

Approximation algorithms for minimizing average distortion

Kedar Dhamdhere^{*1}, Anupam Gupta², and R. Ravi^{**3}

¹ School of Computer Science, Carnegie Mellon University. kedar@cs.cmu.edu

² School of Computer Science, Carnegie Mellon University. anupamg@cs.cmu.edu

³ Graduate School of Industrial Administration, Carnegie Mellon University.
ravi@cmu.edu

Abstract. We study the problem of embedding arbitrary finite metrics into a line metric in a non-contracting fashion to approximate the minimum average distortion. Since a path metric (or a line metric) is quite restricted, these embeddings could have high average distortions ($\Omega(n)$, where n is the number of points in the original metric). Furthermore, we prove that finding best embedding of even a tree metric into a line to minimize average distortion is NP-hard. Hence, we focus on *approximating* the best possible embedding for given input metric.

We give a constant-factor approximation for the problem of embedding general metrics into the line metric. For the case of the metrics which can be represented as trees, we provide improved approximation ratios in polynomial time as well as a QPTAS (Quasi-Polynomial Time Approximation Scheme).

1 Introduction

Metric embeddings have recently attracted much attention in theoretical computer science because of their many algorithmic applications. These range from simplifying the structure of the input data for approximation and online problems [5, 8, 9, 15, 18, 24], serving as a well-roundable relaxation of important NP-hard problems [7, 11–13, 17, 27] or simply by being the object of study [1, 16] arising from applications such as computational biology. Embedding techniques have thus become an indispensable addition to the algorithms toolbox, providing powerful and elegant solutions to many algorithmic problems (see, e.g., [29, Chapter 15] and [22]).

An embedding of a metric (V, d) into a “simpler” host metric (H, δ) is a map $f : V \rightarrow H$; the embedding is a good one if the distances between points in d closely resemble those between their images in δ . An embedding is called *non-contracting* if the map does not decrease any distances, i.e., $d(x, y) \leq \delta(f(x), f(y))$ ⁴ for all $x, y \in V$. We restrict ourselves to non-contracting embeddings in this paper. Perhaps the most popular and useful measure of the

^{*} Supported by NSF ITR grants CCR-0085982 and CCR-0122581.

^{**} Supported in part by NSF grant CCR-0105548 and ITR grant CCR-0122581.

⁴ In the sequel, we will abbreviate $\delta(f(x), f(y))$ to $\delta(x, y)$.

quality of an embedding f is the distortion $\alpha = \alpha(f)$, which is:

$$\text{distortion } \alpha = \max_{x,y \in V} \frac{\delta(x,y)}{d(x,y)}.$$

A closely related measure is that of *average distortion*, which is

$$\text{average distortion } \rho(f) = \frac{\sum_{x,y \in V} \delta(x,y)}{\sum_{x,y \in V} d(x,y)}.$$

While many embedding techniques and algorithms are known, the analyses for these embeddings usually only offer uniform bounds on the distortion of the embeddings; few results which *approximate* the distortion of the embeddings to better than these uniform bounds. This is best shown by a concrete example: Matoušek [28] proved that *any* metric (V, d) can be embedded into the real line with distortion $O(|V|)$; furthermore, the result is existentially tight, as the n -cycle cannot be embedded into the line with distortion $o(|V|)$ (see, e.g., [31, 21]). However, no algorithm is known which offers *per-instance* guarantees; hence, while it may be possible to embed (X, d) into \mathbb{R} with distortion $\alpha = O(1)$, no algorithms are known which give us embeddings with distortion, say, that is within $O(|V|^{1-\epsilon})$ times ρ ! No results are known even when we replace distortion α by average distortion ρ as the measure of goodness⁵.

1.1 Our Results

In this paper, we prove results for approximating the average distortion when embedding metrics into the line \mathbb{R} (while ensuring that the map is non-contracting). We can think of embeddings into a line as defining a tour on the nodes of the original metric. Note that for an embedding to be non-contracting, it is necessary and sufficient to have the distance between adjacent pair of vertices in the tour to be the same as their distance in the input metric. Our results demonstrate a close relationship between minimizing average distortion and the problems of finding short TSP tours [25], minimum latency tours [10, 20, 4], and optimal k -repairmen solutions[14]. In particular, we prove the following results.

- **Hardness for average distortion:** We prove that the problem of finding a minimum average distortion non-contracting embedding of finite metrics into the line is NP-hard, *even when the input metric is a tree metric*. This is proved via a reduction from the Minimum Latency Problem on trees [33].
- **Constant-factor approximations:** We give an algorithm that embeds any metric (V, d) into the line with average distortion that is within a constant of the minimum possible over all non-contracting embeddings. In fact, we prove a slightly more general bound on non-contracting embeddings into k -spiders (i.e., homeomorphs of stars with k leaves). This result uses a lower

⁵ One notable exception is the remark of Linial et al. [27] that the optimal embedding of any finite metric into (unbounded dimensional) Euclidean spaces to minimize distortion can be computed as a solution to a semi-definite program.

bound on the minimum average distortion of a non-contracting embedding into a k -spider in terms of the minimum k -repairmen tour [14] on the metric. We also show a tightened result for the case of 2-spiders using ideas from constructing minimum latency tours [20].

- **QPTAS on trees:** For tree metrics on n nodes, we give an algorithm for finding a $(1 + \epsilon)$ -approximation to the minimum average distortion non-contracting embedding into a line in $n^{O(\log n/\epsilon^2)}$ time. Our algorithm uses a lower bound on the minimum average distortion related to the TSP tour length and latencies of appropriately chosen segments of an optimal tour. In this way, it extends the ideas of Arora & Karakostas [6] for minimizing latency on trees to the more general time-dependent TSPs [10] to provide a QPTAS for the latter problem as well.

Given a tree metric as input, if the minimum average distortion is measured only over the endpoints of the edges of the tree (we call this objective the average tree-edge distortion), we can prove that an embedding following an Euler tour of the tree is optimal. This tour can be found in polynomial time by dynamic programming. We omit the description of this algorithm due to lack of space.

1.2 Related Work

The definition of average distortion is by no means new; e.g., Alon et al. [2] study the question of embedding a metric into a tree with low average distortion. In recent work on average distortion that is closer to our work, Rabinovich [30] proves bounds on average distortion of *non-expanding* embeddings into a line and shows the close connection between this and the max-flow min-cut ratio for concurrent multicommodity flow with applications to finding quotient cuts in graphs [26].

Our problem is similar to that of finding the Minimum Linear Arrangement (MLA), for which Rao & Richa [32] gave an $O(\log n)$ approximation using the notion of spreading metrics. However, while the MLA problem involves minimizing the average stretch of the edges $\sum_{\{u,v\} \in E} |\pi(u) - \pi(v)|$ under all maps $\pi : V \rightarrow [n]$, the mappings in our problem are $f : V \rightarrow \mathbb{R}$, and must ensure that $|f(u) - f(v)| \geq d(u, v) \forall \{u, v\} \in V \times V$.

The problem of finding Minimum Latency tours (a.k.a. the traveling repairman problem) is most relevant to our discussion in terms of techniques used. This problem requires a repairman who starts from a depot on a given finite metric to visit n customers, one at each node of the metric; his goal is to minimize the *average waiting time* or *latency* of the customers, where the waiting time of a customer is the sum of the distances of all edges traversed by the repairman before visiting this customer. The version of this problem with only one repairman (also called the Minimum Latency Problem) is known to be NP-hard even on trees [33] and MAX-SNP hard in general [10]. The first constant-factor approximation for this problem was given by Blum *et al.*[10], which was subsequently improved by Goemans and Kleinberg [20] to the currently best-known

bound of 7.18. Recently, Archer, Levin and Williamson [4, 3] gave faster algorithms obtaining very similar approximation guarantees. For the special cases of the latency problem on trees, and in \mathbb{R}^d for fixed dimension d , Arora and Karakostas [6] gave quasi-polynomial time approximation schemes (QPTAS). The extension of the latency problem to more than one repairmen was recently studied in [14] where the authors show a 16.994-approximation for the general k -repairman case.

Finally, a problem that generalizes both the cost of a tour as well as its latency into one objective is that of finding time dependent TSP tours. A constant factor approximation algorithm is also known for this problem [10].

Outline: The rest of the paper is organized as follows. In Section 2, we argue that the embedding problem is NP-hard, and give the constant-factor approximation algorithm for embedding metrics into the line with constant average distortion. Section 3 shows the QPTAS for the case of trees metrics as inputs.

2 Embedding arbitrary metrics into the line

In this section, we show that we can approximate the average distortion into a line for a given metric to within a constant; to this end, we show that the problem is closely related to that of finding the minimum latency tours and its generalizations in a finite metric space. We omit the proof of the following theorem; the reduction is from Minimum Latency on trees.

Theorem 1. *It is NP-hard to find a non-contracting embedding of a given metric induced by a tree into a line that minimizes the average distortion.*

First, we show a simple 2-approximation for embedding a finite metric into a special kind of tree metric, namely a k -spider. (A k -spider is a tree with all vertices except the *center* having degrees 1 or 2, and hence is a homeomorph of the star with k leaves). The case of a n -spider or a complete star is more natural to argue about, while the 2-spider is a path giving our main result.

Embeddings into trees Consider the problem of embedding the given metric d into a tree metric δ to minimize average distortion. Let $\Delta = \sum_{x,y \in V} d(x,y)$ denote the sum of all the distances in the metric d , and hence $\text{av}(d) = \Delta/n^2$ is the average distance in d . The *median* of the metric d is the point $v \in V$ that minimizes $\Delta_v = \sum_{w \in V} d(v,w)$, and will be denoted by med . Note that we can decompose Δ as follows:

$$\Delta = \sum_{u,v \in V} d(u,v) = \sum_{u \in V} (\sum_{v \in V} d(u,v)) = \sum_{u \in V} \Delta_u \geq n \Delta_{med} \quad (1)$$

since $\Delta_{med} \leq \Delta_v$ for all $v \in V$. Consider a shortest-path tree T (which is a star in a general metric d) rooted at med , and let d_T denote the metric induced by this shortest path tree. Then the total distance in this tree T is

$$\begin{aligned} \Delta_T &= n^2 \cdot \text{av}(d_T) = \sum_{u,v \in V} d_T(u,v) \leq \sum_{u,v \in V} d_T(\text{med}, u) + d_T(\text{med}, v) \\ &= \sum_{u,v \in V} d(\text{med}, u) + d(\text{med}, v) = 2n \Delta_{med} \end{aligned}$$

where the inequality in the second step is just the triangle inequality. This implies that $n\Delta_{med} \leq \Delta \leq \Delta_T \leq 2n\Delta_{med}$, and thus:

Lemma 1 (See also [34]). *Given any graph, the average distance Δ_T for the tree rooted at the median is at most 2Δ , and is a 2-approximation for the problem of embedding the graph into trees.*

Note here that the bound of 2 above is an *absolute* bound on the worst-case ratio between the average distance in the output tree and the graph, and is in the same flavor as the more traditional results on bounding the maximum distortion of embeddings. We next move toward an approximation approach by restricting the class of trees into which we embed.

Embeddings into spiders We now generalize the previous result to the case of embeddings into k -spiders. The vertex of degree k is called the *center* of the spider, and the components obtained by removing the center are called its *legs* [23].

Let d_k^* denote the optimal k -spider embedding. We decompose the sum of distances in d_k^* as the sum of k -repairman paths rooted at each vertex. Recall that, in k -traveling repairman problem, we are given k repairmen starting at a common depot s . The k repairmen are to visit n customers sitting one per node of the input metric space. The goal is to find tours on which to send the repairmen so as to minimize the total time customers have to wait for a repairman to arrive [14].

Let c be the center of the spider in the optimal k -spider embedding. To construct a k -repairman paths starting from a vertex r , we do the following. We send one repairman away from the center along the leg of the spider which contains r . The other $k - 1$ repairmen travel toward the center c of the spider. From the center, they go off, one per remaining leg of the spider. The cost of this k -repairman tour is $\Delta_r^* = \sum_j d_k^*(r, j)$. Summing over all choices of the root we see that this is same as the sum of distances in the embedding d_k^* .

$$\sum_{v \in V} \Delta_v^* = n^2 \cdot \text{av}(d_k^*)$$

Hence, n times the cost of the cheapest k -repairman tour over all choices of the depots (denoted by Δ^{opt}), is a lower bound on the sum of all the distances. i.e.,

$$\sum_{u, v \in V} d_k^*(u, v) \geq n \cdot \min_r \{\Delta_r^{opt}\}.$$

Consider the cheapest k -repairman tour over all choices of centers. Let it be centered at a vertex c . This tour defines a non-contracting embedding into a k -spider with c at the center of the spider. Let $d^c(u)$ denote the distance of vertex u from the center c in the tour. We can bound the sum of distances in this embedding as follows:

$$\sum_{u, v \in V} d_k^c(u, v) \leq \sum_{u, v \in V} d^c(u) + d^c(v) \leq 2n \sum_{u \in V} d^c(u) \leq 2 \sum_{u, v \in V} d_k^*(u, v).$$

Thus, if we could compute the optimal k -repairman tour centered at c exactly, we would obtain a 2-approximation to the problem of embedding the metric

into k -spiders. Although the problem of finding an optimal k -repairman tour is NP-hard, the argument above proves the following.

Theorem 2. *Given a γ -approximation for the minimum k -repairmen problem on a metric d , we can obtain a 2γ -approximation for embedding the metric d into a k -spider in a non-contracting fashion to minimize the average distortion.*

The current best known approximation factor for the k -repairman problem is about 17 (due to Fakcharoenphol et al. [14]), leading to the following corollary.

Corollary 1. *There is a 34-approximation for minimizing the average distortion of a non-contracting embedding of a given finite metric into a k -spider.*

Embeddings into a line: Improved guarantee We can get a better approximation factor for embeddings into the line by employing a slightly different strategy. Instead of using the result of Fakcharoenphol *et al.* as a black box, we instead give an algorithm to find a 1-repairman tour (i.e., a minimum latency tour) that is within a factor of 14.36 of the optimum 2-repairmen tour in the given metric. Since a 1-repairman tour is also a 2-repairmen tour (with the second repairman doing nothing), we can then apply Theorem 2 to bring down the overall approximation ratio to 28.72.

The idea behind the algorithm is the same as in scaled search, due to Blum *et al.* [10]; here is an outline. To find a 1-repairman solution centered at r :

```

for  $j = 0, 1, 2, 3, \dots$ , do
     $T_j \leftarrow$  tree rooted at  $r$  spanning the most vertices among those
    with cost  $\leq 2^{j+2}$ .
Concatenate Euler tours of the trees  $T_j$  (in increasing order of  $j$ ), to form
a 1-repairman path.

```

Lemma 2. *The cost of the 1-repairman tour produced by the preceding algorithm is within a factor 32 of the cheapest 2-repairman tour.*

Proof. Let vertex v be the i^{th} closest vertex to root r in the optimal 2-repairman tour. Let the distance of v from the root r in the tour be between $[2^j, 2^{j+1})$ in the optimal solution. Consider the tree T_j of cost 2^{j+2} constructed by our algorithm. We claim that T_j spans at least i vertices. Thus cost of i^{th} vertex in our tour has latency at most

$$\sum_{i=0}^j (\text{cost of } i^{\text{th}} \text{ tour}) \leq \sum_{i=0}^j 2 \cdot 2^{i+2} \leq 2^{j+4}$$

Hence, the distance of the i^{th} vertex in our 1-repairman tour is at most 16 times its counterpart in the optimal 2-repairmen tour.

Although the problem of finding the largest tree with cost at most 2^{j+2} is NP-hard, we can find a tree having as many vertices as the this optimal tree instead (but with cost at most $2 \cdot 2^{j+2}$ using Garg's [19] algorithm for i -MST). This increases the overall approximation factor to $16 \cdot 2 = 32$.

Lemma 3. *We can find a 1-repairman tour with cost ≤ 14.36 times the cost of the cheapest 2-repairman tour.*

Proof. (Sketch) Let b be a real number greater than 1 to be chosen later. Let $c = b^U$, where U is a real number chosen uniformly at random from the interval $[0, 1]$. Instead of finding the trees of cost $2, 4, 8, \dots$ which cover the most vertices, we will find the trees of cost at most $2c, 2cb, 2cb^2, \dots$ which cover the most vertices. Using the methods of Goemans and Kleinberg, we can show that the approximation ratio of the previous proof can be improved to 14.36.

Note that this improves on the result of Fakcharoenphol et al. [14] for the special case of the 2-repairman problem. An application of Theorem 2 now gives us the following:

Theorem 3. *There exists a 28.72-approximation algorithm to embed a given (weighted) metric into a line in a non-contracting fashion to minimize the average distortion.*

As a consequence of the analysis in Lemma 2, we also get the following result:

Lemma 4. *For $l \leq k$, we can find an l -repairman tour with cost at most $17(k/l)$ times that of the optimal k -repairman tour.*

We note that the factor $\frac{k}{l}$ in the above Lemma is necessary as demonstrated by the metric induced by an unweighted star graph. Compare the above result to that of Fakcharoenphol et al. [14] which outputs a k -repairmen tour of cost $O(\frac{k}{l})$ times the minimum l -repairmen tour for $k \geq l$ (where the factor $\frac{k}{l}$ is not necessary since the algorithm delivers a solution with more repairmen than the optimal compared against).

3 Approximation Schemes for trees

In this section, we restrict our attention to the special case of tree metrics. We give a quasi-polynomial time approximation scheme for minimizing the average distortion for embeddings into the line metric. Our algorithm is based on the QPTAS given by Arora and Karakostas for the minimum latency problem [6]. They proved that a near-optimal latency tour can be constructed by concatenating $O(\log |V|/\epsilon)$ optimal TSP subtours, and the best such solution can be found by dynamic programming.

For an embedding $f : V \rightarrow \mathbb{R}$ into the line, let the *span* of the embedding be defined as $\max_{x,y} |f(x) - f(y)|$, the maximum distance between two points on the line. We note that an embedding with the shortest span is just the optimal TSP tour. While embedding a given metric into the line metric, minimizing the span of the embedding could result in very high average distortion. However, we show that it suffices to minimize the span locally to find near optimal embedding. In particular, our solution within $(1 + \epsilon)$ of optimal minimum average distortion is to find an embedding that is the union of $O(\log |V|/\epsilon^2)$ TSP tours with geometrically decreasing number of vertices.

In the sequel, we use n to denote $|V|$, the number of vertices. For our algorithm, we assume that all the edge lengths are in the range $[1, n^2/\epsilon]$. Indeed, if D is the diameter of the metric space and u and v are two vertices such that $d(u, v) = D$, then $\sum_{x, y \in V} d(x, y) \geq \sum_{x \in V} d(x, u) + d(x, v) \geq nD$. We can then merge all pairs of nodes with inter-node distance at most $\epsilon D/n^2$, which affects the sum of distance by at most ϵnD . Hence the ratio of maximum to minimum nonzero distance in the metric can be assumed to be n^2/ϵ .

Relation to TDTSPs We first show that the Arora-Karakostas QPTAS works also for the case of Time Dependent Traveling Salesman Problem (TDTSP) defined by Blum *et al.*. In the TDTSP, the objective is to minimize a positive linear combination of the TSP tour value and the total latency of the tour. The intuition behind this is that adding a component of TSP in the objective value preserves the property that the tour composed of TSP tours continues to remain near-optimal.

We now describe how to break up an optimal tour into locally optimal segments. Let \mathcal{T} denote the optimal tour for the objective function $\alpha TSP + \beta LAT$ where TSP and LAT denote the span and latency objective values of the tour respectively. We break this tour into k segments (k depends on the input parameter ϵ). In segment i we visit n_i nodes, where

$$n_i = \lceil (1 + \epsilon)^{k-1-i} \rceil \text{ for } i = 1, \dots, k-1; \quad n_k = \lceil 1/\epsilon \rceil$$

Note that these n_i 's are chosen in such a way that $n_i \leq \epsilon \sum_{j>i} n_j$. Denote $\sum_{j>i} n_j$ by r_i . Replace the optimal tour in each segment, except the last one, by the minimum-distance traveling salesman tour for that segment. The new tour now consists of the concatenation of $O(\log n/\epsilon)$ locally optimal TSP tours. This gives us the following lemma.

Lemma 5. *There is a tour that is a concatenation of $O(\log n/\epsilon)$ TSP tours that has $\alpha TSP + \beta LAT$ objective value at most $(1 + \epsilon)$ times the minimum.*

We now use the Lemma 5 to show the following theorem for average distance.

Theorem 4. *Any finite metric has a non-contracting embedding into a line that is composed of $O(\log n/\epsilon^2)$ minimum TSP tour segments with average distortion no more than $(1 + \epsilon)$ times the minimum possible over all such embeddings.*

Proof. Our strategy is same as in Lemma 5. Consider the optimal embedding of the input tree into a line. We break this embedding up into $O(\log n/\epsilon)$ segments. Let n_i be the size of i th segment defined as before. We now divide the objective function value according to the segments, so that only the share C_i of segment i changes, if we replace the embedding of segment i with a different embedding.

Let T_i be the length of the embedding of segment i . If i_0 is the left-most node in the embedding of the segment i , then let $L_i = \sum_{j \in n_i} l(i_0, j)$ be the sum of the distances of all nodes in segment i from node v . Note that L_i is the total

latency of vertices in segment i with i_0 as root. And let $D_i = \sum_{u,v \in n_i} l(u,v)$ be the sum of all the pairwise distances in segment i .

Let $q_i = \sum_{j < i} n_j$ and $r_i = \sum_{j > i} n_j$ be the number of total nodes to the left and right of segment i respectively.

The contribution of the segment i to the objective comes from the following distinct terms.

1. If a vertex u is to the left of the segment i and a vertex v is to the right, then the segment i adds T_i to the distance between them.
2. If a vertex u is to the left and w is in the segment i , then the contribution is $l(i_0, w) =$ the distance from the left most vertex i_0 of the segment i to w .
3. If a vertex v is to the right and w is in the segment i , then the contribution is $T_i - l(i_0, w)$.
4. If both the vertices w and w' are in the segment i , then the contribution is $l(w, w')$.

These contributions, when summed up over all pairs of vertices, give:

$$C_i = q_i r_i T_i + q_i L_i + r_i (n_i T_i - L_i) + D_i \quad (2)$$

Note that $D_i \leq n_i^2 T_i$. For $i = 2, \dots, k$, we know that $n_i \leq q_i$ and $n_i \leq \epsilon \cdot r_i$. Hence, comparing D_i with the first term in (2), we get

$$(1 + \epsilon)(q_i r_i T_i + q_i L_i + r_i (n_i T_i - L_i)) \geq C_i \geq q_i r_i T_i + q_i L_i + r_i (n_i T_i - L_i) \quad (3)$$

To prove the statement in Theorem 4, it suffices to find a tour that is within $(1 + \epsilon)$ of the lower bound in the RHS of the above inequality 3. The expression for the lower bound on the RHS of inequality 3 is a linear combination of TSP and Latency values of the tour in segment i . We can apply Lemma 5 to obtain a tour composed of $O(\log n_i / \epsilon)$ TSP tours. This tour is within $(1 + \epsilon)$ factor of the lower bound on C_i .

A technical detail in this argument is that the coefficient of L_i could be negative. Lemma 5 does not handle this case. But note that $n_i T_i - L_i$ is the total ‘‘reverse’’ latency in segment i with the rightmost endpoint being the root. Thus we can rewrite the lower bound as a linear combination of T_i and $n_i T_i - L_i$ with positive coefficients.

We can thus replace each segment i , with a concatenation of $O(\log n_i / \epsilon)$ TSP tours, without increasing the cost by more than a factor of $(1 + \epsilon)$. Since there are $O(\log n / \epsilon)$ segments in all, it follows that there is an embedding consisting of $O(\log^2 n / \epsilon^2)$ shortest TSP tours.

Finally, we show how to reduce this number down to $O(\log n / \epsilon^2)$. Let us rewrite the lower bound in (3) as $(q_i - r_i)L_i + (q_i + n_i)r_i T_i$. Note that $L_i \leq n_i T_i$. This gives us that the term $(q_i - r_i)L_i$ is at most $\epsilon \cdot (q_i + n_i)r_i T_i$, whenever $q_i - r_i$ is positive. Hence, if we replace the segment i with a shortest TSP tour on those vertices, the cost will be within $(1 + \epsilon)$ of the lower bound in (3). It is easy to check that, for $i \geq 1/\epsilon$, we have $q_i \geq r_i$. Hence for $i = 1, \dots, 1/\epsilon$, using Lemma 5, we replace each segment by a concatenation of $O(\log n / \epsilon)$ tours each.

Then for the segments i and above, we use only one minimum TSP tour. Overall this results in a concatenation of $O(\log n/\epsilon^2)$ tours with near-optimal average distortion.

Note that, an optimal TSP tour of the tree is an Euler tour. In other words, each edge is crossed exactly twice, once in each direction. As a consequence, we have the following.

Theorem 5. *There exists a non-contracting embedding of a tree metric into a line with average distortion at most $(1 + \epsilon)$ times the minimum possible that, when viewed as a walk, crosses every tree edge $O(\log n/\epsilon^2)$ times.*

Now using dynamic programming using these structural results proves the following theorem.

Theorem 6. *For any given $\epsilon > 0$, there is an algorithm that runs in time $n^{O(\log n/\epsilon^2)}$ and computes a non-contracting embedding of a given input tree metric into a line with average distortion at most $(1 + \epsilon)$ -times the minimum.*

Proof. (Sketch)

We now describe the quasi-polynomial-time approximation scheme based on dynamic programming. Theorem 5 can be restated in terms of crossings of vertices. Consider a separator vertex for the tree. We will denote the partition of the tree at the centroid as the left and right parts. There exists a near optimal embedding that, when viewed as a tour, crosses the separator node from left half to right half $O(\log n/\epsilon^2)$ times.

We develop a dynamic program based on the above observation. Given the input tree, we try each vertex as the starting point of our tour. In order to compute the tour, we first find a separator node in the tree. For the dynamic program, we maintain the following state space. Consider the sub-tours formed between successive places where we cross the separator node. We guess the number of nodes and the length for each of these sub-tours. Note that since there are only $O(\log n/\epsilon^2)$ crossings, there are only $n^{O(\log n/\epsilon^2)}$ choices for the number of nodes. Moreover, the length of each tour can take at most $O(\log n/\epsilon)$ different values. Thus the number of choices for the length are bounded by about $O((\log n)^{\log n})$. Thus the total size of state space is $n^{O(\log n/\epsilon^2)}$. Finding the best tour given the lengths of sub-tours can be done by recursing on the left and right parts independently. For each of these sub-tours, we want to visit all the vertices while staying on one side throughout. The total running time of this procedure is $n^{O(\log n/\epsilon^2)}$.

4 Open Problems and Discussion

It is important to note that a non-contracting embedding can be converted to a non-expanding embedding by scaling down all the distances. However, the converse is not true, since in non-expanding embeddings, the host metric could be a semi-metric. In other words, mapping two points in the guest metric to

the same point in the host metric is allowed. This represents a crucial difference between the two problems.

For the case of non-contracting embeddings considered in the paper, here are some open questions :

- (1) Is there a simpler and better approximation algorithm for minimizing average distortion in trees?
- (2) Can the Quasi-PTAS be extended to (outer)planar graphs?
- (3) A different but related objective function is sum of the distortions of all pairs over all non-contracting embeddings. Are there approximation algorithms for this objective?
- (4) For the case of weighted average distance, we can write a linear program based on the spreading metric LP for minimum linear arrangement (à la Rao & Richa [32]). However, the integrality gap of this LP is as yet unknown.

References

1. Richa Agarwala, Vineet Bafna, Martin Farach, Babu O. Narayanan, Mike Paterson, and Mikkel Thorup. On the approximability of numerical taxonomy (fitting distances by tree metrics). In *SODA*, pages 365–372, 1996.
2. Noga Alon, Richard M. Karp, David Peleg, and Douglas West. A graph-theoretic game and its application to the k-server problem. *SIAM J. Comput.*, 24(1):78–100, 1995.
3. Aaron Archer, Asaf Levin, and David P. Williamson. A faster, better approximation algorithm for the minimum latency problem. 2003.
4. Aaron Archer and David P. Williamson. Faster approximation algorithms for the minimum latency problem. In *SODA*, 2003.
5. Sanjeev Arora. Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. *Journal of the ACM*, 45(5):753–782, 1998.
6. Sanjeev Arora and George Karakostas. Approximation schemes for minimum latency problems. In *Proceedings of the ACM STOC*, pages 688–693, 1999.
7. Yonatan Aumann and Yuval Rabani. An $o(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301, 1998.
8. Yair Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *IEEE FOCS*, pages 184–193, 1996.
9. Yair Bartal, Avrim Blum, Carl Burch, and Andrew Tomkins. A polylog(n)-competitive algorithm for metrical task systems. In *Proceedings of STOC*, pages 711–719, 1997.
10. Avrim Blum, Prasad Chalasani, Don Coppersmith, Bill Pulleyblank, Prabhakar Raghavan, and Madhu Sudan. The minimum latency problem. In *Proceedings of the ACM STOC*, pages 163–171, 1994.
11. Avrim Blum, Goran Konjevod, R. Ravi, and Santosh Vempala. Semi-definite relaxations for minimum bandwidth and other vertex-ordering problems. In *Proceedings of the 30th ACM STOC*, pages 100–105, 1998.
12. J. Bourgain. On lipshitz embedding of finite metric spaces in hilbert space. *Israel J. Math.*, 52:46–52, 1985.
13. Gruia Calinescu, Howard J. Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *SODA*, pages 8–16, 2001.

14. Jittat Fakcharoenphol, Chris Harrelson, and Satish Rao. The k-traveling repairman problem. In *SODA: 14th ACM-SIAM Symposium on Discrete Algorithms*, 2003.
15. Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the 35th Annual ACM STOC*, pages 448–455, 2003.
16. Martin Farach, Sampath Kannan, and Tandy Warnow. A robust model for finding optimal evolutionary trees. *Algorithmica*, 13(1/2):155–179, 1995.
17. Uriel Feige. Approximating the bandwidth via volume respecting embeddings (extended abstract). In *Proc. 30th ACM STOC*, pages 90–99, 1998.
18. Garg, Konjevod, and Ravi. A polylogarithmic approximation algorithm for the group steiner tree problem. In *SODA*, 1998.
19. N. Garg. Personal communication, Spetember 2000.
20. Goemans and Kleinberg. An improved approximation ratio for the minimum latency problem. In *SODA*, 1996.
21. Anupam Gupta. Steiner nodes in trees don't (really) help. In *SODA*, 2001.
22. Piotr Indyk. Algorithmic aspects of geometric embeddings. In *IEEE FOCS*, 2001.
23. Philip N. Klein and R. Ravi. A nearly best-possible approximation algorithm for node-weighted steiner trees. *J. Algorithms*, 19(1):104–115, 1995.
24. Jon M. Kleinberg and Eva Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and markov random fields. In *IEEE FOCS*, pages 14–23, 1999.
25. E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and editors D. B. Shmoys. *The Traveling Salesman Problem*. John Wiley & Sons, 1985.
26. F. T. Leighton and S. Rao. An approximate max-flow mincut theorem for uniform multicommodity flow problems with applications to approximation algorithms. In *Proc. of the 29th IEEE FOCS*, pages 422–431, 1988.
27. Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15:215–245, 1995.
28. J. Matoušek. Bi-lipschitz embeddings into low dimensional euclidean spaces. *Comment. Math. Univ. Carolinae*, 31(3):589–600, 1990.
29. J. Matoušek. *Lectures on Discrete Geometry*. Springer-Verlag, 2002.
30. Y. Rabinovich. On average distortion of embedding metrics into l_1 and into the line. In *35th Annual ACM STOC*, 2003.
31. Y. Rabinovich and R. Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *GEOMETRY: Discrete & Computational Geometry*, 19, 1998.
32. Satish Rao and Andrea Richa. New approximation techniques for some ordering problems. In *SODA*, 1998.
33. R.A. Sitters. The minimum latency problem is np-hard for weighted trees. In *W.J. Cook, A.S. Schulz (eds.), Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science 2337*, pages 230–239, 2002.
34. R.T. Wong. Worst-case analysis of network design problem heuristics. *SIAM Journal Alg. Disc. Math*, 1(1):51–63, 1980.