On Earthmover Distance, Metric Labeling, and 0-Extension *

Howard Karloff[†] Subhash Khot[‡] Aranyak Mehta[§] Yuval Rabani[¶]

Abstract

We study the fundamental classification problems 0-Extension and Metric Labeling. A generalization of Multiway Cut, 0-Extension is closely related to partitioning problems in graph theory and to Lipschitz extensions in Banach spaces; its generalization Metric Labeling is motivated by applications in computer vision. Researchers had proposed using earthmover metrics to get polynomial time-solvable relaxations for these problems. A conjecture that has attracted much attention recently is that the integrality ratio for these relaxations is constant. We prove

1. that the integrality ratio of the earthmover relaxation for Metric Labeling is $\Omega(\log k)$ (which is asymptotically tight), k being the number of labels, whereas the best previous lower bound on the integrality ratio was only constant;
2. that the integrality ratio of the earthmover relaxation for 0-Extension is $\Omega(\sqrt{\log k})$, k being the number of terminals (it was known to be $O((\log k)/\log\log k)$), whereas the best previous lower bound was only constant;
3. that for no $\epsilon > 0$ is there a polynomial-time $O((\log n)^{1/4-\epsilon})$-approximation algorithm for 0-Extension, n being the number of vertices, unless NP$\subseteq$DTIME($n^{\text{poly}(\log n)}$), whereas the strongest inapproximability result known before was only MAX SNP-hardness; and
4. that there is a polynomial-time approximation algorithm for 0-Extension with performance ratio $O(\sqrt{\text{diam}(d)})$, where diam(d) is the ratio of the largest to smallest nonzero distances in the terminal metric.

1 Introduction

Originally suggested by Karzanov [14], 0-Extension takes as input an undirected graph G with a nonnegative weight function w on the edges, a subset T of the node set $V(G)$ (the elements of T being called *terminals*), and a metric d on T. The goal is to assign each node $v \in V(G)$ to a terminal $t(v) \in T$ (with $t(v) = v$ for every $v \in T$), minimizing the total cost of the assignment, which is defined to be $\sum_{\{u,v\}\in E(G)} w(u,v)d(t(u),t(v))$. We are partitioning the graph into $|T|$ pieces, the ith piece containing terminal i, where the cost of sending endpoints u and v of an edge to different terminals *depends on the terminals to which u and v are assigned*. It is the fact that the cost associated with edge $\{u,v\}$ depends on the terminals to which u and v are assigned, and not just on whether $t(u) = t(v)$ or not, that makes it more challenging than easier problems like Multiway Cut. Multiway Cut asks for the minimum cost of partitioning a graph into $|T|$ parts, the ith part including the ith terminal. In other words, Multiway Cut asks for the minimum cost of an assignment of $V(G)$ to T, with $t(v) = v$ for all $v \in T$, of $\sum_{\{u,v\}\in E(G)} w(u,v) \cdot [1 \text{ if } t(u) \neq t(v),\ 0 \text{ otherwise}]$. Thus Multiway Cut is precisely 0-Extension when the metric d is the uniform metric.

0-Extension takes its (unfortunate) name from the fact that we wish to extend the metric d on T to a semimetric on all of $V(G)$ subject to the restriction that every nonterminal must be at distance 0 from some terminal; such extensions are called 0-*extensions*.

*A preliminary version of this paper appeared in the 2006 ACM Symposium on the Theory of Computing.

[†]AT&T Labs—Research, 180 Park Ave., Florham Park, NJ 07932, howard@research.att.com.

[‡]College of Computing, Georgia Institute of Technology, Atlanta, GA 30332–0280, khot@cc.gatech.edu.

[§]IBM Almaden Research Center, San Jose, CA 95120. aranyak@cc.gatech.edu. This work was done while this author was at the Georgia Institute of Technology.

[¶]Computer Science Department, Technion—Israel Institute of Technology, Haifa 32000, Israel, rabani@cs.technion.ac.il.
Part of this work was done while this author was on sabbatical leave at Cornell University. Work supported by Israel Science Foundation grant number 52/03 and by United States-Israel Binational Science Foundation grant number 2002282.

Calinescu et al. [6] gave a $O(\log |T|)$-approximation algorithm for 0-Extension. The better analysis of Fakcharoenphol et al. [10] improved the guarantee to $O(\log |T|/\log\log |T|)$. The underlying idea in [6, 10] is to solve a linear programming relaxation that optimizes over all metric extensions (rather than just 0-extensions), and then to "round" the solution using a new partitioning procedure. Lee and Naor [18] later showed that this partitioning procedure can be used to improve the bounds on Lipschitz extensions in Banach spaces. Krauthgamer, Lee, Mendel, and Naor [15] used the 0-Extension partitioning techniques of [6, 10] in their *measured descent* embedding method.

Metric Labeling takes as input an undirected graph G with a nonnegative weight function w on the edges, a metric space (T, d) (the elements of T are called *labels*), and a nonnegative cost function c on node-label pairs. The goal is to assign, for every node $v \in V(G)$, a label $t(v) \in T$, minimizing the total cost of the assignment, which is $\sum_{v\in V(G)} c(v, t(v)) + \sum_{\{u,v\}\in E(G)} w(u, v)d(t(u), t(v))$. (Notice that this problem generalizes 0-Extension by allowing an arbitrary assignment cost function c.) Motivated by applications to segmentation problems in computer vision, this problem was introduced by Kleinberg and Tardos [17], who proposed an approximation algorithm based on the approximate representation, due to Bartal [3], of (T, d) as a combination of dominating tree metrics. Using the recent improved representation of metrics as combinations of dominating tree metrics due to Fakcharoenphol et al. [11], the Kleinberg-Tardos algorithm guarantees a $O(\log |T|)$ approximation factor, which is the best general result to date. Constant-factor approximations are known for some special cases [17, 12, 8, 1]. (Indeed, the result of [11] was also achieved by modifying the 0-Extension partitioning procedure of [6, 10].)

An obvious question emerges from the above discussion: Can the upper bounds of $O(\log |T|)$ and $O(\log |T|/\log\log |T|)$ for Metric Labeling and 0-Extension, respectively, be improved? As shown above, past experience indicates that pursuing this question may produce results whose impact goes beyond solving the specific optimization problems. Unfortunately, improving the approximation guarantees for these problems is impossible using the methods that were used by the above-mentioned algorithms. Specifically, the bound on embedding a metric into a combination of dominating tree metrics is asymptotically tight (a lower bound follows from [2, 19]), and the diameter-times-boundary volume bound of the partitioning is also tight (proof omitted). The metric relaxation of 0-Extension was shown to have integrality ratio $\Omega\left(\sqrt{\log |T|}\right)$ [6]. (A different earlier construction of Johnson et al. [13] done in the context of Lipschitz extensions implies a somewhat weaker bound.)

A promising direction was suggested independently by Charikar [7] and by Chekuri et al. [8]. They suggested a new linear programming relaxation, motivated by a successful relaxation for the special case Multiway Cut of 0-Extension. The same relaxation, with different objective functions, can be used for 0-Extension and for Metric Labeling. The idea is to find an optimal *transportation metric* extending d (instead of an *arbitrary metric* extending d). It is often called the *earthmover relaxation*. (Transportation metrics are called *earthmover distance* in the computer vision literature, where they are used as a standard metric to compare histograms. In other fields where they are applied, including analysis and information theory, they are often called by other names.) Chekuri et al. [8] showed that the earthmover relaxation for Metric Labeling has integrality ratio at least as good as the performance ratio of the Kleinberg-Tardos algorithm; see [8] for details. Archer et al. [1] gave an earthmover relaxation-based Metric Labeling algorithm whose performance depends on the decomposability of the metric d. Furthermore, the previously known bad examples for the metric relaxation for 0-Extension actually have only *constant* integrality ratio in the earthmover relaxation (proofs omitted). Despite these positive indications and significant attention, no progress has been reported on improving the upper bounds in the general case for either 0-Extension or Metric Labeling. In fact, Chuzhoy and Naor [9] recently published a disturbing result. They proved that unless NP $\subseteq$ DTIME $\left(n^{poly(\log n)}\right)$, there is no polynomial-time algorithm that approximates Metric Labeling within a factor of $O\left((\log |T|)^{\frac{1}{2}-\epsilon}\right)$, for any $\epsilon > 0$. Their result does not apply to 0-Extension.

In this paper we resolve many of the questions mentioned above. In Section 3 we prove an $\Omega(\log |T|)$ integrality ratio for the earthmover relaxation for Metric Labeling, in contrast to the previous constant lower bound. In view of the known upper bounds [8], this result is asymptotically tight. In Section 4 we prove an $\Omega\left(\sqrt{\log |T|}\right)$ integrality ratio for the earthmover relaxation for 0-Extension. This matches the lower bound known for the metric relaxation. We also provide in the appendix, an alternate construction to prove both these integrality ratios.

(A result of Bourgain [4] implies that the transportation metric over a Hamming cube cannot be embedded into a convex combination of 0-extensions of the Hamming cube with distortion which is bounded by an absolute constant as the cube dimension increases. The bound was improved significantly (and stated explicitly) in a very recent paper of Khot and Naor [16]. However, this does not imply an integrality ratio, as we are interested in the Lipschitz constant of the embedding rather than the product of the Lipschitz constants of the embedding and its inverse. In fact, when d is a Hamming cube metric, the earthmover relaxation gives an *optimal* integral solution value (proof omitted)!)

In Section 5 we prove that unless NP $\subseteq$ DTIME $\left(n^{poly(\log n)}\right)$, there is no polynomial-time algorithm that approximates 0-Extension within a factor of $O\left((\log n)^{\frac{1}{4}-\epsilon}\right)$, $n = |V|$, for any $\epsilon > 0$. On a more optimistic note, in Section 6 we give an algorithm for rounding the earthmover solution for 0-Extension that guarantees a $O\left(\sqrt{\operatorname{diam}(d)}\right)$ approximation. Such a bound is not known for the metric relaxation. In the appendix we give alternative constructions, based on linear codes, of Metric Labeling and 0-Extension instances with large integrality ratios. These constructions were inspired by the results in [16].

Through this work we develop new techniques for analyzing transportation metrics, which we hope will find further use in the numerous areas in which earthmover metrics are applied.

2 Preliminaries

We often use k to denote $|T|$. For $v \in V(G)$ let $N(v)$ denote the set of neighbors of v. Informally, the earthmover relaxation for 0-Extension assigns to each $v \in V(G)$ a probability distribution x^v over the set of terminals. In other words, $x^v \in \mathbb{R}^k$ is a nonnegative vector with $\|x^v\|_1 = 1$. An edge $\{u, v\} \in E(G)$ gets stretched by the minimum cost of transporting mass to convert x^u into x^v (or vice versa), where the cost of transporting a unit of mass from terminal i to terminal j is $d(i, j)$. This is simply a flow computation, the vector $f^{uv} \in \mathbb{R}^{k\times k}$ denoting this flow. Formally, the relaxation is the following linear program.

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2}\sum_{u\in V}\sum_{v\in N(u)} w(u,v)\cdot\left[\sum_{i\in T}\sum_{j\in T} d(i,j)(f^{uv})_{ij}\right] \\ \text{such that} & (x_j^u - x_j^v) + \sum_{i\in T}\left((f^{uv})_{ij} - (f^{uv})_{ji}\right) = 0 \quad \forall u\in V, \forall v\in N(u), \forall j\in T \\ & \sum_{j\in T} x_j^u = 1 \quad \forall u \in V \\ & x, f \geq 0, \end{array}$$

where we put $x_j^j = 1$ for all $j \in T$. The earthmover relaxation for Metric Labeling is identical, except the objective function has an additional term of $\sum_{v\in V}\sum_{i\in T} c(v,i)x_i^v$ and we don't put $x_j^j = 1$.

3 Integrality Ratio for Metric Labeling

Consider an infinite family of (bounded-degree) expanders. Let H be a member of this family and let k be the number of nodes in H. We define the following instance of Metric Labeling:

The label set T is the set $V(H)$ of vertices of H. The metric on the label set is the shortest path metric of H. The input graph G has $V(G) = \{\{i,j\} : i,j \in T, i \neq j\}$ and $E(G) = \{\{\{i,j\},\{i,j'\}\} : \{j,j'\} \in E(H)\}$. All edges have weight 1. The cost of assigning a label t to a node $\{i,j\}$ is 0 if $t \in \{i,j\}$ and ∞ otherwise.

Consider the fractional solution that assigns to every node $\{i,j\}$ a vector $x^{\{i,j\}}$ where $x_i^{\{i,j\}} = x_j^{\{i,j\}} = \frac{1}{2}$, and the other entries are 0. Notice that the length of every edge in $E(G)$ is exactly $\frac{1}{2}$, so the cost of this feasible solution is $|E(G)|/2 = k|E(H)|/2$.

To bound the cost of an integral solution we need the following lemma.

Lemma 1 *Consider a tournament over k nodes. At least half the nodes have both their indegree and their outdegree between $k/8$ and $7k/8$.*

Proof: Suppose there are m nodes with indegree exceeding $7k/8$. In the subtournament on these m nodes there must be a node with indegree at most $m/2$, since the sum of the indegrees is $\binom{m}{2}$. This node's total indegree, which exceeds $7k/8$, is at most $k - m + m/2 = k - m/2$. Thus $m/2 < k/8$ and hence $m < k/4$.

The same argument can be applied to outdegree. □

Theorem 2 *Any integral solution to the above instance has cost $\Omega(|E(G)| \log k)$. Thus, the integrality ratio for* Metric Labeling *is $\Omega(\log k)$.*

Proof: An integral solution must assign to a node $\{i, j\}$ either label i or label j. Consider the tournament on the label set T where there is an arc (i, j) if $\{i, j\}$ is assigned to j and the reverse arc otherwise. Call a label *balanced* if and only if both its indegree and its outdegree in the tournament are between $k/8$ and $7k/8$. By Lemma 1, at least half the labels are balanced.

Let t be a balanced label. Put $I = \{i : \{t, i\} \text{ is assigned } t\}$ and $J = \{j : \{t, j\} \text{ is assigned } j\}$. By definition we have $k/8 \leq |I|, |J| \leq 7k/8$. Therefore, there are at least ck expander edges $\{i, j\}$ for which $i \in I$ and $j \in J$, where $8c$ is the expansion constant. Let G_t denote the subgraph of G that is induced by the set of nodes $\{\{t, i\} : i \in T\}$. Clearly, $E(G)$ is the disjoint union of all G_t's. Every G_t is just a copy of H. For some constant a, the number of terminals j at distance at most $a \lg k$ from t is $o(k)$. Thus, $\Omega(k)$ edges in G_t are stretched to $\Omega(\log k)$. Summing over all balanced labels, we get that the total cost is $\Omega(k^2 \log k)$. □

4 Integrality Ratio for 0-Extension

If the metric d on the terminals is the shortest-path metric of a high-girth expander, the earthmover relaxation guarantees a constant integrality ratio for 0-Extension (proof omitted). Therefore, the Metric Labeling construction does not work for 0-Extension. An obvious suggestion is to insist on a small girth expander, for example, by taking the Cartesian product of an expander with itself. We don't know if this works; however, the following modification does work.

Consider an infinite family of (bounded-degree) expanders. Let H be a member of this family. The terminal set T is $V(H) \times V(H)$. Let $k = |V(H)|^2$ denote the number of terminals. The metric d on the terminals is given by

$$d\left((u, v), (u', v')\right) = \sqrt{\lg k} \cdot \mathrm{emd_H}\left(\{u, v\}, \{u', v'\}\right) + d_H(u, u');$$

here d_H is the shortest path metric on H, "$\{u, v\}$" denotes the probability distribution on vertices which assigns mass $1/2$ to each of u and v (likewise for "$\{u', v'\}$"), and $\mathrm{emd_H}(\{u, v\}, \{u', v'\})$ is the earthmover distance between the two probability distributions, the underlying metric being d_H.

The set V' of nonterminals is $V' = \binom{V(H)}{2}$. Notice that $|V'|$ is approximately $\frac{k}{2}$. The input graph G has node set $V = V(G) = T \cup V'$. To define the edge set, put

$$E_1 = \left\{\{\{u, v\}, \{u, v'\}\} : \ u, v, v' \in V(H) \text{ and } \{v, v'\} \in E(H)\right\},$$

$$E_2 = \left\{\{(u, v), \{u, v\}\} : \ u, v \in V(H)\right\},$$

and put $E = E(G) = E_1 \cup E_2$. Edges in E_1 join pairs of nonterminals and have weight $\sqrt{\lg k}$. Edges in E_2 join terminals to nonterminals and have weight 1.

Lemma 3 *The cost of the fractional solution for this instance is $O(k \log k)$.*

Proof: Consider the fractional solution that puts, for every $\{u, v\} \in V'$,

$$x^{\{u,v\}}_{(u,v)} = x^{\{u,v\}}_{(v,u)} = \frac{1}{2}.$$

By definition of d, the cost of an edge $\{(u,v),\{u,v\}\} \in E_2$ (which has weight 1) is $\frac{1}{2}d_H(u,v)$, which is $O(\log k)$. There are $O(k)$ such edges. The cost of an edge $\{\{u,v\},\{u,v'\}\} \in E_1$, not including its weight $\sqrt{\lg k}$, is the earthmover distance *over* d (not over d_H) between the configuration which splits its mass uniformly between (u,v) and (v,u) and the configuration which splits its mass uniformly between (u,v') and (v',u). This is at most $(1/2)d((u,v),(u,v')) + (1/2)d((v,u),(v',u)) \le (1/2)[\sqrt{\lg k}\cdot 1 + 0] + (1/2)[\sqrt{\lg k} + 1] = \sqrt{\lg k} + 1/2$. Hence its cost, including its weight, is $O(\log k)$. There are $O(k)$ such edges as well. □

Theorem 4 *The integrality ratio for this instance is* $\Omega\left(\sqrt{\log k}\right)$.

Proof: We will show that every integral solution must cost $\Omega\left(k(\log k)^{3/2}\right)$. Together with Lemma 3, this implies the lower bound.

Consider an arbitrary integral solution, where every $\{u,v\} \in V'$ is assigned to $\varphi(\{u,v\}) \in V(H) \times V(H)$. Let $\gamma > 0$ be a sufficiently small constant, and let

$$V_1 = \{\{u,v\} \in V' : \ \mathrm{emd_H}\left(\{u,v\},\{u',v'\}\right) \ge \gamma \lg k,$$

where $(u',v') := \varphi(u,v)\}$. For every $\{u,v\} \in V_1$, the edges in E_2 incident to $\{u,v\}$ (one or two such edges) cost vertices at least $\sqrt{\lg k}\cdot(\gamma \lg k)$, by definition of the metric d on the terminals. If $V_1 \ge \frac{k}{64}$ then the total cost is $\Omega\left(k(\log k)^{3/2}\right)$.

Otherwise, define a directed graph on $V(H)$ with no loops, parallel or antiparallel arcs as follows. Every node $e = \{u,v\} \in V' \setminus V_1$ contributes an arc. Let $(u',v') = \varphi(\{u,v\})$. If

$$\mathrm{emd_H}\left(\{u,v\},\{u',v'\}\right) = \frac{1}{2}d_H(u,u') + \frac{1}{2}d_H(v,v'),$$

then add the arc (u,v). Otherwise,

$$\mathrm{emd_H}\left(\{u,v\},\{u',v'\}\right) = \frac{1}{2}d_H(u,v') + \frac{1}{2}d_H(v,u')$$

(this is not obvious, but true); add the arc (v,u). (In other words, given $e = \{u,v\}$, choose y,z such that $\{y,z\} = \{u,v\}$ and

$$\mathrm{emd_H}(e,\{u',v'\}\} = (1/2)d_H(y,u') + (1/2)d_H(z,v')$$

and then add arc (y,z).) Unless $V_1 = \emptyset$, the resulting graph is not a tournament. Hence add arbitrary dummy arcs to make a tournament. The number of arcs that need to be added is $|V_1| < \frac{k}{64}$. By Lemma 1, at least $\frac{V(H)}{2} = \frac{1}{2}\sqrt{k}$ tournament nodes have both indegree and outdegree between $\frac{1}{8}\sqrt{k}$ and $\frac{7}{8}\sqrt{k}$. If we now remove the dummy arcs, at least $\frac{1}{4}\sqrt{k}$ tournament nodes have both indegree and outdegree between $\frac{1}{16}\sqrt{k}$ and $\frac{7}{8}\sqrt{k}$. (One has to remove $(1/16)\sqrt{k}$ arcs to "ruin" two vertices.) Consider such a node $u \in V(H)$. Let

$$O_u = \{v \in V(H) : \ (u,v) \text{ is in the partial tournament}\},$$

and

$$I_u = \{v \in V(H) : \ (v,u) \text{ is in the partial tournament}\}.$$

As $|I_u| \ge \frac{1}{16}\sqrt{k}$, there is a constant $\epsilon > 0$ such that $I'_u = \{v \in I_u : \ d_H(u,v) \ge \epsilon \lg k\}$ satisfies $|I'_u| \ge \frac{1}{32}\sqrt{k}$. We need $\gamma \le \frac{\epsilon}{4}$. As H is a bounded degree expander, there are $\Theta(\sqrt{k})$ constant-length, edge-disjoint paths between O_u and I'_u. Consider any such path, and let $v_1 \in O_u$ and $v_2 \in I'_u$ be its endpoints. Notice that $\varphi(\{u,v_1\}) = (u',v'_1)$, where $d_H(u,u') < \gamma \lg k$. Similarly, $\varphi(\{u,v_2\}) = (v'_2,u'')$, where $d_H(v_2,v'_2) < \gamma \lg k$. So, $d_H(u',v'_2) > (\epsilon - 2\gamma)\lg k \ge \frac{\epsilon}{2}\lg k$. Therefore, $d\left(\varphi(\{u,v_1\}),\varphi(\{u,v_2\})\right)$ is $\Omega(\log k)$. By the triangle inequality, there must be an edge $\{v,v'\} \in E(H)$ along the path (which has constant length) such that $d\left(\varphi(\{u,v\}),\varphi(\{u,v'\})\right)$ is $\Omega(\log k)$. Recall that every edge $\{\{u,v\},\{u,v'\}\} \in E_1$ has weight $\sqrt{\lg k}$. Therefore, the total cost of such edges, fixing $u \in V(H)$ with both indegree and outdegree at least $\frac{1}{16}\sqrt{k}$, is $\Omega\left(\sqrt{k\log k}\cdot \log k\right)$. Summing over all such u (each edge is counted at most twice), we get a total cost of $\Omega\left(k(\log k)^{3/2}\right)$. □

5 Hardness of 0-Extension

To prove the hardness of 0-Extension we start with the construction of [9] for the hardness of Metric Labeling and modify this construction so that it works for 0-Extension. We achieve this by applying a technique similar to the one applied in Section 4 to the Metric Labeling instance of Section 3.

Let us first recall the k-prover protocol of [9]. We start with a Gap-3SAT(5) formula ϕ. There are k provers $P_1, ..., P_k$ (k will be chosen later to be poly($\log n$), where n is the size of ϕ).

- For each (i, j), $1 \le i < j \le k$, the verifier chooses, randomly and independently, a clause C_{ij} and a distinguished variable x_{ij} from the clause. P_i is sent C_{ij} (and is expected to return an assignment to all variables of the clause), P_j is sent x_{ij} (and is expected to return an assignment to this variable), and every other prover is sent both C_{ij} and x_{ij} (and is expected to return an assignment to all variables of the clause). Thus the query sent to each prover has $\binom{k}{2}$ coordinates.
- The verifier checks, for each pair (i, j), that the answers of all the provers are consistent.

We denote the set of random strings used by the verifier by R. Given $r \in R$, and $1 \le i \le k$, let $q_i(r)$ be the query sent to P_i when the verifier chooses the random string r. Let $Q_i = \cup_r \{q_i(r)\}$ be the set of all possible queries to P_i. For $q \in Q_i$, let $\mathcal{A}_i(q)$ be the set of all possible answers of the ith prover to q which satisfy all the clauses appearing in the query.

Consider any pair P_i and P_j of provers. Let $q_i \in Q_i$ and $q_j \in Q_j$ be a pair of queries such that for some $r \in R$, $q_i = q_i(r)$ and $q_j = q_j(r)$. Let A_i and A_j denote the answers of provers P_i and P_j, respectively, to the queries. We say that the answers are *weakly consistent* if the assignments to C_{ij} in A_i and to x_{ij} in A_j are consistent. The answers are called *strongly consistent* if they are also consistent in every coordinate $(a, b) \neq (i, j)$.

We use the following theorem of Chuzhoy and Naor [9].

Theorem 5 *(Theorem 4.2 in [9]). There is a constant $0 < \epsilon < 1$ such that if ϕ is a Yes instance, then there is a strategy of the k provers such that the verifier always accepts, and if ϕ is a No instance, then for any strategy of the provers, for every pair P_i, P_j of provers, $i < j$, the probability that their answers are weakly consistent is at most $1 - \frac{\epsilon}{3}$.*

We now construct a 0-Extension instance from an instance of Gap-3SAT(5) based on the k-prover system described above. Recall that an instance of 0-Extension consists of a graph $G(V' \dot\cup T, E)$, where the set of vertices consists of two parts, the terminals T and the nonterminals V'. Each edge is between a terminal and a nonterminal or between two nonterminals. Every edge has a weight, which is the factor by which it contributes to the cost. Also provided is a metric on the set of terminals.

Our 0-Extension instance is based on the Metric Labeling instance in [9], with additional edges between nonterminals and terminals, and a special distance metric on the terminals. To define our instance, we proceed thus. We first define the set V' of nonterminals and the set T of terminals. The set of nonterminals (resp., terminals) is precisely the set of vertices (resp., labels) in the construction of [9]. We also define a graph $G_{V'}$ on V' and a graph G_T on T. Finally we define the weighted graph $G(V' \dot\cup T, E)$ of the input instance.

Nonterminals: V' consists of two types of nonterminals.

- For each i, $1 \le i \le k$, and each query $q \in Q_i$ there is a *query nonterminal* $v(i, q)$.
- For each random string r, there is a *constraint nonterminal* $v(r)$.

The graph $G_{V'}$ on V' is defined by placing, for each i and r, an edge between constraint nonterminal $v(r)$ and query nonterminal $v(i, q_i(r))$.[1] Each edge in $G_{V'}$ has length $\frac{1}{2}$.

[1] We assume without loss of generality that $G_{V'}$ is connected. In general it may be disconnected if the SAT formula we start with itself has disconnected components of variables, where the connectivity is via common clauses. But we can add dummy clauses to connect all variables, and this will yield a connected $G_{V'}$.

Terminals: T also consists of two types of terminals.

- For each i such that $1 \le i \le k$, each query $q \in Q_i$, and each answer $A_i \in \mathcal{A}_i(q)$ to the query q, there is a query terminal $(v(i,q), A_i)$.
- For every random string r of the verifier, for every k-tuple $(A_1, A_2, ..., A_k)$ of pairwise strongly consistent answers satisfying $A_i \in \mathcal{A}_i(q_i(r))$ for $1 \le i \le k$, there is a constraint terminal $(v(r), (A_1, A_2, ..., A_k))$.

Note that for every nonterminal x, there is a set of terminals of the form $(x, \cdot)$ derived from x. In what follows we will represent a generic terminal by (x, y). The graph G_T on T, defined only for the purpose of defining the metric on T, is defined by the following edges: incident on every constraint terminal $(v(r), (A_1, A_2, ..., A_k))$ is, for each i, an edge of length $\frac{1}{2}$ to query terminal $(v(i, q_i(r)), A_i)$.

Metric on terminals: We now use the graphs G_T and $G_{V'}$ to define the metric d_T on T. To do so, we first define two different metrics, Δ on T and M_Δ on V'.

For $t, t' \in T$, let $\Delta(t, t')$ equal the minimum of k and the distance between t and t' in G_T. Note that this is indeed a metric. For $x, x' \in V'$, let $M(x, x')$ be the minimum of k and the distance between x and x' in $G_{V'}$. Now we can define the metric on the set T of terminals. For two terminals (x, y) and (x', y'), define $d_T((x,y),(x',y')) = \sqrt{k} \cdot M(x,x') + \Delta((x,y),(x',y'))$.

Input graph: The input graph consists of the set of nonterminals and terminals. There are two kinds of edges. The first kind consists of edges between two nonterminals. These are precisely the edges of the graph $G_{V'}$ and have weight $\sqrt{k}$. The second kind consists of those between a nonterminal and a terminal, and are defined as follows: for every $r \in R$, and for every k-tuple $(A_1, A_2, ..., A_k)$ of strongly consistent answers, with $A_i \in \mathcal{A}_i(q_i(r))$ for $1 \le i \le k$, there is an edge between constraint nonterminal $v(r)$ and constraint terminal $(v(r), (A_1, ..., A_k))$. Similarly, for every $r \in R$, $i = 1, ..., k$, and every possible answer A_i of prover P_i to $q_i(r)$, there is an edge between query nonterminal $v(i, q_i(r))$ and query terminal $(v(i, q_i(r)), A_i)$. To define the weight of an edge of the second kind, we define the following. For a nonterminal v, let d_v be the number of nonterminal-nonterminal edges incident on v, and let z_v be the number of nonterminal-terminal edges incident on v. Then the weight of every nonterminal-terminal edge incident on v is $w_v = d_v/z_v$. (If $z_v = 0$ there are no edges awaiting weights.) Define $y = \frac{1}{2}\sum_{v \in V'} d_v$, the total number of nonterminal-nonterminal edges. Note that y is also equal to $k|R|$.

5.1 Yes Instance

We assume now that the SAT formula is a Yes instance. Then there is a strategy of the provers so that the verifier accepts with probability 1. For $i = 1, ..., k$ and query $q_i \in Q_i$, let $f_i(q_i) \in \mathcal{A}_i(q_i)$ be the answer of prover P_i to query q_i under this strategy. Note that for each random string r, $f_1(q_1(r)), f_2(q_2(r)), ..., f_k(q_k(r))$ are pairwise strongly consistent. From this strategy, we can define the following assignment of nonterminals to terminals.

For every random string r, assign constraint nonterminal $v(r)$ to constraint terminal $(v(r), (f_1(q_1(r)), f_2(q_2(r)), ..., f_k(q_k(r))))$. For every random string r and $i = 1, ..., k$, assign query nonterminal $v(i, q_i(r))$ to query terminal $(v(i, q_i(r)), f_i(q_i(r)))$.

Consider an edge between two nonterminals, say, constraint nonterminal $v(r)$ and query nonterminal $v(i, q_i(r))$. Let $a = (v(r), (f_1(q_1(r)), f_2(q_2(r)), ..., f_k(q_k(r))))$ and $b = (v(i, q_i(r)), f_i(q_i(r)))$. Since $v(r)$ is assigned to terminal a and $v(i, q_i(r))$ is assigned to terminal b, the distance to which this edge is stretched is

$$d_T(a,b) = \sqrt{k} \cdot M(v(r), v(i, q_i(r))) + \Delta(a,b) \le \sqrt{k}/2 + 1/2.$$

This is because $v(r)$ and $v(i, q_i(r))$ are neighbors in $G_{V'}$ and a and b are neighbors in G_T. The weight of the edge between the nonterminals $v(r)$ and $v(i, q_i(r))$ is $\sqrt{k}$; hence the contribution to the cost is at most $\sqrt{k}((1/2)\sqrt{k} + 1/2)$. Since there are a total of y number of edges of this type, the total contribution of such edges to the cost is at most yk.

Consider an edge between a nonterminal, say, a constraint nonterminal $v(r)$, and a constraint terminal $b = (v(r), (A_1, A_2, ..., A_k))$. (The case of an edge between a query nonterminal and a query terminal is identical.) Let $a = (v(r), (f_1(q_1(r)), f_2(q_2(r)), ..., f_k(q_k(r))))$. Since $v(r)$ is assigned to a, the distance to which this edge is stretched is

$$d_T(a, b) = \sqrt{k} \cdot M(v(r), v(r)) + \Delta(a, b) \leq \sqrt{k} \cdot 0 + k.$$

The inequality follows because the distances under Δ are at most k. This is true for all the $z_{v(r)}$ nonterminal-terminal edges incident on $v(r)$. Since the weight of each such edge is $w_{v(r)} = d_{v(r)}/z_{v(r)}$, the contribution to the cost of all these edges is at most $w_{v(r)} z_{v(r)} k = d_{v(r)} k$. Summing over all nonterminals v, we get that the total contribution to the cost of all nonterminal-terminal edges is $\sum_{v \in V'} d_v k = 2yk$. Thus the total cost in the Yes case is at most $yk + 2yk = 3yk$.

5.2 No Instance

Let $f : V' \rightarrow T$ be any assignment of nonterminals to terminals. For $v \in V'$, define $g(v), h(v)$ by $f(v) = (g(v), h(v))$. Let $V_1 = \{v \in V' : M(v, g(v)) \geq \gamma k\}$ for some small constant $0 < \gamma < 1$ to be chosen later. We also pick a constant $\alpha > 0$, to be fixed later. We consider two cases.

Case 1: $\sum_{v \in V_1} d_v > \alpha \sum_{v \in V'} d_v = \alpha(2y)$. Take any nonterminal in V_1, say, a constraint nonterminal $v(r)$ (the case of a query nonterminal is identical), and consider any terminal $a = (v(r), (A_1, A_2, ..., A_k))$ such that there is an edge between $v(r)$ and a. Then the distance to which this edge is stretched is

$$d_T(f(v(r)), a) \;=\; \sqrt{k} \cdot M(g(v(r)), v(r)) + \Delta(f(v(r)), a) \;\geq\; \sqrt{k} \cdot (\gamma k) + 0 \;=\; \gamma k^{3/2}.$$

The inequality follows from the fact that $v(r) \in V_1$. This is true for all the $z_{v(r)}$ nonterminal-terminal edges incident on $v(r)$. Each such edge has a weight of $w_{v(r)} = d_{v(r)}/z_{v(r)}$. Hence the contribution to the cost incurred by the nonterminal-terminal edges incident on nonterminals in V_1 is at least

$$\begin{aligned} \sum_{v \in V_1} z_v w_v (\gamma k^{3/2}) &= \gamma k^{3/2} \sum_{v \in V_1} d_v \\ &> \gamma k^{3/2} [\alpha(2y)] \\ &= 2\gamma\alpha y k^{3/2}. \end{aligned}$$

The inequality follows because we are in Case 1. Hence the total cost in this case is at least $2\gamma\alpha y k^{3/2}$.

Case 2: $\sum_{v \in V_1} d_v \leq \alpha \sum_{v \in V'} d_v = \alpha(2y)$. We will first change the assignment $f = (g, h)$ to an assignment $f' = (g', h')$ such that for all $v \in V'$, $g'(v) = v$. (Such a "natural" assignment corresponds to the Metric Labeling, not 0-Extension, work of Chuzhoy and Naor [9]. Once we have such an assignment we will be able to invoke the main lemma of [9].) Furthermore, we will not change f much in going to f': we will have $\Delta(f(v), f'(v)) < \gamma k$ for all $v \in V' \backslash V_1$. We get the "natural" assignment simply by changing the assignment of the nonterminal v from $f(v)$ to that terminal of the form $(v, \cdot)$ which is closest, according to distance Δ on T, to $f(v)$.

But have we changed the assignments too much? Recall that by definition, for every $v \in V' \backslash V_1$, $M(v, g(v)) < \gamma k$. That is, it takes fewer than $2\gamma k$ steps in the graph $G_{V'}$ on nonterminals to move from $g(v)$ to v (the factor of 2 appears because every edge in $G_{V'}$ is of length $\frac{1}{2}$). But this implies that it takes fewer than $2\gamma k$ steps in the graph G_T on terminals to move from $(g(v), h(v))$ to some terminal of the form $(v, \cdot)$; that is, $\Delta(f(v), f'(v)) \leq M(v, g(v)) < \gamma k$. (This follows from the structure of the graph. If x and x' are adjacent nonterminals and (x, y) is any terminal, then there is a terminal (x', y') adjacent to (x, y); y' "gives the same answer to the question" as y.)

Now, because $g'(v) = v$ for all $v \in V'$, we have a valid assignment in the sense of the Metric Labeling, not 0-Extension, instance of [9]. An edge between two nonterminals, say $v(r)$ and $v(i, q_i(r))$, is stretched to

$$d_T\left(f'(v(r)), f'(v(i, q_i(r)))\right) = \sqrt{k} \cdot M\left(g'(v(r)), g'(v(i, q_i(r)))\right) + \Delta\left(f'(v(r)), f'(v(i, q_i(r)))\right). \tag{1}$$

By summing over all nonterminal-nonterminal edges, and ignoring the first term on the right-hand side of (1), we get

$$\sum_{r,i} d_T\left(f'(v(r)), f'(v(i, q_i(r)))\right) \geq \sum_{r,i} \Delta\left(f'(v(r)), f'(v(i, q_i(r)))\right). \tag{2}$$

Here is the key point. By Proposition 4.4 and Lemma 4.5 in [9], we know that the right-hand side of (2) is at least $\binom{k}{2}\frac{\epsilon}{3}|R|$. (While [9] does not "truncate" the distance metric on the terminal graph at k, as we do for Δ, it can be shown that their proof works even when such truncation is done). Since the total number of nonterminal-nonterminal edges is $y = k|R|$, we get that the total stretch of nonterminal-nonterminal edges is at least $(\epsilon' k)y$, for some constant $\epsilon' > 0$.

We now wish to compare this to the total stretch of these edges in the original assignment f. In transforming f to f', we may have increased the total stretch. Call a nonterminal-nonterminal edge *bad* if it is incident to at least one nonterminal in V_1. Thus there are at most $\sum_{v \in V_1} d_v$ bad edges. Since we are in Case 2, this means that there are at most $(2\alpha)y$ bad edges, i.e., at most a 2α-fraction of all nonterminal-nonterminal edges is bad. Call the nonterminal-nonterminal edges which are not bad *good*. Then, when we go back from f' to f, $|\Delta(f'(v), f'(u)) - \Delta(f(v), f(u))| \leq 2\gamma k$; this is true because if an edge with endpoints u, v is good, both u and v are not in V_1, and hence $\Delta(f'(v), f(v)) \leq \gamma k$, and $\Delta(f'(u), f(u)) \leq \gamma k$. For a bad edge $|\Delta(f'(v), f'(u)) - \Delta(f(v), f(u))| \leq k$. Hence the total stretch of nonterminal-nonterminal edges in the assignment f is at least $(\epsilon' k)y - (2\alpha y)k - y(2\gamma k) = yk(\epsilon' - 2\alpha - 2\gamma)$, where the first subtracted term corresponds to the bad edges, and the second subtracted term corresponds to the good edges.

We choose α and γ small enough so that, for f, the total stretch of nonterminal-nonterminal edges becomes at least $\epsilon'' yk$, for some constant $\epsilon'' > 0$. Since each nonterminal-nonterminal edge has a weight of $\sqrt{k}$, the total contribution of these edges to the cost, and hence the total cost in Case 2, is at least $\epsilon'' y k^{3/2}$.

Thus the ratio between the costs in the No and Yes cases is $\Omega(\sqrt{k})$. As in [9], the size N of the instance that we constructed is $n^{O(k^2)}$, where n is the size of the formula ϕ from which we started. Choosing k to be $\text{poly}(\lg n)$, which is $(\lg N)^{\frac{1}{2}-\delta}$ for an arbitrarily small constant $\delta > 0$, we get

Theorem 6 *For any constant $\delta > 0$, there is no $O((\log N)^{\frac{1}{4}-\delta})$-approximation algorithm for* 0-Extension *unless $NP \subseteq DTIME(n^{\text{poly}(\log n)})$.*

6 A $O(\sqrt{\text{diam}(d)})$-Approximation Algorithm for 0-Extension

We define the *diameter* of (T, d) as $\text{diam}(d) = \frac{\max_{i,j} d(i,j)}{\min_{i \neq j} d(i,j)}$. Alternatively, we can scale d so that the minimum distance between different points is 1, and then the diameter is simply the largest distance. We describe a rounding algorithm that guarantees a ratio of $O(\sqrt{\text{diam}(d)})$ between the costs of the fractional solution and of the rounded solution. Let $G = (V, E)$ be an input graph, and let $T \subseteq V$ denote the set of terminals. Given a solution x for the earthmover relaxation, let $\text{emd}(v, T) = \min_{j \in T}\{\text{emd}(x^v, x^j)\}$. The algorithm uses the following lemma.

Lemma 7 *(Archer et al. [1]). There exist $c_1, c_2 > 0$ such that for every input graph $G = (V, E)$, for every set T of terminals, and for every solution x for the earthmover relaxation, there exists a distribution on solutions y such that for every $v \in V$, if $\text{emd}(x^v, x^j) > c_1 \cdot \text{emd}(v, T)$, then $y^v_j = 0$, and furthermore, for every $u, v \in V$, $\text{E}_y[\text{emd}(y^u, y^v)] \leq c_2 \cdot \text{emd}(x^u, x^v)$.*

We use the rounding algorithm of [17], designed for the case in which d is a uniform metric.

Lemma 8 *(Kleinberg and Tardos [17].) There is a probabilistic polynomial-time rounding algorithm that, given a feasible solution x to the earthmover relaxation, generates a probability distribution over assignments $\varphi : V \to T$ satisfying $\varphi(v) = v$ if $v \in T$ such that for every $u, v \in V$, $\mathrm{E}[d(\varphi(u), \varphi(v)] \leq \|x^u - x^v\|_1$; and for every $v \in V$ and $i \in T$, $\Pr[\varphi(v) = i] \leq 2x_i^v$.*

We are now ready to describe the algorithm. We assume that d is scaled so that the minimum distance between different terminals is 1. Thus $\mathrm{diam}(d)$ is the maximum distance between terminals. First, pick α in the range $\sqrt{\mathrm{diam}(d)} < \alpha < 2\sqrt{\mathrm{diam}(d)}$ uniformly at random. Assign to terminal 1 all nodes $v \in V$ such that $\mathrm{emd}(v, T) > \alpha$. Second, truncate x to a (random) solution y using Lemma 7, ignoring the nodes already assigned. Use the uniform metric-case rounding algorithm of Lemma 8 on y to assign the remaining nodes to terminals.

Theorem 9 *The expected cost of the rounded solution is $O(\sqrt{\mathrm{diam}(d)})$ times the cost of x.*

Proof: We do an edge-by-edge analysis. We calculate the expected cost incurred in the first phase and that incurred in the second phase, showing that both are $O(\sqrt{\mathrm{diam}(d)})$ times the cost of x in the linear program.

Take any edge $\{u, v\}$; first, we calculate the expected cost it incurs in the first phase. Choose u so that $\mathrm{emd}(u, T) \leq \mathrm{emd}(v, T)$. By the triangle inequality, $\mathrm{emd}(x^u, x^v) \geq \mathrm{emd}(v, T) - \mathrm{emd}(u, T)$. As we draw α uniformly in a range of length $\sqrt{\mathrm{diam}(d)}$, we have that the probability that $\alpha \in [\mathrm{emd}(u, T), \mathrm{emd}(v, T)]$ is at most $\mathrm{emd}(x^u, x^v)/\sqrt{\mathrm{diam}(d)}$. If this happens, $\{u, v\}$ is stretched to a length of at most $\mathrm{diam}(d)$, so the expected contribution of these edges to the cost is at most the cost of x times $O(\sqrt{\mathrm{diam}(d)})$.

Now we calculate the expected cost incurred in the second phase by an edge $\{u, v\}$. For it to be positive, it must be the case that $\mathrm{emd}(u, T), \mathrm{emd}(v, T) < 2\sqrt{\mathrm{diam}(d)}$, for otherwise either one or both of u and v was already assigned in the first phase, and hence there is no cost to charge to $\{u, v\}$ in the second phase. Hence we may assume $\mathrm{emd}(u, T), \mathrm{emd}(v, T) < 2\sqrt{\mathrm{diam}(d)}$. We condition on $\alpha \geq \max\{\mathrm{emd}(u, T), \mathrm{emd}(v, T)\}$. This happens with probability at most 1, so we are perhaps overestimating the cost of the rounding via the uniform-case algorithm. Suppose u, v are assigned to terminals t_u, t_v, respectively. By Lemma 8, the guarantee of the uniform-case rounding rounding algorithm is that $\Pr[t_u \neq t_v] \leq \|y^u - y^v\|_1$. Notice that by the triangle inequality, $d(t_u, t_v) \leq \mathrm{emd}(t_u, x^u) + \mathrm{emd}(x^u, x^v) + \mathrm{emd}(x^v, t_v)$. Further notice that by Lemma 8, $y_{t_u}^u \neq 0$, so by Lemma 7, $\mathrm{emd}(t_u, x^u) \leq c_1 \,\mathrm{emd}(x^u, T) \leq 2c_1\sqrt{\mathrm{diam}(d)}$. The term $\mathrm{emd}(x^v, t_v)$ can be bounded similarly, so $d(t_u, t_v) \leq \mathrm{emd}(x^u, x^v) + 4c_1\sqrt{\mathrm{diam}(d)}$. Also, $\mathrm{emd}(x^u, x^v) \geq \mathrm{E}_y[\mathrm{emd}(y^u, y^v)]/c_2 \geq E_y[\|y^u - y^v\|_1]/(2c_2)$ (as the minimum distance between different terminals is 1). Fix y. We have

$$\begin{aligned}
\mathrm{E}[d(t_u, t_v)] &\leq \Pr[t_u \neq t_v] \cdot (\mathrm{emd}(x^u, x^v) + 4c_1\sqrt{\mathrm{diam}(d)}) \\
&\leq \|y^u - y^v\|_1 \cdot (\mathrm{emd}(x^u, x^v) + 4c_1\sqrt{\mathrm{diam}(d)}) \\
&\leq 2 \cdot \mathrm{emd}(x^u, x^v) + 4c_1\sqrt{\mathrm{diam}(d)} \cdot \|y^u - y^v\|_1,
\end{aligned}$$

as $\|y^u - y^v\|_1 \leq 2$. Taking the expectation over y, we get

$$\begin{aligned}
\mathrm{E}_y[\mathrm{E}[d(t_u, t_v)]] &\leq 2 \cdot \mathrm{emd}(x^u, x^v) + 4c_1\sqrt{\mathrm{diam}(d)} \cdot \mathrm{E}_y[\|y^u - y^v\|_1] \\
&\leq (2 + 8c_1c_2\sqrt{\mathrm{diam}(d)}) \cdot \mathrm{emd}(x^u, x^v),
\end{aligned}$$

using the above inequalities. Adding together the costs in the two phases gives us a ratio of $O(\sqrt{\mathrm{diam}(d)})$. □

References

[1] A. Archer, J. Fakcharoenphol, C. Harrelson, R. Krauthgamer, K. Talwar, and É. Tardos, Approximate classification via earthmover metrics, in *Proc. SODA '04*.

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

[3] Y. Bartal, On approximating arbitrary metrics by tree metrics, in *Proc. STOC '98*.

[4] J. Bourgain, The metrical interpretation of superreflexivity in Banach spaces, *Israel J. Math.*, 56(2):222–230, 1986.

[5] G. Calinescu, H. J. Karloff, and Y. Rabani, An improved approximation algorithm for MULTIWAY CUT, *J. Computing and Syst. Sci.*, 60(3):564–574, 2000 (preliminary version in *STOC '98*).

[6] G. Calinescu, H. J. Karloff, and Y. Rabani, Approximation algorithms for the 0-EXTENSION problem, *SIAM J. Computing*, 34(2):358–372, 2004 (preliminary version in *SODA '01*).

[7] M. Charikar, private communication, 2000.

[8] C. Chekuri, S. Khanna, J. Naor, and L. Zosin, Approximation algorithms for the METRIC LABELING problem via a new linear programming formulation, to appear in *SIAM J. on Discrete Math* (preliminary version in *SODA '01*).

[9] J. Chuzhoy and J. Naor, The hardness of METRIC LABELING, in *Proc. FOCS '04*, 108–114.

[10] J. Fakcharoenphol, C. Harrelson, S. Rao, and K. Talwar, An improved approximation algorithm for the 0-EXTENSION PROBLEM, in *Proc. SODA '03*, 342–352.

[11] J. Fakcharoenphol, S. Rao, and K. Talwar, A tight bound on approximating arbitrary metrics by tree metrics, in *Proc. STOC '03*, 448–455.

[12] A. Gupta and E. Tardos, A constant factor approximation algorithm for a class of classification problems, in *Proc. STOC '00*, pages 652–658.

[13] W. B. Johnson, J. Lindenstrauss, and G. Schechtman, Extensions of Lipschitz maps into Banach space, *Israel J. Math.*, 54(2):129–138, 1986.

[14] A. V. Karzanov, Minimum 0-extension of graph metrics, *Europ. J. Combinat.*, 19:71–101, 1998.

[15] R. Krauthgamer, J. Lee, M. Mendel, and A. Naor, "Measured Descent: A New Embedding Method For Finite Metrics," *Geometric and Functional Analysis* 15 (4), 839-858, 2005 (preliminary version in *FOCS '04*).

[16] S. Khot and A. Naor, Nonembeddability theorems via Fourier analysis, in *Proc. FOCS '05*.

[17] J. Kleinberg and É. Tardos. Approximation algorithms for classification problems with pairwise relationships: METRIC LABELING and Markov random fields, *J. Assoc. Comput. Mach.*, 49:616–639, 2002 (preliminary version in *FOCS '99*).

[18] J. R. Lee and A. Naor, Extending Lipschitz functions via random metric partitions, *Math. Invent.*, 160(1): 59-95, 2005.

[19] N. Linial, E. London, and Y. Rabinovich, The geometry of graphs and some of its algorithmic applications, *Combinatorica* 15(2):215–245, 1995 (preliminary version in *FOCS '94*).

A Alternative Integrality Ratio Constructions

In this section we provide an alternative construction of a METRIC LABELING instance for which the earth-mover relaxation has an integrality ratio of $\Omega(\log |T|)$. We also show how the method used in Section 4 can be applied to this construction to find an instance of 0-EXTENSION with integrality ratio $\Omega(\sqrt{\log |T|})$. This construction uses properties of certain linear codes, and is inspired by the results in [16].

Fix a linear code $C \subseteq \{0,1\}^n$ with distance at least ηn and rate at least $(1-\delta)n$ where η and δ are sufficiently small constants (the Gilbert-Varshamov bound shows that such codes exist with $\delta \approx H(\eta)$ where $H()$ is the binary entropy function).

Let T_0 be the orthogonal subspace of the code, i.e., $T_0 = C^\perp$. Let $\mathcal{T}$ denote the set of all cosets of T_0, i.e.,

$$\mathcal{T} = \{T_0 + v \ : \ v \in \{0,1\}^n\}.$$

Note that every coset has cardinality $2^{\delta n}$ and the number of cosets is $2^{(1-\delta)n}$.

Let Δ denote the Hamming metric on $\{0,1\}^n$. Consider the following metric d_{EM} on $\mathcal{T}$: For $T, T' \in \mathcal{T}$,

$$d_{EM}(T,T') := \min_{u\in T, v \in T'} \Delta(u,v).$$

In fact, for any fixed $u_0 \in T, v_0 \in T'$, we have

$$d_{EM}(T,T') = \min_{v \in T'} \Delta(u_0, v) = \min_{u\in T} \Delta(u, v_0).$$

This is indeed the earthmover distance. Consider the uniform probability distribution on T and T', respectively. The earthmover distance between these distributions (with underlying Hamming metric) is exactly $d_{EM}(T,T')$. This is because there is a "matching" between points in T and T' such that the Hamming distance between every pair of matched points is exactly d_{EM}.

The following two lemmas are from [16]. We provide the proofs here for completeness.

Lemma 10 *Let $\theta > 0$ be a sufficiently small constant. Then if two cosets T, T' are picked at random from $\mathcal{T}$, then with high probability $d_{EM}(T,T') \geq \ \theta n$.*

Proof: Fix coset T and fix any $u_0 \in T$. Consider the process of picking another random coset T'. One can pick a $y \in \{0,1\}^n$ at random and define $T' = T + y$. Clearly,

$$\begin{aligned}
\Pr[\exists v \in T' \text{ such that } \Delta(u_0, v) \leq \theta n] &= \Pr[\exists u \in T \text{ such that } \Delta(u_0, u+y) \leq \theta n] \\
&\leq \sum_{u \in T} \Pr[\Delta(u_0, u+y) \leq \theta n] \\
&\leq |T| \cdot 2^{-(1-H(\theta))n} \\
&= 2^{-(1-H(\theta)-\delta)n},
\end{aligned}$$

where the inequality on the penultimate line follows because $u+y$ is a random vector and its distance from u_0 has binomial distribution with mean $n/2$. □

Lemma 11 *Let $f : \mathcal{T} \mapsto \mathbf{R}^m$ be any assignment of vectors to points in $\mathcal{T}$. Let $\|\cdot\|$ denote the ℓ_2-norm. Then*

$$E_{T,i}[\|f(T) - f(T+e_i)\|^2] \geq 2\eta \cdot E_{T,T'}[\|f(T) - f(T')\|^2],$$

where e_i denotes a vector whose i^{th} coordinate is equal to 1 and the rest are 0. T and T' are random cosets picked independently and i is picked randomly (and independently of T) from $1 \leq i \leq n$.

Proof: Clearly, it suffices to prove this when f is a real-valued function (i.e., $m = 1$), since the desired inequality can be "split" into separate inequalities for every dimension. So assume f is a real-valued function. Let f' be a function on $\{0,1\}^n$ that is constant on every coset T and whose value on this coset equals $f(T)$. Clearly,

$$\begin{aligned}
E_{T,e_i}[|f(T) - f(T+e_i)|^2] &= E_{x\in\{0,1\}^n, e_i}[|f'(x) - f'(x+e_i)|^2] \\
&= E_{x,e_i}[f'(x)^2 + f'(x+e_i)^2 - 2f'(x)f'(x+e_i)].
\end{aligned}$$

Note that

$$E_x[f'(x)^2] = E_{x,e_i}[f'(x+e_i)^2] = \sum_{S\subseteq[n]} \hat{f}'(S)^2.$$

Also, using Fourier expansion,

$$\begin{aligned} E_{x,e_i}[f'(x)f'(x+e_i)] &= E_{x,e_i}[\sum_{S,S'} \hat{f}'(S)\hat{f}'(S')\chi_S(x)\chi_{S'}(x+e_i)] \\ &= \sum_S \hat{f}'(S)^2 E_{e_i}[\chi_S(e_i)] \\ &= \sum_S \hat{f}'(S)^2(1-2|S|/n). \end{aligned}$$

Combining these, we get

$$E_{T,e_i}[|f(T)-f(T+e_i)|^2] = 4\sum_{S\subseteq[n]} \hat{f}'(S)^2 \cdot \frac{|S|}{n}.$$

Now note that the function f' is constant on every coset of T_0 and hence only those Fourier coefficients are non-zero that are in $T_0^\perp$, i.e., those that are codewords in C. Thus either $S=\emptyset$ or $|S|\geq \eta n$ since the code has distance ηn. The lemma follows by observing that the total Fourier mass on non-empty coefficients is given by

$$\begin{aligned} \sum_{S\subseteq n} \hat{f}'(S)^2 - \hat{f}'(\emptyset)^2 &= E_x[f'(x)^2] - E_x[f'(x)]^2 \\ &= \frac{1}{2}E_{x,x'}[f'(x)^2 + f'(x')^2 - 2f'(x)f'(x')] \\ &= \frac{1}{2}E_{x,x'}[|f'(x)-f'(x')|^2] \\ &= \frac{1}{2}E_{T,T'}[|f(T)-f(T')|^2. \end{aligned}$$

□

A.1 Metric Labeling

Consider the following Metric Labeling instance. The label set is $\{0,1\}^n$. The distance between two labels is the Hamming distance between them. The input graph has vertices corresponding to cosets of T_0, i.e., the set of vertices is $\mathcal{T}$. There is an edge between T and $T+e_i$ for every coset T and a coordinate vector e_i. All edges have weight 1. The cost of assigning a vertex corresponding to a coset T to a label $x\in\{0,1\}^n$ is 0 if $x\in T\subseteq\{0,1\}^n$, and ∞ otherwise.

Fractional Solution: The fractional solution assigns to every vertex T a uniform probability distribution on labels in T. The earthmover distance between such distributions is exactly d_{EM}. For every edge $(T,T+e_i)$, we have $d_{EM}(T,T+e_i)=1$. Thus the average cost per edge in the fractional case is 1.

Integral Solutions: Now we will prove an $\Omega(n)$ lower bound on the average cost per edge of any integral solution.

Take any labeling of vertices, i.e., a map $h:\mathcal{T}\mapsto\{0,1\}^n$ such that for every coset T, $h(T)\in T$. We also think of values of h as vectors in $\mathbf{R}^n$. The following series of inequalities gives the desired lower bound.

$$\begin{aligned} E_{T,i}[\Delta(h(T), h(T+e_i)] &= E_{T,i}[\|h(T) - h(T+e_i)\|^2] \\ &\geq 2\eta \cdot E_{T,T'}[\|h(T) - h(T')\|^2] \\ &= 2\eta \cdot E_{T,T'}[\Delta(h(T), h(T'))] \\ &\geq 2\eta \cdot E_{T,T}[\min_{u \in T, v \in T'} \Delta(u,v)] \\ &= 2\eta \cdot E_{T,T'}[d_{EM}(T,T')], \end{aligned}$$

which is $\Omega(n)$, where on the second line we used Lemma 11 and at the end we used Lemma 10. Thus the integrality ratio of the earthmover relaxation for this instance is $\Omega(n)$, which is $\Omega(\log |T|)$.

A.2 0-Extension

We define an instance of 0-Extension as follows. The set X of terminals is defined as

$$X := \{(T, x) \ : \ T \in \mathcal{T},\ x \in \{0,1\}^n,\ x \in T\}.$$

The metric d_X on X is defined as

$$d_X((T,x),(T',x')) := L \cdot d_{EM}(T,T') + \Delta(x,x'),$$

where L will be chosen to be $\sqrt{n}$ later.

The set of nonterminals is defined to be $\mathcal{T}$. The input graph has as its vertex set the union of the set of terminals and the set of nonterminals. There is an edge between nonterminals T and $T + e_i$. These edges have a weight of K (which we will choose to be $\sqrt{n}$ later). There are edges from a nonterminal T to all terminals $\{(T,x) : x \in T\}$. These edges have a weight of 1.

Thus, the cost of an assignment $f : \mathcal{T} \mapsto X$ of nonterminals to terminals is

$$cost(f) := K \cdot E_{T,T+e_i}[d_X(f(T), f(T+e_i))] + E_{T,x\in T}[d_X(f(T),(T,x))].$$

Call the two components of the cost function as $cost_1$ and $cost_2$, respectively.

Fractional Solution:

We construct a fractional solution whose cost is at most $K \cdot (L+1) + n$.

Assign to a nonterminal T the uniform probability distribution on the set $\{(T,x) : x \in T\}$ of terminals. Clearly, the "movement" $(T,x) \mapsto (T+e_i, x+e_i)$ "moves" this distribution to the uniform probability distribution on the set of terminals $\{(T+e_i, x') : x' \in T+e_i\}$. Therefore, the contribution to the $cost_1$ component of the cost is

$$\begin{aligned} K \cdot d_X((T,x),(T+e_i,x+e_i)) &= K \cdot (L \cdot d_{EM}(T, T+e_i) + \Delta(x, x+e_i)) \\ &= K \cdot (L+1). \end{aligned}$$

Also for any T and $x_0 \in T$, the earthmover distance between the uniform distribution on set $\{(T,x) : x \in T\}$ and the distribution "concentrated" at (T, x_0) is at most n. This is an upper bound on the $cost_2$ component of the fractional cost.

Integral Solutions:

We will prove a $\min\{\Omega(nK), \Omega(nL)\}$ lower bound on the cost of any integral solution.

Let $f : \mathcal{T} \mapsto X$ be any assignment of nonterminals to terminals. Denote $f(T) = (g(T), h(T))$, where $g(T) \in \mathcal{T},\ h(T) \in \{0,1\}^n$ and $h(T) \in g(T)$.

We consider two cases.

Case (i): $E_{T,T'}[d_{EM}(g(T),g(T'))] \le \gamma n$ where $\gamma > 0$ is a small constant to be chosen later.
Applying the triangle inequality, we have

$$d_{EM}(T,T') \le d_{EM}(T,g(T)) + d_{EM}(g(T),g(T')) + d_{EM}(g(T'),T').$$

Taking expectation over random T, T' and using Lemma 10, we see that

$$\theta n \le 2 \cdot E_T[d_{EM}(g(T),T)] + \gamma n.$$

Assuming $\gamma \le \theta/2$, we have

$$E_T[d_{EM}(g(T),T)] \ge \theta n/4. \tag{3}$$

Now we will show that the $cost_2$-component of the cost is at least $L \cdot \theta n/4$. Indeed,

$$\begin{aligned} cost_2 &= E_{T,x\in T}[d_X(f(T),(T,x))] \\ &= E_{T,x\in T}[d_X((g(T),h(T)),(T,x))] \\ &= E_{T,x\in T}[L \cdot d_{EM}(g(T),T) + \Delta(h(T),x)] \\ &\ge E_T[L \cdot d_{EM}(g(T),T)] \\ &\ge L \cdot \theta n/4. \end{aligned}$$

Case (ii): $E_{T,T'}[d_{EM}(g(T),g(T'))] \ge \gamma n$.
From $h(T) \in g(T)$, $h(T') \in g(T')$, and the definition of d_{EM}, we get

$$E_{T,T'}[\Delta(h(T),h(T')] \ge \gamma n.$$

We will show a lower bound of $K \cdot 2\eta\gamma n$ on the $cost_1$-component of the cost. We see that $cost_1$ equals

$$\begin{aligned} K \cdot E_{T,T+e_i}[d_X((g(T),h(T)),(g(T+e_i),h(T+e_i)))] &\ge K \cdot E_{T,T+e_i}[\Delta(h(T),h(T+e_i))] \\ &= K \cdot E_{T,T+e_i}[\|h(T)-h(T+e_i)\|^2] \\ &\ge K \cdot 2\eta \cdot E_{T,T'}[\|h(T)-h(T')\|^2] \\ &= K \cdot 2\eta \cdot E_{T,T'}[\Delta(h(T),h(T'))] \\ &\ge K \cdot 2\eta \cdot \gamma n, \end{aligned}$$

where we used Lemma 11 again.

Choosing $K = L = \sqrt{n}$ gives an upper bound of $O(n)$ on the fractional cost and a lower bound of $\Omega(n^{3/2})$ on the cost of any integral solution. This proves the $\Omega(\sqrt{\log |T|})$ integrality ratio for 0-Extension. Observe the tradeoff between the two cost components $cost_1$ and $cost_2$ that limits to $\Omega(\sqrt{\log N})$ the lower bound that we can prove.