for every $x, y \in X$ and $\omega \in A$,

$$\|x - y\|_p \le c_p (\log n)^{\frac{1}{p} - \frac{1}{2}} \|S_\omega(x) - S_\omega(y)\|_2. \tag{1}$$

By the Johnson-Lindenstrauss dimension reduction lemma [11] there is an integer $d = O(\log n)$ such that for every $\omega \in A$ there is a function $g_\omega : S_\omega(X) \to \ell_2^d$ satisfying for all $x, y \in X$,

$$\|S_\omega(x) - S_\omega(y)\|_2 \le \|g_\omega(S_\omega(x)) - g_\omega(S_\omega(y))\|_2 \le 2\|S_\omega(x) - S_\omega(y)\|_2. \tag{2}$$

Now, by the Theorem 5.2 there is a distribution μ_ω over partitions of $g_\omega(S_\omega(X))$ which is $\Delta/\left(c_p (\log n)^{\frac{1}{p} - \frac{1}{2}}\right)$ bounded and for every $x, y \in X$,

$$\begin{aligned} \mu_\omega\left[g_\omega(S_\omega(x)) \text{ and } g_\omega(S_\omega(y)) \text{ are separated}\right] &\le O\left(\frac{c_p (\log n)^{\frac{1}{p}}}{\Delta}\right) \|g_\omega(S_\omega(x)) - g_\omega(S_\omega(y))\|_2 \\ &\le O\left(\frac{c_p (\log n)^{\frac{1}{p}}}{\Delta}\right) \|S_\omega(x) - S_\omega(y)\|_2, \end{aligned}$$

where we have used the upper bound in (2).

Define a probability measure P' on A by $P' = \frac{1}{P(A)} \cdot P$. Every partition $\{C_i\}$ of $g_\omega(S_\omega(X))$ can be pulled back to a partition of X given by $\{S_\omega^{-1}(g_\omega^{-1}(C_i))\}$, so that we get an induced distribution, ν, over partitions of X. Observe that if the partition $\{C_i\}$ is in the support of μ_ω then for every i,

$$\operatorname{diam}(S_\omega^{-1}(g_\omega^{-1}(C_i))) \le c_p (\log n)^{\frac{1}{p}-\frac{1}{2}} \operatorname{diam}(C_i) \le \Delta,$$

where we have used (1). So, the distribution ν is Δ bounded. Finally, for every $x, y \in X$, since $P(A) \ge \frac{1}{2}$,

$$\begin{aligned}
\nu\left[x \text{ and } y \text{ are separated}\right] &= \int_A \mu_\omega \left[g_\omega(S_\omega(x)) \text{ and } g_\omega(S_\omega(y)) \text{ are separated}\right] dP'(\omega) \\
&\le 2\int_A \mu_\omega \left[g_\omega(S_\omega(x)) \text{ and } g_\omega(S_\omega(y)) \text{ are separated}\right] dP(\omega) \\
&\le O\left(\frac{c_p (\log n)^{\frac{1}{p}}}{\Delta}\right) \int_A \|S_\omega(x) - S_\omega(y)\|_2 dP(\omega) \\
&= O\left(\frac{c_p (\log n)^{\frac{1}{p}}}{\Delta}\right) \cdot \|x-y\|_p \cdot \mathbb{E}X,
\end{aligned}$$

and we conclude since for $p > 1$, $\mathbb{E}X = C_p < \infty$.

□

Acknowledgements

The first author thanks Elaine Park for an enlightening discussion.

References

[1] K. Ball. An elementary introduction to modern convex geometry. In *Flavors of geometry*, volume 31 of *Math. Sci. Res. Inst. Publ.*, pages 1–58. Cambridge Univ. Press, Cambridge, 1997.

[2] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science*, pages 184–193. IEEE, 1996.

[3] Y. Bartal. On approximating arbitrary metrices by tree metrics. In *30th Annual ACM Symposium on Theory of Computing*, pages 161–168. ACM, 1998.

[4] G. Calinescu, H. Karloff, and Y. Rabani. Approximation algorithms for the 0-extension problem. In *SODA*, 2001.

[5] M. Charikar, C. Chekuri, T. Cheung, Z. Dai, A. Goel, S. Guha, and M. Li. Approximation algorithms for directed Steiner problems. *J. Algorithms*, 33(1):73–91, 1999.

[6] J. Fakcharoenphol, C. Harrelson, S. Rao, and K. Talwar. An improved approximation algorithm for the 0-extension problem. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*. ACM, 2003.

[7] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *35th Annual ACM Symposium on Theory of Computing*. ACM, 2003. To appear.

[8] J. Fakcharoenphol and K. Talwar. An improved decomposition theorem for graphs excluding a fixed minor. In *7th International Workshop on Randomization and Approximation Techniques in Computer Science*. Springer-Verlag, 2003.

[9] A. Gutpa, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *Proceedings of the 44th Annual Symposium Foundations of Computer Science, to appear*, 2003.

[10] J. Heinonen. *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, New York, 2001.

[11] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[12] A. Karzanov. Minimum 0-extensions of graph metrics. *European Journal of Combinatorics*, 19(1):71–101, 1998.

[13] P. Klein, S. A. Plotkin, and S. Rao. Excluded minors, network decomposition, and multicommodity flow. In *25th Annual ACM Symposium on Theory of Computing*, pages 682–690, May 1993.

[14] G. Konjevod, R. Ravi, and F. S. Salman. On approximating planar metrics by tree metrics. *Inform. Process. Lett.*, 80(4):213–219, 2001.

[15] R. Krauthgamer and J. R. Lee. The intrinsic dimensionality of graphs. In *35th Annual ACM Symposium on Theory of Computing*. ACM, 2003. To appear.

[16] R. Krauthgamer and J. R. Lee. Navigating nets: Simple algorithms for proximity search. Submitted, 2003.

[17] N. Linial and M. Saks. Low diameter graph decompositions. *Combinatorica*, 13(4):441–454, 1993.

[18] M. B. Marcus and G. Pisier. Characterizations of almost surely continuous p-stable random Fourier series and strongly stationary processes. *Acta Math.*, 152(3-4):245–301, 1984.

[19] Y. Rabinovich. On average distortion of embedding metrics into the line and into L_1. In *35th Annual ACM Symposium on Theory of Computing*. ACM, 2003.

[20] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the 15th Annual Symposium on Computational Geometry*, pages 300–306. ACM, 1999.

[21] A. L. Vol'berg and S. V. Konyagin. On measures with the doubling condition. *Math. USSR-Izv.*, 30:629–638, 1988.