Maximum Gradient Embeddings and Monotone Clustering

Extended abstract*

Manor Mendel[1] and Assaf Naor[2]

[1] The Open University of Israel
[2] Courant Institute

Abstract. Let (X, d_X) be an n-point metric space. We show that there exists a distribution $\mathscr{D}$ over non-contractive embeddings into trees $f : X \to T$ such that for every $x \in X$,

$$\mathbb{E}_{\mathscr{D}}\left[\max_{y\in X\setminus\{x\}} \frac{d_T(f(x), f(y))}{d_X(x, y)}\right] \le C(\log n)^2,$$

where C is a universal constant. Conversely we show that the above quadratic dependence on $\log n$ cannot be improved in general. Such embeddings, which we call *maximum gradient embeddings*, yield a framework for the design of approximation algorithms for a wide range of clustering problems with monotone costs, including fault-tolerant versions of k-median and facility location.

1 Introduction

We introduce a new notion of embedding, called *maximum gradient embeddings*, which is "just right" for approximating a wide range of clustering problems. We then provide optimal maximum gradient embeddings of general finite metric spaces, and use them to design approximation algorithms for several natural clustering problems. Rather than being encyclopedic, the main emphasis of the present paper is that these embeddings yield a generic approach to many problems.

Due to their special structure, it is natural to try to embed metric spaces into trees. This is especially important for algorithmic purposes, as many hard problems are tractable on trees. Unfortunately, this is too much to hope for in the bi-Lipschitz category: As shown by Rabinovich and Raz [22] the n-cycle incurs distortion $\Omega(n)$ in any embedding into a tree. However, one can relax this idea and look for a *random* embedding into a tree which is faithful on average. Such an approach has been developed in recent years by mathematicians and computer scientists. In the mathematical literature this is referred to as embeddings into products of trees, and it is an invaluable tool in the study of negatively curved spaces (see for example [7, 10, 20]).

* Full version appears in [19].

Probabilistic embeddings into *dominating* trees became an important algorithmic paradigm due to the work of Bartal [3, 4] (see also [1, 11] for the related problem of embedding graphs into distributions over spanning trees). This work led to the design of many approximation algorithms for a wide range of NP hard problems. In some cases the best known approximation factors are due to the "probabilistic tree" approach, while in other cases improved algorithms have been subsequently found after the original application of probabilistic embeddings was discovered. But, in both cases it is clear that the strength of Bartal's approach is that it is generic: For a certain type of problem one can quickly get a polylogarithmic approximation using probabilistic embedding into trees, and then proceed to analyze certain particular cases if one desires to find better approximation guarantees. However, probabilistic embeddings into trees do not always work. In [5] Bartal and Mendel introduced the weaker notion of multi-embeddings, and used it to design improved algorithms for special classes of metric spaces. Here we *strengthen* this notion to maximum gradient embeddings, and use it to design approximation algorithms for harder problems to which regular probabilistic embeddings do not apply.

Let (X, d_X) and (Y, d_Y) be metric spaces, and fix a mapping $f : X \to Y$. The mapping f is called *non-contractive* if for every $x, y \in X$, $d_Y(f(x), f(y)) \geq d_X(x, y)$. The *maximum gradient* of f at a point $x \in X$ is defined as

$$|\nabla f(x)|_\infty = \sup_{y \in X \setminus \{x\}} \frac{d_Y(f(x), f(y))}{d_X(x, y)}. \tag{1}$$

Thus the *Lipschitz constant* of f is given by $\|f\|_{\mathrm{Lip}} = \sup_{x \in X} |\nabla f(x)|_\infty$.

In what follows when we refer to a tree metric we mean the shortest-path metric on a graph-theoretical tree with weighted edges. Recall that (U, d_U) is an ultrametric if for every $u, v, w \in U$ we have $d_U(u, v) \leq \max\{d_U(u, w), d_U(w, v)\}$. It is well known that ultrametrics embed isometrically into tree metrics. The following result is due to Fakcharoenphol, Rao and Talwar [12], and is a slight improvement over an earlier theorem of Bartal [4]. For every n-point metric space (X, d_X) there is a distribution $\mathscr{D}$ over non-contractive embeddings into ultrametrics $f : X \to U$ such that

$$\max_{\substack{x, y \in X \\ x \neq y}} \mathbb{E}_{\mathscr{D}} \frac{d_U(f(x), f(y)}{d_X(x, y)} = O(\log n). \tag{2}$$

The logarithmic upper bound in (2) cannot be improved in general.

Inequality (2) is extremely useful for optimization problems whose objective function is linear in the distances, since by linearity of expectation it reduces such tasks to trees, with only a logarithmic loss in the approximation guarantee. When it comes to non-linear problems, the use of (2) is very limited. We will show that this issue can be addressed using the following theorem, which is our main result.

Theorem 1. *Let (X, d_X) be an n-point metric space. Then there exists a distribution $\mathscr{D}$ over non-contractive embeddings into ultrametrics $f : X \to U$ (thus* *both the ultrametric* (U, d_U) *and the mapping* f *are random) such that for every* $x \in X$, $\mathbb{E}_{\mathscr{D}}|\nabla f(x)|_\infty \le C(\log n)^2$, *where* C *is a universal constant.*

On the other hand, there exists a universal constant $c > 0$ *and arbitrarily large* n*-point metric spaces* Y_n *such that for any distribution over non-contractive embeddings into trees* $f : Y_n \to T$ *there is necessarily some* $x \in Y_n$ *for which* $\mathbb{E}_{\mathscr{D}}|\nabla f(x)|_\infty \ge c(\log n)^2$.

We call embeddings as in Theorem 1, i.e. embeddings with small expected maximum gradient, *maximum gradient embeddings into distributions over trees* (in what follows we will only deal with distributions over trees, so we will drop the last part of this title when referring to the embedding, without creating any ambiguity). The proof of the upper bound in Theorem 1 is a modification of an argument of Fakcharoenphol, Rao and Talwar [12], which is based on ideas from [3, 8]. Alternative proofs of the main technical step of the proof of the upper bound in Theorem 1 can be also deduced from the results of [18] or an argument in the proof of Lemma 2.1 in [13].

The heart of this paper is the lower bound in Theorem 1. The metrics Y_n in Theorem 1 are the diamond graphs of Newman and Rabinovich [21], which have been previously used as counter-examples in several embedding problems— see [6, 14, 17, 21].

1.1 A framework for clustering problems with monotone costs

We now turn to some algorithmic applications of Theorem 1. The general reduction in Theorem 2 below should also be viewed as an explanation why maximum gradient embeddings are so natural— they are precisely the notion of embedding which allows such reductions to go through. . In the full version of this paper we also analyze in detail two concrete optimization problems which belong to this framework.

A very general setting of the clustering problem is as follows. Let X be an n-point set, and denote by $\mathrm{MET}(X)$ the set of all metrics on X. A *possible clustering solution* consists of sets of the form $\{(x_1, C_1), \ldots, (x_k, C_k)\}$ where $x_1, \ldots, x_k \in X$ and $C_1, \ldots, C_k \subseteq X$. We think of $C_1, \ldots, C_k$ as the clusters, and x_i as the "center" of C_i. In this general framework we do not require that the clusters cover X, or that they are pairwise disjoint, or that they contain their centers. Thus the space of possible clustering solutions is $\mathcal{P} = 2^{X \times 2^X}$ (though the exact structure of $\mathcal{P}$ does not play a significant role in the proof of Theorem 2). Assume that for every point $x \in X$, every metric $d \in \mathrm{MET}(X)$, and every possible clustering solution $P \in \mathcal{P}$, we are given $\mathit{\Gamma}(x, d, P) \in [0, \infty]$, which we think of as a measure of the dissatisfaction of x with respect to P and d. Our goal is to minimize the average dissatisfaction of the points of X. Formally, given a measure of dissatisfaction (which we also call in what follows a *clustering cost function*) $\mathit{\Gamma} : X \times \mathrm{MET}(X) \times \mathcal{P} \to [0, \infty]$, we wish to compute for a given metric $d \in \mathrm{MET}(X)$ the value

$$\mathrm{Opt}_{\mathit{\Gamma}}(X, d) \stackrel{\mathrm{def}}{=} \min \left\{ \sum_{x \in X} \mathit{\Gamma}(x, d, P) : P \in \mathcal{P} \right\}$$

(Since we are mainly concerned with the algorithmic aspect of this problem, we assume from now on that Γ can be computed efficiently.)

We make two natural assumptions on the cost function Γ. First of all, we will assume that it scales homogeneously with respect to the metric, i.e. for every $\lambda > 0$, $x \in X$, $d \in \mathrm{MET}(X)$ and $P \in \mathcal{P}$ we have $\Gamma(x, \lambda d, P) = \lambda\Gamma(x, d, P)$. Secondly we will assume that Γ is monotone with respecting to the metric, i.e. if $d, \overline{d} \in \mathrm{MET}(X)$ and $x \in X$ satisfy $d(x, y) \le \overline{d}(x, y)$ for every $y \in X$ then $\Gamma(x, d, P) \le \Gamma(x, \overline{d}, P)$. In other words, if all the points in X are further with respect to $\overline{d}$ from x then they are with respect to d, then x is more dissatisfied. This is a very natural assumption to make, as most clustering problems look for clusters which are small in various (metric) senses. We call clustering problems with Γ satisfying these assumptions *monotone clustering problems*. A large part of the clustering problems that have been considered in the literature fall into this framework.

The following theorem is a simple application of Theorem 1. It shows that it is enough to solve monotone clustering problems on ultrametrics, with only a polylogarithmic loss in the approximation factor.

Theorem 2 (Reduction to ultrametrics). *Let X be an n-point set and fix a homogeneous monotone clustering cost function $\Gamma : X \times \mathrm{MET}(X) \times \mathcal{P} \to [0, \infty]$. Assume that there is a randomized polynomial time algorithm which approximates $\mathrm{Opt}_\Gamma(X, \rho)$ to within a factor $\alpha(n)$ on any ultrametric $\rho \in \mathrm{MET}(X)$. Then there is a polynomial time randomized algorithm which approximates $\mathrm{Opt}_\Gamma(X, d)$ on any metric $d \in \mathrm{MET}(X)$ to within a factor of $O\left(\alpha(n)(\log n)^2\right)$.*

Proof. Let (X, d) be an n-point metric space and let $\mathscr{D}$ be the distribution over random ultrametrics ρ on X from Theorem 1 (which is computable in polynomial time, as follows directly from our proof of Theorem 1). In other words, $\rho(x, y) \ge d(x, y)$ for all $x, y \in X$ and

$$\max_{x \in X} \mathbb{E}_{\mathscr{D}} \max_{y \in X \setminus \{x\}} \frac{\rho(x, y)}{d(x, y)} \le C(\log n)^2.$$

Let $P \in \mathcal{P}$ be a clustering solution for which $\mathrm{Opt}_\Gamma(X, d) = \sum_{x \in X} \Gamma(x, d, P)$. Using the monotonicity and homogeneity of Γ we see that

$$\begin{aligned}
\mathbb{E}_{\mathscr{D}} \mathrm{Opt}_\Gamma(X, \rho) &\le \mathbb{E}_{\mathscr{D}} \sum_{x \in X} \Gamma(x, \rho, P) && (P \text{ is sub-optimal in } \rho) \\
&\le \mathbb{E}_D \sum_{x \in X} \Gamma\left(x, \left[\max_{y \in X \setminus \{x\}} \frac{\rho(x, y)}{d(x, y)}\right] \cdot d, P\right) && (\text{Monotonicity of } \Gamma) \\
&= \mathbb{E}_{\mathscr{D}} \sum_{x \in X} \left[\max_{y \in X \setminus \{x\}} \frac{\rho(x, y)}{d(x, y)}\right] \cdot \Gamma(x, d, P) && (\text{Homogeneity of } \Gamma) \\
&= \sum_{x \in X} \left(\mathbb{E}_{\mathscr{D}} \left[\max_{y \in X \setminus \{x\}} \frac{\rho(x, y)}{d(x, y)}\right]\right) \cdot \Gamma(x, d, P) \\
&\le C(\log n)^2 \cdot \mathrm{Opt}_\Gamma(X, d) && (\text{Theorem 1}).
\end{aligned}$$

Hence, with probability at least $\frac{1}{2}$, $\mathrm{Opt}_{\Gamma}(X,\rho) \le 2C(\log n)^2 \cdot \mathrm{Opt}_{\Gamma}(X,d)$. For such ρ compute a clustering solution $Q \in \mathcal{P}$ satisfying

$$\sum_{x\in X} \Gamma(x,\rho,Q) \le \alpha(n)\mathrm{Opt}_{\Gamma}(X,\rho) \le 2C\alpha(n)(\log n)^2 \cdot \mathrm{Opt}_{\Gamma}(X,d).$$

Since $\rho \ge d$ it remains to use the monotonicity of Γ once more to deduce that

$$\sum_{x\in X} \Gamma(x,\rho,Q) \ge \sum_{x\in X} \Gamma(x,d,Q) \ge \mathrm{Opt}_{\Gamma}(X,d).$$

Thus Q is a $O\left(\alpha(n)(\log n)^2\right)$ approximate solution to the clustering problem on X with cost Γ. □

Due to Theorem 2 we see that the main difficulty in monotone clustering problems lies in the design of good approximation algorithms for them on ultrametrics. This is a generic reduction, and in many particular cases it might be possible use a case-specific analysis to improve the $O\left((\log n)^2\right)$ loss in the approximation factor. However, as a general reduction paradigm for clustering problems, Theorem 2 makes it clear why maximum gradient embeddings are so natural.

We next demonstrate the applicability of the monotone clustering framework to a concrete example called *fault-tolerant k-median*. In the full version of the paper, we analyze another clustering problem, called $\Sigma\ell_p$ clustering.

Fault-tolerant k-median. Fix $k \in \mathbb{N}$. The k-median problem is as follows. Given an n-point metric space (X, d_X), find $x_1, \ldots, x_k \in X$ that minimize the objective function

$$\sum_{x\in X} \min_{j\in\{x_1,\ldots,x_k\}} d_X(x,x_j). \tag{3}$$

This very natural and well studied problem can be easily cast as monotone clustering problem by defining $\Gamma(x,d,\{(x_1,C_1),\ldots,(x_m,C_m)\})$ to be ∞ if $m \ne k$, and otherwise $\Gamma(x,d,\{(x_1,C_1),\ldots,(x_m,C_m)\}) = \min_{j\in\{x_1,\ldots,x_k\}} d(x,x_j)$.

The linear structure of (3) makes it a prime example of a problem which can be approximated using Bartal's probabilistic embeddings. Indeed, the first non-trivial approximation algorithm for k-median clustering was obtained by Bartal in [4]. Since then this problem has been investigated extensively: The first constant factor approximation for it was obtained in [9] using LP rounding, and the first combinatorial (primal-dual) constant-factor algorithm was obtained in [15]. In [2] an analysis of a natural local search heuristic yields the best known approximation factor for k-median clustering.

Here we study the following fault-tolerant version of the k-median problem. Let (X, d) be an n-point metric space and fix $k \in \mathbb{N}$. Assume that for every $x \in X$ we are given an integer $j(x) \in X$ (which we call the fault-tolerant parameter of x). Given $x_1, \ldots, x_k$ and $x \in X$ let $x_j^*(x; d)$ be the j-th closest point to x in $\{x_1, \ldots, x_k\}$. In other words, $\{x_j^*(x; d)\}_{j=1}^k$ is a re-ordering of $\{x_j\}_{j=1}^k$ such that $d(x, x_1^*(x; d)) \le \cdots \le d(x, x_k^*(x; d))$. Our goal is to minimize the objective function

$$\sum_{x\in X} d\left(x, x_{j(x)}^*(x; d)\right). \tag{4}$$

To understand (4) assume for the sake of simplicity that $j(x) = j$ for all $x \in X$. If $\{x_j\}_{j=1}^k$ minimizes (4) and $j-1$ of them are corrupted (due to possible noise), then the optimum value of (4) does not change. In this sense the clustering problem in (4) is fault-tolerant. In other words, the optimum solution of (4) is insensitive to (controlled) noise. Observe that for $j = 1$ we return to the k-median clustering problem.

We remark that another fault-tolerant version of k-median clustering was introduced in [16]. In this problem we connect each point x in the metric space X to $j(x)$ centers, but the objective function is the sum over $x \in X$ of the sum of the distances from x to all the $j(x)$ centers. Once again, the linearity of the objective function seems to make the problem easier, and in [23] a constant factor approximation is achieved (this immediately implies that our version of fault-tolerant k-median clustering, i.e. the minimization of (4), has a $O\left(\max_{x\in X} j(x)\right)$ approximation algorithm). In particular, the LP that was previously used for k-median clustering naturally generalizes to this setting. This is not the case for our fault-tolerant version in (4). Moreover, the local search techniques for k-median clustering (see for example [2]) do not seem to be easily generalizable to the case $j > 1$, and in any case seem to require $n^{\Omega(j)}$ time, which is not polynomial even for moderate values of j.

Arguing as above in the case of k-median clustering we see that the fault-tolerant k-median clustering problem in (4) is a monotone clustering problem. In the full version of this paper we show that it can be solved exactly in polynomial time on ultrametrics. Thus, in combination with Theorem 2, we obtain a $O\left((\log n)^2\right)$ approximation algorithm for the minimization of (4) on general metrics.

2 Proof of Theorem 1

We begin by sketching the proof of the upper bound in Theorem 1. The full version of this paper has a complete self-contained proof of it.

By the arguments appearing in [13, 18], for every N-point metric spacemetric space (X, d_X) there exist a distribution $\mathscr{D}$ of non-contractive embeddings $f : X \to U$ such that for every $x \in X$, and $t \ge 1$,

$$\Pr_{\mathcal{D}}\left[\exists y \in X, \frac{d_U(f(x), f(y))}{d_X(x, y)} \ge t\right] \le \frac{128 \log_2 n}{t}.$$

By using a known trick one can modify the construction of $\mathscr{D}$ to also satisfy for every $x \in X$,

$$\Pr_{\mathscr{D}}\left[\exists y \in X, \frac{d_U(f(x), f(y))}{d_X(x, y)} \ge 4n\right] = 0.$$

Hence for every $x \in X$,

$$\begin{aligned}\mathbb{E}_{\mathscr{D}}|\nabla f|_\infty &= \mathbb{E}_{\mathscr{D}} \sup_{y \neq x} \frac{d_U(f(x), f(y))}{d_X(x,y)} \\ &\leq \sum_{i=0}^{\infty} 2^{i+1} \cdot \Pr\left[\exists y \in X, \frac{d_U(f(x), f(y))}{d_X(x,y)} \geq 2^i\right] \\ &\leq \sum_{i=0}^{2+\log_2 n} 2^i \cdot \frac{128 \log_2 n}{2^i} = O\left((\log n)^2\right).\end{aligned}$$

We next prove a matching lower bound. As mentioned in the introduction, the metrics Y_n in Theorem 1 are the diamond graphs of Newman and Rabinovich [21], which will be defined presently. Before passing to this more complicated lower bound, we will analyze the simpler example of cycles.

Let C_n, $n > 3$, be the unweighted path on n-vertices. We will identify C_n with the group $\mathbb{Z}_n$ of integers modulo n. We first observe that in this special case the upper bound in Theorem 1 can be improved to $O(\log n)$. This is achieved by using Karp's embedding of the cycle into spanning paths— we simply choose an edge of C_n uniformly at random and delete it. Let $f : C_n \to \mathbb{Z}$ be the randomized embedding thus obtained, which is clearly non-contractive.

Karp noted that it is easy to see that as a probabilistic embedding into trees f has distortion at most 2. We will now show that as a maximum gradient embedding, f has distortion $\Theta(\log n)$. Indeed, fix $x \in C_n$, and denote the deleted edge by $\{a, a+1\}$. Assume that $d_{C_n}(x,a) = t \leq n/2 - 1$. Then the distance from $a+1$ to x changed from $t+1$ in C_n to $n-t-1$ in the path. It is also easy to see that this is where the maximum gradient is attained. Thus

$$\mathbb{E}|\nabla f(x)|_\infty \approx \frac{2}{n} \sum_{0 \leq t \leq n/2} \frac{n-t-1}{t+1} = \Theta(\log n).$$

We will now show that any maximum gradient embedding of C_n into a distribution over trees incurs distortion $\Omega(\log n)$. For this purpose we will use the following lemma from [22].

Lemma 1. *For any tree metric T, and any non-contractive embedding $g : C_n \to T$, there exists an edge $(x, x+1)$ of C_n such that $d_T(g(x), g(x+1)) \geq \frac{n}{3} - 1$.*

Now, let $\mathscr{D}$ be a distribution over non-contractive embeddings of C_n into trees $f : C_n \to T$. By Lemma 1 we know that there exists $x \in C_n$ such that $d_T(f(x), f(x+1)) \geq \frac{n-3}{3}$. Thus for every $y \in C_n$ we have that

$$\max\{d_T(f(y), f(x)), d_T(f(y), f(x+1))\} \geq \frac{n-3}{6}.$$

On the other hand $\max\{d_{C_n}(y,x), d_{C_n}(y, x+1)\} \leq d_{C_n}(x,y) + 1$. It follows that,

$$|\nabla f(y)|_\infty \geq \frac{n-3}{6 d_{C_n}(x,y) + 6}.$$

Summing this inequality over $y \in C_n$ we see that

$$\sum_{y \in C_n} |\nabla f(y)|_\infty \geq \sum_{0 \leq k \leq n/2} \frac{n-3}{6k+6} = \Omega(n \log n).$$

Thus

$$\max_{y \in C_n} \mathbb{E}_{\mathscr{D}} |\nabla f(y)|_\infty \geq \frac{1}{n} \sum_{y \in C_n} \mathbb{E}_{\mathscr{D}} |\nabla f(y)|_\infty = \Omega(\log n),$$

as required.

We now pass to the proof of the lower bound in Theorem 1. We start by describing the diamond graphs $\{G_k\}_{k=1}^\infty$, and a special labelling of them that we will use throughout the ensuing arguments. The first diamond graph G_1 is a cycle of length 4, and G_{k+1} is obtained from G_k by replacing each edge by a quadrilateral. Thus G_k has 4^k edges and $\frac{2 \cdot 4^k + 4}{3}$ vertices. As we have done before, the required lower bound on maximum gradient embeddings of G_k into trees will be proved if we show that for every tree T and every non-contractive embedding $f : G_k \to T$ we have

$$\frac{1}{4^k} \sum_{e \in E(G_k)} \sum_{x \in e} |\nabla f(x)|_\infty = \Omega\left(k^2\right). \tag{5}$$

Note that the inequality (5) is different from the inequality that we proved in the case of the cycle in that the weighting on the vertices of G_k that it induces is not uniform— high degree vertices get more weight in the average in the left-hand side of (5).

We will prove (5) by induction on k. In order to facilitate such an induction, we will first strengthen the inductive hypothesis. To this end we need to introduce a useful labelling of G_k. For $1 \leq i \leq k$ the graph G_k contains 4^{k-i} canonical copies of G_i, which we index by elements of $\{1,2,3,4\}^{k-i}$, and denote $\left\{G_{[\alpha]}^{(k)}\right\}_{\alpha \in \{1,2,3,4\}^{k-i}}$. These graphs are defined as follows. For $k = 2$ they are shown in Figure 1.

For $k = 3$ these canonical subgraphs are shown in Figure 2.

Formally, we set $G_{[\emptyset]}^{(k)} = G_k$, and assume inductively that the canonical subgraphs of G_{k-1} have been defined. Let H_1, H_2, H_3, H_4 be the top-right, top-left, bottom-right and bottom-left copies of G_{k-1} in G_k, respectively. For $\alpha \in \{1,2,3,4\}^{k-1-i}$ and $j \in \{1,2,3,4\}$ we denote the copy of G_i in H_j corresponding to $G_{[\alpha]}^{(k-1)}$ by $G_{[j\alpha]}^{(k)}$.

For every $1 \leq i \leq k$ and $\alpha \in \{1,2,3,4\}^{k-i}$ let $T_{[\alpha]}^{(k)}, B_{[\alpha]}^{(k)}, L_{[\alpha]}^{(k)}, R_{[\alpha]}^{(k)}$ be the topmost, bottom-most, left-most, and right-most vertices of $G_{[\alpha]}^{(k)}$, respectively. Fixing an embedding $f : G_k \to T$, we will construct inductively a set of simple cycles $\mathscr{C}_{[\alpha]}$ in $G_{[\alpha]}^{(k)}$ and for each $C \in \mathscr{C}_{[\alpha]}$ an edge $\varepsilon_C \in E\left(\mathscr{C}_{[\alpha]}\right)$, with the following properties.

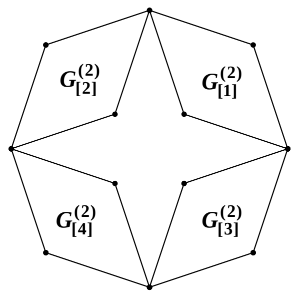

Fig. 1. The graph G_2 and the labelling of the canonical copies of G_1 contained in it.

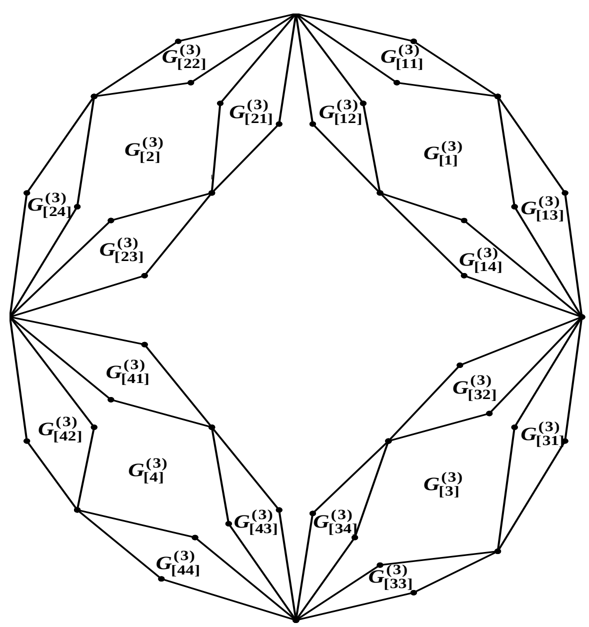

Fig. 2. The graph G_3 and the induced labeling of canonical copies of G_1 and G_2.

1. The cycles in $\mathscr{C}_{[\alpha]}$ are edge-disjoint, and they all pass through the vertices $T^{(k)}_{[\alpha]}, B^{(k)}_{[\alpha]}, L^{(k)}_{[\alpha]}, R^{(k)}_{[\alpha]}$. There are 2^{i-1} cycles in $\mathscr{C}_{[\alpha]}$, and each of them contains 2^{i+1} edges. Thus in particular the cycles in $\mathscr{C}_{[\alpha]}$ form a disjoint cover of the edges in $G^{(k)}_{[\alpha]}$.
2. If $C \in \mathscr{C}_{[\alpha]}$ and $\varepsilon_C = \{x, y\}$ then $d_T(f(x), f(y)) \geq \frac{2^{i+1}}{3} - 1$.
3. Denote $E_{[\alpha]} = \{\varepsilon_C : \ C \in \mathscr{C}_{[\alpha]}\}$ and $\Delta_i = \bigcup_{\alpha\in\{1,2,3,4\}^{k-i}} E_{[\alpha]}$. The edges in Δ_i will be called the *designated edges* of level i. For $\alpha \in \{1, 2, 3, 4\}^{k-i}$, $C \in \mathscr{C}_{[\alpha]}$ and $j < i$ let $\Delta_j(C) = \Delta_j \cap E(C)$ be the designated edges of level j on C. Then we require that each of the two paths $T^{(k)}_{[\alpha]} - L^{(k)}_{[\alpha]} - B^{(k)}_{[\alpha]}$ and $T^{(k)}_{[\alpha]} - R^{(k)}_{[\alpha]} - B^{(k)}_{[\alpha]}$ in C contain exactly 2^{i-j-1} edges from $\Delta_j(C)$.

The construction is done by induction on i. For $i = 1$ and $\alpha \in \{1, 2, 3, 4\}^{k-1}$ we let $\mathscr{C}_{[\alpha]}$ contain only the 4-cycle $G^{(k)}_{[\alpha]}$ itself. Moreover by Lemma 1 there is and edge $\varepsilon_{G^{(k)}_{[\alpha]}} \in E\left(G^{(k)}_{[\alpha]}\right)$ such that if $\varepsilon_{G^{(k)}_{[\alpha]}} = \{x, y\}$ then $d_T(f(x), f(y)) \geq \frac{1}{3}$. This completes the construction for $i = 1$. Assuming we have completed the construction for $i-1$ we construct the cycles at level i as follows. Fix arbitrary cycles $C_1 \in \mathscr{C}_{[1\alpha]}$, $C_2 \in \mathscr{C}_{[2\alpha]}$, $C_3 \in \mathscr{C}_{[3\alpha]}$, $C_4 \in \mathscr{C}_{[4\alpha]}$. We will use these four cycles to construct two cycles in $\mathscr{C}_{[\alpha]}$. The first one consists of the $T^{(k)}_{[\alpha]} - R^{(k)}_{[\alpha]}$ path in C_1 which contains the edge ε_{C_1}, the $R^{(k)}_{[\alpha]} - B^{(k)}_{[\alpha]}$ path in C_3 which does not contain the edge ε_{C_3}, the $B^{(k)}_{[\alpha]} - L^{(k)}_{[\alpha]}$ path in C_4 which contains the edge ε_{C_4}, and the $L^{(k)}_{[\alpha]} - T^{(k)}_{[\alpha]}$ path in C_2 which does not contain the edge ε_{C_2}. The remaining edges in $E(C_1) \cup E(C_2) \cup E(C_3) \cup E(C_4)$ constitute the second cycle that we extract from C_1, C_2, C_3, C_4. Continuing in this manner by choosing cycles from $\mathscr{C}_{[1\alpha]} \setminus \{C_1\}$, $\mathscr{C}_{[2\alpha]} \setminus \{C_2\}$, $\mathscr{C}_{[3\alpha]} \setminus \{C_3\}$, $\mathscr{C}_{[4\alpha]} \setminus \{C_4\}$ and repeating this procedure, and then continuing until we exhaust the cycles in $\mathscr{C}_{[1\alpha]} \cup \mathscr{C}_{[2\alpha]} \cup \mathscr{C}_{[3\alpha]} \cup \mathscr{C}_{[4\alpha]}$, we obtain the set of cycles $\mathscr{C}_{[\alpha]}$. For every $C \in \mathscr{C}_{[alpha]}$ we then apply Lemma 1 to obtain an edge ε_C with the required property.

For each edge $e \in E(G_k)$ let $\alpha \in \{1, 2, 3, 4\}^{k-i}$ be the unique multi-index such that $e \in E\left(G^{(k)}_{[\alpha]}\right)$. We denote by $C_i(e)$ the unique cycle in $\mathscr{C}_{[\alpha]}$ containing e. We will also denote $\widehat{e}_i(e) = \varepsilon_{C_i(e)}$. Finally we let $a_i(e) \in e$ and $b_i(e) \in \widehat{e}_i(e)$ be vertices such that

$$d_T(f(a_i(e)), f(b_i(e))) = \max_{\substack{a\in e\\ b\in \widehat{e}_i(e)}} d_T(f(a), f(b)).$$

Note that by the definition of $\widehat{e}_i(e)$ and the triangle inequality we are assured that

$$d_T(f(a_i(e)), f(b_i(e))) \geq \frac{1}{2}\left(\frac{2^{i+1}}{3} - 1\right) \geq \frac{2^i}{12}. \tag{6}$$

Recall that we plan to prove (5) by induction on k. Having done all of the above preparation, we are now in position to strengthen (5) so as to make the inductive argument easier. Given two edges $e, h \in G_k$ we write $e \frown_i h$ if both e, h are on the same canonical copy of G_i in G_k, $C_i(e) = C_i(h) = C$, and furthermore e and h on the same side of C. In other words, $e \frown_i h$ if there is $\alpha \in \{1,2,3,4\}^{k-i}$ and $C \in \mathscr{C}_{[\alpha]}$ such that if we partition the edges of C into two disjoint $T_{[\alpha]}^{(k)} - B_{[\alpha]}^{(k)}$ paths, then e and h are on the same path.

Let $m \in \mathbb{N}$ be a universal constant that will be specified later. For every integer $\ell \le k/m$ and any $\alpha \in \{1,2,3,4\}^{k-m\ell}$ define

$$L_\ell(\alpha) = \frac{1}{4^{m\ell}} \sum_{e \in E\left(G_{[\alpha]}^{(k)}\right)} \max_{\substack{i \in \{1,\ldots,\ell\} \\ e \frown_{im} \widehat{e}_{im}(e)}} \frac{d_T(f(a_{im}(e)), f(b_{im}(e))) \wedge 2^{im}}{d_{G_k}(e, \widehat{e}_{im}(e)) + 1}.$$

We also write $L_\ell = \min_{\alpha \in \{1,2,3,4\}^{k-m\ell}} L_\ell(\alpha)$. We will prove that $L_\ell \ge L_{\ell-1} + c\ell$, where $c > 0$ is a universal constant. This will imply that for $\ell = \lfloor k/m \rfloor$ we have $L_\ell = \Omega(k^2)$ (since m is a universal constant). By simple arithmetic (5) follows.

Observe that for every $\alpha \in \{1,2,3,4\}^{k-m\ell}$ we have

$$\begin{aligned}
L_\ell(\alpha) &= 4^{-m} \sum_{\beta \in \{1,2,3,4\}^m} 4^{-m(\ell-1)} \sum_{e \in E\left(G_{[\beta\alpha]}^{(k)}\right)} \max_{\substack{i \in \{1,\ldots,\ell\} \\ e \frown_{im} \widehat{e}_{im}(e)}} \frac{d_T(f(a_{im}(e)), f(b_{im}(e))) \wedge 2^{im}}{d_{G_k}(e, \widehat{e}_{im}(e)) + 1} \\
&= 4^{-m} \sum_{\beta \in \{1,2,3,4\}^m} 4^{-m(\ell-1)} \sum_{e \in E\left(G_{[\beta\alpha]}^{(k)}\right)} \max_{\substack{i \in \{1,\ldots,\ell-1\} \\ e \frown_{im} \widehat{e}_{im}(e)}} \frac{d_T(f(a_{im}(e)), f(b_{im}(e))) \wedge 2^{im}}{d_{G_k}(e, \widehat{e}_{im}(e)) + 1} \\
&\quad + \frac{1}{4^{m\ell}} \sum_{e \in E\left(G_{[\alpha]}^{(k)}\right)} \max\left\{0, \frac{d_T(f(a_{\ell m}(e)), f(b_{\ell m}(e))) \wedge 2^{\ell m}}{d_{G_k}(e, \widehat{e}_{\ell m}(e)) + 1} \cdot \mathbf{1}_{\{e \frown_{\ell m} \widehat{e}_{\ell m}(e)\}} \right. \\
&\qquad\qquad \left. - \max_{\substack{i \in \{1,\ldots,\ell-1\} \\ e \frown_{im} \widehat{e}_{im}(e)}} \frac{d_T(f(a_{im}(e)), f(b_{im}(e))) \wedge 2^{im}}{d_{G_k}(e, \widehat{e}_{im}(e)) + 1} \right\} \\
&= \frac{1}{4^m} \sum_{\beta \in \{1,2,3,4\}^m} L_{\ell-1}(\beta\alpha) \\
&\quad + \frac{1}{4^{m\ell}} \sum_{e \in E\left(G_{[\alpha]}^{(k)}\right)} \max\left\{0, \frac{d_T(f(a_{\ell m}(e)), f(b_{\ell m}(e))) \wedge 2^{\ell m}}{d_{G_k}(e, \widehat{e}_{\ell m}(e)) + 1} \cdot \mathbf{1}_{\{e \frown_{\ell m} \widehat{e}_{\ell m}(e)\}} \right. \\
&\qquad\qquad \left. - \max_{\substack{i \in \{1,\ldots,\ell-1\} \\ e \frown_{im} \widehat{e}_{im}(e)}} \frac{d_T(f(a_{im}(e)), f(b_{im}(e))) \wedge 2^{im}}{d_{G_k}(e, \widehat{e}_{im}(e)) + 1} \right\}
\end{aligned}$$

$$\geq L_{\ell-1}$$

$$+\frac{1}{4^{m\ell}}\sum_{e\in E\left(G_{[\alpha]}^{(k)}\right)}\max\left\{0,\frac{d_T(f(a_{\ell m}(e)),f(b_{\ell m}(e)))\wedge 2^{\ell m}}{d_{G_k}(e,\widehat{e}_{\ell m}(e))+1}\cdot\mathbf{1}_{\{e\frown_{\ell m}\widehat{e}_{\ell m}(e)\}}\right.$$

$$\left.-\max_{\substack{i\in\{1,\ldots,\ell-1\}\\ e\frown_{im}\widehat{e}_{im}(e)}}\frac{d_T(f(a_{im}(e)),f(b_{im}(e)))\wedge 2^{im}}{d_{G_k}(e,\widehat{e}_{im}(e))+1}\right\}.$$

Thus it is enough to show that

$$A\stackrel{\mathrm{def}}{=}4^{-m\ell}\sum_{e\in E\left(G_{[\alpha]}^{(k)}\right)}\max\left\{0,\frac{d_T(f(a_{\ell m}(e)),f(b_{\ell m}(e)))\wedge 2^{\ell m}}{d_{G_k}(e,\widehat{e}_{\ell m}(e))+1}\cdot\mathbf{1}_{\{e\frown_{\ell m}\widehat{e}_{\ell m}(e)\}}\right.$$

$$\left.-\max_{\substack{i\in\{1,\ldots,\ell-1\}\\ e\frown_{im}\widehat{e}_{im}(e)}}\frac{d_T(f(a_{im}(e)),f(b_{im}(e)))\wedge 2^{im}}{d_{G_k}(e,\widehat{e}_{im}(e))+1}\right\}=\Omega(\ell). \quad (7)$$

To prove (7), denote for $C\in\mathscr{C}_{[\alpha]}$

$$S_C=\left\{e\in E(C):\ \varepsilon_C\frown_{\ell m}e\ \text{ and }\right.$$

$$\max_{\substack{i\in\{1,\ldots,\ell-1\}\\ e\frown_{im}\widehat{e}_{im}(e)}}\frac{d_T(f(a_{im}(e)),f(b_{im}(e)))\wedge 2^{im}}{d_{G_k}(e,\widehat{e}_{im}(e))+1}$$

$$\left.\geq\frac{1}{2}\cdot\frac{d_T(f(a_{\ell m}(e)),f(b_{\ell m}(e)))\wedge 2^{\ell m}}{d_{G_k}(e,\widehat{e}_{\ell m}(e))+1}\right\}.$$

Then using (6) we see that

$$A\geq\frac{1}{2\cdot 4^{m\ell}}\sum_{C\in\mathscr{C}_{[\alpha]}}\sum_{\substack{e\in E(C)\setminus S_C\\ \varepsilon_C\frown_{\ell m}e}}\frac{d_T(f(a_{\ell m}(e)),f(b_{\ell m}(e)))\wedge 2^{\ell m}}{d_{G_k}(e,\widehat{e}_{\ell m}(e))+1}$$

$$\geq\frac{1}{2\cdot 4^{m\ell}}\sum_{C\in\mathscr{C}_{[\alpha]}}\sum_{\substack{e\in E(C)\\ \varepsilon_C\frown_{\ell m}e}}\frac{d_T(f(a_{\ell m}(e)),f(b_{\ell m}(e)))\wedge 2^{\ell m}}{d_{G_k}(e,\widehat{e}_{\ell m}(e))+1}$$

$$-\frac{1}{2\cdot 4^{m\ell}}\sum_{C\in\mathscr{C}_{[\alpha]}}\sum_{e\in S_C}\frac{d_T(f(a_{\ell m}(e)),f(b_{\ell m}(e)))\wedge 2^{\ell m}}{d_{G_k}(e,\widehat{e}_{\ell m}(e))+1}$$

$$\geq\frac{1}{2\cdot 4^{m\ell}}\sum_{C\in\mathscr{C}_{[\alpha]}}\sum_{i=1}^{2^{m\ell-1}}\frac{2^{m\ell}}{12i}-\frac{1}{2\cdot 4^{m\ell}}\sum_{C\in\mathscr{C}_{[\alpha]}}\sum_{e\in S_C}\frac{d_T(f(a_{\ell m}(e)),f(b_{\ell m}(e)))\wedge 2^{\ell m}}{d_{G_k}(e,\widehat{e}_{\ell m}(e))+1}$$

$$= \Omega\left(\frac{1}{4^{m\ell}} \cdot |\mathscr{C}_{[\alpha]}| \cdot 2^{m\ell} \cdot m\ell\right)$$
$$- \frac{1}{2 \cdot 4^{m\ell}} \sum_{C \in \mathscr{C}_{[\alpha]}} \sum_{e \in S_C} \frac{d_T(f(a_{\ell m}(e)), f(b_{\ell m}(e))) \wedge 2^{\ell m}}{d_{G_k}(e, \widehat{e}_{\ell m}(e)) + 1}$$
$$= \Omega(m\ell) - \frac{1}{2 \cdot 4^{m\ell}} \sum_{C \in \mathscr{C}_{[\alpha]}} \sum_{e \in S_C} \frac{d_T(f(a_{\ell m}(e)), f(b_{\ell m}(e))) \wedge 2^{\ell m}}{d_{G_k}(e, \widehat{e}_{\ell m}(e)) + 1}. \tag{8}$$

To estimate the negative term in (8) fix $C \in \mathscr{C}_{[\alpha]}$. For every edge $e \in S_C$ (which implies in particular that $\widehat{e}_{\ell m}(e) = \varepsilon_C$) we fix an integer $i < \ell$ such that $e \frown_{im} \widehat{e}_{im}(e)$ and

$$\begin{aligned} \frac{2^{im}}{d_{G_k}(e, \widehat{e}_{im}(e)) + 1} &\geq \frac{d_T(f(a_{im}(e)), f(b_{im}(e))) \wedge 2^{im}}{d_{G_k}(e, \widehat{e}_{im}(e)) + 1} \\ &\geq \frac{1}{2} \cdot \frac{d_T(f(a_{\ell m}(e)), f(b_{\ell m}(e))) \wedge 2^{\ell m}}{d_{G_k}(e, \widehat{e}_{\ell m}(e)) + 1} \\ &\geq \frac{1}{12} \cdot \frac{2^{\ell m}}{d_{G_k}(e, \varepsilon_C) + 1}, \end{aligned}$$

or

$$d_{G_k}(e, \widehat{e}_{im}(e)) + 1 \leq 2^{(i-\ell)m+4} \left[d_{G_k}(e, \varepsilon_C) + 1\right]. \tag{9}$$

We shall call the edge $\widehat{e}_{im}(e)$ the designated edge that inserted e into S_C. For a designated edge $\varepsilon \in E(C)$ of level im (i.e. $\varepsilon \in \Delta_{im}(C)$) we shall denote by $\mathscr{E}_C(\varepsilon)$ the set of edges of C which ε inserted to S_C. Denoting $D_\varepsilon = d_{G_k}(\varepsilon, \varepsilon_C) + 1$ we see that (9) implies that for $e \in \mathscr{E}_C(\varepsilon)$ we have

$$\left| D_\varepsilon - \left[d_{G_k}(e, \varepsilon_C) + 1\right] \right| \leq 2^{(i-\ell)m+4} \left[d_{G_k}(e, \varepsilon_C) + 1\right]. \tag{10}$$

Assuming that $m \geq 5$ we are assured that $2^{(i-\ell)m+4} \leq \frac{1}{2}$. Thus (10) implies that

$$\frac{D_\varepsilon}{1 + 2^{(i-\ell)m+4}} \leq d_{G_k}(e, \varepsilon_C) + 1 \leq \frac{D_\varepsilon}{1 - 2^{(i-\ell)m+4}}.$$

Hence

$$\begin{aligned} &\sum_{e \in S_C} \frac{d_T(f(a_{\ell m}(e)), f(b_{\ell m}(e))) \wedge 2^{\ell m}}{d_{G_k}(e, \widehat{e}_{\ell m}(e)) + 1} \\ &\qquad \leq \sum_{i=1}^{\ell-1} \sum_{\varepsilon \in \Delta_{im}(C)} \sum_{e \in \mathscr{E}_C(\varepsilon)} \frac{2^{\ell m}}{d_{G_k}(e, \varepsilon_C) + 1} \\ &\qquad \leq 2 \sum_{i=1}^{\ell-1} \sum_{\varepsilon \in \Delta_{im}(C)} \sum_{\substack{j \in \mathbb{N} \\ \frac{D_\varepsilon}{1+2^{(i-\ell)m+4}} \leq j \leq \frac{D_\varepsilon}{1-2^{(i-\ell)m+4}}}} \frac{2^{\ell m}}{j} \end{aligned}$$

$$
\begin{aligned}
&= O(1) \cdot 2^{\ell m} \sum_{i=1}^{\ell-1} |\Delta_{im}(C)| \cdot \log \left(\frac{1+2^{(i-\ell)m+4}}{1-2^{(i-\ell)m+4}} \right) \\
&= O(1) \cdot 2^{\ell m} \ell \cdot 2^{(\ell-i)m} \cdot 2^{(i-\ell)m} = O(1) \cdot 2^{\ell m} \ell .
\end{aligned}
$$

Thus, using (8) we see that

$$
A = \Omega(m\ell) - O(1) \cdot \frac{1}{4^{\ell m}} \cdot \left|\mathscr{C}_{[\alpha]}\right| 2^{m\ell} \ell = \Omega(m\ell) - O(1)\ell = \Omega(\ell),
$$

provided that m is a large enough absolute constant. This completes the proof of the lower bound in Theorem 1. □

Bibiliography

[1] N. Alon, R. Karp, D. Peleg, and D. West, *A graph-theoretic game and its application to the k-server problem*, SIAM J. Comput. **24** (1995), no. 1.

[2] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit, *Local search heuristics for k-median and facility location problems*, SIAM J. Comput. **33** (2004), no. 3, 544–562.

[3] Y. Bartal, *Probabilistic approximation of metric spaces and its algorithmic applications*, 37th Annual Symposium on Foundations of Computer Science (1996), IEEE Comput. Soc. Press, Los Alamitos, CA, 1996, pp. 184–193.

[4] ———, *On approximating arbitrary metrics by tree metrics*, Proceedings of the 30th Annual ACM Symposium on Theory of Computing, ACM, New York, 1998, pp. 161–168.

[5] Y. Bartal and M. Mendel, *Multi-embedding of metric spaces*, SIAM J. Comput. **34** (2004), no. 1, 248–259.

[6] B. Brinkman and M. Charikar, *On the impossibility of dimension reduction in l_1*, J. ACM **52** (2005), no. 5, 766–788).

[7] S. Buyalo and V. Schroeder, *Embedding of hyperbolic spaces in the product of trees*, Geom. Dedicata **113** (2005), 75–93.

[8] G. Calinescu, H. Karloff, and Y. Rabani, *Approximation algorithms for the 0-extension problem*, SIAM J. Comput. **34** (2004/05), no. 2, 358–372.

[9] M. Charikar, S. Guha, É. Tardos, and D. B. Shmoys, *A constant-factor approximation algorithm for the k-median problem*, J. Comput. System Sci. **65** (2002), no. 1, 129–149.

[10] A. N. Dranishnikov, *On hypersphericity of manifolds with finite asymptotic dimension*, Trans. Amer. Math. Soc. **355** (2003), no. 1, 155–167.

[11] M. Elkin, Y. Emek, D. A. Spielman, and S.-H. Teng, *Lower-stretch spanning trees*, Proceedings of the 37th Annual ACM Symposium on Theory of Computing, ACM, New York, 2005, pp. 494–503.

[12] J. Fakcharoenphol, S. Rao, and K. Talwar, *A tight bound on approximating arbitrary metrics by tree metrics*, Proceedings of the 35th annual ACM Symposium on Theory of Computing, 2003, pp. 448–455.

[13] A. Gupta, M. Hajiaghayi, and H. Räcke, *Oblivious Network Design*, Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms (Miami, FL, 2006), ACM, New York, 2006, pp. 970–979.

[14] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair, *Cuts, trees and l_1-embeddings of graphs*, Combinatorica **24** (2004), no. 2, 233–269.

[15] K. Jain and V. Vazirani, *Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation*, J. ACM **48** (2001), no. 2, 274–296.

[16] K. Jain and V. V. Vazirani, *An approximation algorithm for the fault tolerant metric facility location problem*, Algorithmica **38** (2004), no. 3, 433–439. Approximation algorithms.

[17] J. R. Lee and A. Naor, *Embedding the diamond graph in L_p and dimension reduction in L_1*, Geom. Funct. Anal. **14** (2004), no. 4, 745–747.

[18] M. Mendel and A. Naor, *Ramsey partitions and proximity data structures*, J. European Math. Soc. **9** (2007), no. 2, 253–275, available at http://arxiv.org/cs/0511084.

[19] ———, *Maximum gradient embedding and monotone clustering*, available at http://arxiv.org/cs.DS/0606109.

[20] A. Naor, Y. Peres, O. Schramm, and S. Sheffield, *Markov chains in smooth Banach spaces and Gromov hyperbolic metric spaces*, available at http://arxiv.org/math/0410422.

[21] I. Newman and Y. Rabinovich, *A lower bound on the distortion of embedding planar metrics into Euclidean space*, Discrete Comput. Geom. **29** (2003), no. 1, 77–81.

[22] Y. Rabinovich and R. Raz, *Lower bounds on the distortion of embedding finite metric spaces in graphs*, Discrete Comput. Geom. **19** (1998), no. 1, 79–94.

[23] C. Swamy and D. B. Shmoys, *Fault-tolerant facility location*, Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (2003), ACM, New York, 2003, pp. 735–736.