On Approximating Arbitrary Metrics by Tree Metrics

Yair Bartal *

Abstract

This paper is concerned with probabilistic approximation of metric spaces. In previous work we introduced the method of efficient approximation of metrics by more simple families of metrics in a probabilistic fashion. In particular we study probabilistic approximations of arbitrary metric spaces by "hierarchically well-separated tree" metric spaces. This has proved as a useful technique for simplifying the solutions to various problems.

In this paper we improve the result by proving an approximation factor of $O(\log n \log\log n)$ getting the gap to the lower bound within lower order factors. We also give a deterministic version of the result which gives a tree with low average distortion of distances.

The results have applications in a variety of areas including approximation algorithms, on-line algorithms and distributed computation and hence we obtain new approximation bounds for these applications.

1 Introduction

1.1 Probabilistic Metric Approximations and HSTs

We study the problem of approximating arbitrary metric spaces by distances in more simple metric spaces. This allows reducing the solution of problems to the more simple spaces.

Approximations of metric spaces by more simple metric spaces has been intensively studied in various areas of mathematics and computer science. We mention but a few. Johnson and Lindenstrauss [JL84] and Bourgain [Bour85] study embeddings in Hilbert spaces (from a perspective of functional analysis). Graham and Winkler [GW85] study embeddings in $\mathbb{Z}^d$ (from a graph theoretic motivation). Algorithmic applications in distributed computation and graph algorithms, have led to the notion of graph spanners, introduced by Peleg and Ullman [PU89] and later studied in many papers including [PS89, ADDJS90, CDNS92], and to low-distortion embeddings in low-dimensional real normed spaces by Linial, London and Rabinovich [LLR94].

In [Bart96] we define *probabilistic approximation* of metric spaces by a set of "simpler" metric spaces $\mathcal{S}$. Given a metric space M over a finite set V, we consider sets $\mathcal{S}$ of "simple" metric spaces over V such that distances in M are dominated by the corresponding distances in metric spaces in $\mathcal{S}$. M is said to be α-*probabilistically-approximated* by $\mathcal{S}$, if there exists a probability distribution over $\mathcal{S}$ such that for *every* pair of nodes in V, the expected ratio between the distance in a metric space in $\mathcal{S}$ chosen from the probability distribution, and the distance in M is at most α.

Regarding a randomized algorithm as a distribution over deterministic algorithms, its performance ratio is the expected performance ratio of these algorithms over its own coin tosses. The probabilistic metric approximation approach provides a *novel technique* for bounding the performance ratio of *randomized* algorithms for optimization problems on metric spaces where a solution can be expressed as a linear combination of distances in the metric space. If every metric space is α-probabilistically-approximated by $\mathcal{S}$, and the performance ratio of randomized algorithms for $\mathcal{S}$ is at most β, then the performance ratio of randomized algorithms for any metric space is at most $\alpha\beta$.

The advantage of using probabilistic metric space approximation is that certain classes of simple metric spaces which cannot provide a good deterministic approximation of arbitrary metric spaces may provide a good probabilistic approximation.

A natural candidate for a class of "simple" metric spaces is the class of tree metric spaces. Indeed, inten-

*Bell-Labs, Lucent Technologies, 600 Mountain Avenue, Murray Hill, NJ 07974. *E-mail:* yair@research.bell-labs.com

sive recent research has been concerned with approximating metric spaces by tree metrics, including work on *numerical taxonomy* [BG92, ABFNPT96] and *tree spanners* [Cai92, CC95].

However, there exist simple metric spaces for which every specific tree dominating the distances of the metric space has distortion $\Omega(n)$ (trivially matched by the upper bound given by MST in a weighted graph defined by the metric space). In particular this is true for the n-cycle as follows from the recent work of Rabinovich and Raz [RR95]. This justifies the idea of *probabilistic* metric space approximations.

Seminal work along these lines begun with Karp's observation [Karp89] that the n-cycle can be 2 probabilistically approximated by n-line spaces, motivated by its application to the K-server problem. Alon, Karp, Peleg and West [AKPW91] have taken this idea forward in considering probabilistic-approximations of graphs by spanning trees. Their result implies that every metric space can be $2^{O(\sqrt{\log n \log\log n})}$ probabilistically approximated by a tree metric.

The approach taken in [Bart96] exploits the fact that the class of metric spaces used for the approximation need not be induced by subgraphs. This allows us to consider a class of metric spaces that are tree metrics and also have other desirable properties. This class of tree metric spaces are called *k-hierarchically well-separated trees* (k-HST). They have the property that the metric space can be decomposed into submetric spaces each of diameter smaller by a factor of at least k (for some parameter $k > 1$) from the diameter of the complete metric space, and this property holds recursively. This property may allow a divide-and-conquer approach to the solution of particular optimization problem. It is worth to comment that even though the trees we contruct are not contrained to be subgraphs of the oriinal graph if we have a complete graph with edge weights corresponding to a metric (a "metric graph") this property can be easily guarantied.

The main result of [Bart96] states that for any k, every metric space M over a set V of size n can be α-probabilistically-approximated by k-HST's, where $\alpha = O(k \log n \log_k n)$.

In this paper we improve the approximation factor to $\alpha = O(k \log n \log\log n)$.

We note that in [KRS98] it has been shown that for graphs without small $K_{s,s}$-minors (for constant s) the approximation factor is $O(k \log n)$.

1.2 Deterministic Approximations

The problem of finding a good probabilistic metric approximation can be viewed as a game between a "metric player" who is trying to provide the approximating metric space and an "edge player" who wishes to find the pair that receives the worst distortion under the given metric. Probabilistic approximation of a metric space corresponds to a mixed strategy of the metric player.

It follows that an equivalent problem is to find a single metric space that minimizes a weighted average of the distortions of all edges. Indeed in this paper we study both problems simultaneously and give algorithmic results for both.

The deterministic result can be equivalently formulated as follows: given capacities on the edges and edge lengths find a metric space that preserves the total volume up to some approximation factor. One place where this version of the method comes up is when an off-line optimization problem can be formulated as an LP and the objective function forms a metric. In that case the metric can be transformed into a tree metric, and then the solution may be rounded on the tree. This approach has been recently applied by Charikar, Chandra, Goel and Guha [CCGG98] to derandomize some of the applications discussed in the sequel.

The duality of the two problems described above can be formulated in terms of Linear Programming. In recent work by Charikar, Chandra, Goel, Guha and Plotkin [CCGGP98], it is shown that standard Linear Programming techniques can be used to provide a different construction of a probabilistic approximation of metrics by trees that promises the number of trees used to be polynomial. Thus this implies that by computing solutions for all these trees and choosing the best solution one can obtain deterministic approximation algorithms.

1.3 Hierarchical Partition Metrics and HSTs

Our new construction is based on what we call a hierarchical partition metric of a weighted graph which is an assignment of labels to the edge in a manner so that the graph can be separated into clusters so that the edges in the cut are labeled with weight higher than the diameter. This structure is then shown to be close to an HST up to a constant factor. The HPMs are constructed using a similar approach to that of Seymour [Seym95] later used in the context of *spreading metrics* by Even, Naor, Rao and Schieber [ENRS95] to obtain approximation algorithms for a large set of optimization problems.

1.4 Applications

As mentioned the HST approximation theorems have already found several applications in a variety of fields. Our results obtain a further improvement on the ratios obtained for the particular problems.

1.4.1 Approximation Algorithms

• *The k-median* problem. Given a metric space find k median locations so that if every node is assigned to the nearest median point the total sum of distances is minimized. Optimal algorithms are known for trees [KH79, Tami96] and therefore a randomized $O(\log n \log\log n)$-approximation algorithm follows directly. In fact, a simple reduction can be applied to achieve $O(\log k \log\log k)$ [CCGG98]. No previous approximation result was formerly known for this problem.
• *The group Steiner tree* problem. Given a weighted graph and a collection of k sets, find a Steiner tree connecting at least one element from each set. Garg, Konjevod and Ravi [GKR98] give an $O(\log k \log n)$ approximation algorithm for trees and thus we obtain a randomized $O(\log k \log^2 n \log\log n)$ approximation algorithm.
• *Buy-at-bulk network design*. Given a weighted graph and a sub-linear function of the cost describing the edge cost as a function of the load on it. We are given a sequence of pairs of nodes. For each pair we should establish a path connecting them. The goal is to minimize the total load. Awerbuch and Azar [AA97] give an $O(1)$-approximation algorithm on trees and thus for general graphs we obtain a randomized $O(\log n \log\log n)$ approximation ratio. The algorithm is also on-line.
• *Vehicle Routing*. Given a weighted graph and a sequence of commodities, each associated with a source-destination pair. The problem is to construct a tour of minimal length for a vehicle such that every commodity is picked up from its source and delivered to its destination under the constraint that at most k commodities can be carried by the vehicle at any point in time. Charikar et al. [CCGG98] give an $O(1)$-approximation algorithm on trees which implies for general graphs a randomized $O(\log n \log\log n)$ approximation ratio.
• *The communication spanning tree* [Hu74] problem on metric graphs. Given a network with costs on the edges The communication spanning tree problem is to find a spanning tree as to minimize a weighted sum of the tree distances over all pairs of nodes. As mentioned before our result can be adjusted to obtain a spanning tree if the original network is a metric graph. The application for communication spanning tree has been observed in [WLBCRT98]. Since this problem is trivial for tree metrics we get an $O(\log n \log\log n)$ approximation ratio for arbitrary metric graphs.

All the results above can be derandomized using the techniques of Charikar et al. [CCGG98, CCGGP98].

1.4.2 On-line Algorithms

Probabilistic approximation of metric spaces is of particular importance in the case of *on-line* problems, where the randomization against *oblivious* adversaries (I.e., adversaries that cannot see the algorithm's coin tosses [BBKTW90]) is very powerful.
• *Metrical task systems*, due to Borodin, Linial and Saks [BLS87], consist of a set of states forming a metric space and a set of tasks associated with costs in the different states. The goal is to schedule state transitions in order to minimize the total move and task costs. The problem has deterministic competitive ratio of $2n-1$. The metrical task system problem is closely related to the the $(n-1)$-*server problem* [MMS88] on n points. Both upper and lower bounds for one problem translate to the other. Bartal, Blum, Burch and Tomkins [BBBT97] give an $O(\log^2 n)$-competitive algorithm for $\Omega(\log^2 n)$-HSTs and thus we get a randomized $O(\log^5 n \log\log n)$-competitive algorithm for arbitrary spaces.
• *Distributed paging*. The *constrained file migration* problem, due to Bartal, Fiat and Rabani [BFR92], is the problem of migrating files in a network with limited memory capacity at the processors in order to minimize file access and migration costs. This is a special case of the *distributed paging* problem [BFR92, ABF93b, ABF96] where files may also be replicated. In [Bart96] an $O(\log m)$-competitive algorithm for 2-HSTs is given. Thus we get a randomized algorithm for general networks with competitive ratio of $O(\log m \log n \log\log n)$.
• *The K-server problem*, due to Mannasse, McGeoch, and Sleator [MMS88] consists of K-servers in a metric space. Points are requested over time and a server must be moved to the request location. The K-server problem has been extensively studied. Koutsoupias and Papadimitriou [KP94] prove an upper bound of $2K-1$ on the competitive ratio of WFA (the work function algorithm). However WFA requires heavy time and space complexity. Therefore, it may be preferable to have a more efficient algorithm with a higher competitive ratio. By applying our method on the *memoryless* K-server algorithm for trees of Chrobak and Larmore [CL91] we obtain an efficient $O(K \log n \log\log n)$-competitive randomized algorithm.

1.4.3 Distributed Computing

Approximating a network using a tree could be useful to obtain compact approximate representations of a graph. Experimental work on this has be made by Awerbuch and Shavitt [AS98].

Further more the structure of the HST tree can be efficiently kept in the original network, for routing purposes, by using ITR (interval routing scheme), as described in [ABLP89, AP90], with only polylogarithmic memory per processor. Thus our method is also useful for competitive analysis of *distributed* problems [BFR92] in networks.
• *The mobile user* problem [AP91] and *data tracking* [BFR92]. The problem of finding a mobile user or data in a network in a distributed and on-line fashion. This

problem has a trivial $O(1)$ competitive algorithn on a tree. We therefore improve upon the previous results by obtaining a randomized $O(\log n \log\log n)$ competitive ratio.

• *The distributed K-server* problem. Bartal and Rosén [BR92] give a distributed translator for (global-control) K-server algorithms. The competitive ratio of the distributed algorithm is a function of the number of memory states of the original K-server algorithm. From the discussion in the previous paragraph follows a distributed randomized K-server algorithm with only $O(\log n \log\log n)$ overhead.

• *The generalized Steiner* problem and the *network-leasing* problem of [AAB96] and the *file allocation* problem [BFR92, ABF93a]. We obtain better randomized *distributed* on-line algorithms for these problems all with competitive ratio $O(\log n \log\log n)$, by exploiting the simplicity of these problems for trees and k-HSTs.

2 Definitions and Results

2.1 Probabilistic Approximations of Metric Spaces

Let V be a set of n points. We study the set of metric spaces defined over V. Given such a metric space M, the distance between u and v in V is denoted $d_M(u,v)$.

Definition 1 *A metric spaces N over V, dominates a metric space M over V, if for every $u, v \in V$, $d_N(u,v) \geq d_M(u,v)$.*

Definition 2 *A metric spaces N over V, α-approximates a metric space M over V, if it dominates M and for every $u, v \in V$, $d_N(u,v) \leq \alpha \cdot d_M(u,v)$.*

Definition 3 ([Bart96]) *A set of metric spaces $\mathcal{S}$ over V, α-probabilistically-approximates a metric space M over V, if every metric space in $\mathcal{S}$ dominates M and there exists a probability distribution over metric spaces $N \in \mathcal{S}$ such that for every $u, v \in V$, $\mathrm{E}(d_N(u,v)) \leq \alpha \cdot d_M(u,v)$.*

Consider an optimization problem $\mathcal{P}$ defined on metric spaces, where the cost a solution for $\mathcal{P}$ can be expressed as a linear combination of distances between vertices in the metric space.

The following gives the motivation for Definition 3.

Theorem 4 ([Bart96]) *Let $\mathcal{S}$ be a set of metric spaces over V that α-probabilistically-approximates M. If there exists a (randomized) algorithm for metric spaces in $\mathcal{S}$ with performance ratio β for $\mathcal{P}$ then there exists a randomized algorithm for M with performance ratio $\alpha\beta$ for $\mathcal{P}$. Moreover, if the distribution on $\mathcal{S}$ is efficiently computable then so is the algorithm for M.*

2.2 Deterministic Approximation on the Average

The problem of obtaining a probabilistic approximation of a metric space can be viewed as a two person zero-sum game where the "metric player" chooses a metric spaces N out of a fixed set $\mathcal{S}$ and the "edge player" chooses a pair of vertices. The payoff function to the edge player is defined to be the distortion of the distance between the pair (i.e., the ratio between the distance in N and the distance in the original metric M).

Consider the following notion of approximating the average distance in a metric space.

Definition 5 *A set of metric spaces $\mathcal{S}$ over V, α- distributionally approximates a metric space M over V, if every metric space in $\mathcal{S}$ dominates M and for every nonnegative weight function μ on pairs of V (and assume $\sum \mu(u,v) = 1$), there exists a metric space $N \in \mathcal{S}$ such that $\sum_{\{u,v\} \in V^2} \mu(u,v) \frac{d_N(u,v)}{d_M(u,v)} \leq \alpha$.*

Let $\mathcal{P}$ denote the set of mixed strategies for the metric player and $\mathcal{D}$ denote the set of mixed strategies for the edge player.

The von Neumann minimax principle of game theory states that

$$\max_{\mu \in \mathcal{D}} \min_{N \in \mathcal{S}} \sum_{\{u,v\} \in V^2} \mu(u,v) \frac{d_N(u,v)}{d_M(u,v)} =$$

$$\min_{\nu \in \mathcal{P}} \max_{\{u,v\} \in V^2} \sum_{N \in \mathcal{S}} \nu(N) \frac{d_N(u,v)}{d_M(u,v)}.$$

That is a class $\mathcal{S}$ α-probabilistically-approximates M iff $\mathcal{S}$ α-distributionally-approximates M.

Even though the definitions are equivalent it is still interesting to study both the deterministic and probabilistic versions from a complexity perspective. Our goal is to find efficient algorithms to construct the approximating metric space, and therefore we will need to show that both versions can be computed in polynomial time.

2.3 Hierarchically Well-Separated Trees

Any finite metric space can be represented by a weighted connected graph (and vice-versa). For most of the technical part of the paper it will be convenient to use graph terminology. However we will use the definitions of the previous subsection for graphs with the obvious meaning.

For a weighted connected graph $G = (V, E, \omega)$, where $\omega(e)$ is the weight of edge e, $d_G(u,v)$ denotes the sum of edge weights over a shortest path between u and v. Let $\mathrm{diam}(G)$ denote the weighted diameter of the graph G; i.e., the maximum distance between a pair of vertices. When there is no confusion we will use $\Delta = \mathrm{diam}(G)$.

Definition 6 *A tree metric (or additive metric) over V is a metric space corresponding to a weighted tree spanning V.*

The following lower bound gives the justification for considering probabilistic metric approximations.

Theorem 7 ([RR95]) *If a tree metric α (deterministically) approximates the n-cycle then $\alpha = \Omega(n)$.*

Definition 8 ([Bart96]) *A k-hierarchically well separated tree (k-HST) is defined as a rooted weighted tree with the following properties:*

- *The edge weight from any node to each of its children is the same.*
- *The edge weights along any path from the root to a leaf are decreasing by a factor of at least k.*

Our main results are

Theorem 9 *Every weighted connected graph G can be α-probabilistically-approximated by a set of k-HSTs where $\alpha = O(k \log n \log\log n)$. Note that each HST in the set has diameter proportional to that of G. Moreover the probability distribution computation and the HST construction are polynomial time.*

Theorem 10 *Every weighted connected graph G can be α-distributionally-approximated by a set of k-HSTs where $\alpha = O(k \log n \log\log n)$. Note that each HST in the set has diameter proportional to that of G. Moreover the HST construction is polynomial time.*

2.4 Hierarchical Partition Metrics and HSTs

Let $G = (V, E, \omega)$ be a weighted connected graph.

Definition 11 *A hierarchical partition metric (HPM) on a weighted connected graph G is a length function $\ell(e)$ defined on the edges of G according to the following recursive process: let C be a cut in G. For every $e \in C$, $\ell(e) \geq \mathrm{diam}(G)$ and ℓ forms a hierarchical partition metric over each connected component recursively.*

We prove the following theorems

Theorem 12 *Every weighted connected graph G can be α-probabilistically-approximated by the set of HPMs of G where $\alpha = O(\log n \log\log n)$. Moreover the probability distribution computation and the HPM construction is polynomial time.*

Theorem 13 *Every weighted connected graph G can be α-distributionally-approximated by the set of HPMs of G where $\alpha = O(\log n \log\log n)$. Moreover the HPM construction is polynomial time.*

We then obtain the HSTs construction from the following.

Theorem 14 *Given a weighted connected graph G and an HPM ℓ of G. There exists a k-HST that dominates G and for every $u, v \in V$, $d_T(u,v) \leq \frac{k^2}{k-1}\ell(u,v)$. Note that this HST has diameter proportional to that of G. Moreover the HST can be constructed is polynomial time.*

Combining Theorems 12, 13 with Theorem 14 gives our main result Theorems 9, 10.

We note that the ratio of probabilistic approximation of arbitrary metrics by tree metrics has a lower bound of $\Omega(\log n)$ [Bart96] and therefore the same lower bound holds for HPMs.

3 From HPMs to HSTs

In this section we prove Theorem 14. Let $G = (V, E, \omega)$ be a weighted connected graph. Let ℓ be an HPM of G. We build a k-HST, T, recursively as follows. Consider the hierarchical partition metric. Every vertex in the tree corresponds to a subgraph in the construction of the partition. The root corresponds to the entire graph. Set $L = \Delta$.

At a certain level in the recursion the current subgraph H corresponds to a leaf in the current tree T. If $\mathrm{diam}(H) < L/k$ then let v be the leaf corresponding to the subgraph H and decrease L by a factor of k. Otherwise let v be the parent of that leaf. Let C be the cut that defines the HPM and consider the connected components (or clusters) that are left after removing the cut edges. We define a child of v for each of these clusters. The weight of the tree edge to each of the children is equal to $L/2$. Note that the label associated with the cut edges according to ℓ is at least as large as the diameter of the parent subgraph H which is at least L/k. This process is continued recursively on each of clusters.

Obviously the result tree T is a k-HST. Next we shall show that the distances in the HST are dominating the distances according to the original metric and are bounded above by $\frac{k}{k-1}$ times the distances according to the HPM. Consider some pair of vertices u and v. Let w be twice the weight of the heaviest edge in path between them in T. The distance between them in the tree is at least w. Consider the time in the recursive process above when this edge was constructed. The vertices u and v must have been separated into different clusters Let H be the subgraph containing both u and v before that time. By the construction we have that the weight w is at least as large as the diameter of H and therefore at least as large as the original distance between them. Moreover the construction promises that the path between u and v in the original graph must cross an edge with label of the HPM at least as large as the diameter of H which is at least w/k. It is easy to check that the distance between u and v in the HST is at most $\frac{k}{k-1}w$ and thus at most $\frac{k^2}{k-1}$ times the distance according to the HPM.

4 Computing Hierarchical Partition Metrics

Let $G = (V, E, \omega)$ be a weighted connected graph. We are going to show how to generate both the deterministic and probabilistic HPMs for G. For the deterministic HPM assume that μ is a nonnegative weight function μ on the edges of G (and assume $\sum_e \mu(e) = 1$). The construction is done recursively in the following manner.

Let H be the subgraph of G currently served as an input and let U be the set of nodes of H. Let $\phi = \phi(H) = |E(H)|$.

Let λ be a constant to be determined later. We will prove inductively the following

- The exists a poly-time computable probability distribution over HPMs of H such that for all e, $\mathrm{E}(\frac{\ell(e)}{\omega(e)}) \leq \lambda \log(\phi) \log\log(\phi)$.
- For all μ there exists a poly-time computable HPM of H such that $\sum_e \mu(e)\frac{\ell(e)}{\omega(e)} \leq \lambda \log(\phi) \log\log(\phi)$.

which imply Theorems 12, 13.

Let $\Delta = \mathrm{diam}(H)$ and set $T = \phi = |E(H)|$. We slightly modify the graph first by contracting edges e (joining their endpoints) in the set $S = \{e | \omega(e) < \frac{\Delta}{2T}\}$. The resulting graph is denoted $\tilde{H}$ and let $\tilde{\Delta} = \Delta/2$. Note that $\mathrm{diam}(\tilde{H}) \geq \tilde{\Delta}$.

Choose an arbitrary vertex $v \in U$. We choose an integer g in the range $[1, T]$ and set the radius $z_g = \frac{g}{T}\tilde{\Delta}$ (Note that $z_g \in (0, \Delta/2])$. Let the cut C_g be the set of edges (u, w) of $\tilde{H}$ such that $d(v, u) \leq z_g < d(v, w)$. For every $e \in C_g$ set $\ell(e) = \Delta$.

Let $U_g^1 = \{u | d(v, w) < z_g\}$ and $U_g^2 = \{u | d(v, u) \geq z_g\}$ and let H_g^1 and H_g^2 be the corresponding subgraphs of H induced by the two sets and then recurse on both subgraphs (if not empty).

It is left to explain the method for choosing g. The deterministic strategy will choose a specific g while the probabilistic strategy will use a probability distribution p_g. We will do this in conjunction with the analysis. Let $P_e(H)$ denote the expected value of $\frac{\ell(e)}{\omega(e)}$ according to the probability distribution p_g. Let $D(H)$ denote $\sum_e \mu(e)\frac{\ell(e)}{\omega(e)}$. Similarly, define $D(H_g^1)$ and $D(H_g^2)$ where μ is normalized accordingly.

For every $i \in [1, T]$ let $\epsilon_i = \phi(H_i^1)/\phi$.

- First let us consider the deterministic version. If $g = i$ is the value chosen by the algorithm then

$$D(H) = \sum_{e \in C_i} \mu(e) \cdot \frac{\Delta}{\omega(e)} + \sum_{e \in H_i^2} \mu(e) D(H_i^2) + \sum_{e \in H_i^1} \mu(e) D(H_i^1).$$

For $e = (u, w)$, define $\tilde{\omega}(e) = |d(v, u) - d(v, w)|$. Note that $\tilde{\omega}(e) \leq \omega(e)$. For every $e \notin S$, we have $\tilde{\omega}(e) \leq \sum_{j; e \in C_j} 2\frac{\tilde{\Delta}}{T}$ so that $\sum_{e \in H_i^1} \mu(e) \leq \sum_{j<i} \sum_{e \in C_j} 2\frac{\mu(e)\tilde{\Delta}}{T\tilde{\omega}(e)}$. Letting $q_i = 2\sum_{e \in C_i} \frac{\mu(e)\tilde{\Delta}}{T\tilde{\omega}(e)}$ we have $\sum_{e \in H_i^1} \mu(e) \leq \sum_{j<i} q_j$. Similarly, we have $\sum_{e \in H_i^2} \mu(e) \leq \sum_{j>i} q_j$.

By induction $D(H_i^1) \leq \lambda \log(\epsilon_i \phi) \log\log(\phi)$ and $D(H_i^2) \leq \lambda \log((1 - \epsilon_i)\phi) \log\log(\phi)$. Thus we obtain

$$\begin{aligned} D(H) &\leq q_i \cdot T + \sum_{j<i} q_j \cdot \lambda \log(\epsilon_i \phi) \log\log(\phi) \\ &\quad + \sum_{j>i} q_j \cdot \lambda \log((1 - \epsilon_i)\phi) \log\log(\phi) \\ &\leq \lambda \log(\phi) \log\log(\phi) + q_i \cdot T \\ &\quad - \sum_{j<i} q_j \cdot \lambda \log(\frac{1}{\epsilon_i}) \log\log(\phi) \\ &\quad - \sum_{j>i} q_j \cdot \lambda \log(\frac{1}{1 - \epsilon_i}) \log\log(\phi). \end{aligned}$$

It follows that the problem of finding the worst case ratio over all weight functions can be rephrased as a linear programming problem:

$$\begin{aligned} &\max_q s \\ &s.t \quad \forall i \in [1, T], \\ &\quad q_i \cdot T - \sum_{j<i} q_j \cdot \gamma \log(\frac{1}{\epsilon_i}) \log\log(\phi) \\ &- \sum_{j>i} q_j \cdot \gamma \log(\frac{1}{1 - \epsilon_i}) \log\log(\phi) \geq s \\ &and \quad \sum_j q_j = 1, \quad \forall j, q_j \geq 0 \end{aligned} \tag{1}$$

We can show that the solution to this program is ≤ 0 which imply that $D(H) \leq \lambda \log(\phi) \log\log(\phi)$.

- Next consider the probabilistic version. For $e \notin S$, let $\{j | e \in C_j\} = [f, f']$ then

$$\begin{aligned} P_e(H) &= (\sum_{j \in [f, f']} p_j)\frac{\Delta}{\omega(e)} + \sum_{j<f} p_j P_e(H_j^2) \\ &\quad + \sum_{j>f'} p_j P_e(H_j^1). \end{aligned}$$

Let $i = \mathrm{argmax}_{j \in [f, f']} p_j$. Now since $\frac{\Delta}{\omega(e)} \leq 2T$ we have

$$\begin{aligned} (\sum_{j \in [f, f']} p_j)\frac{\Delta}{\omega(e)} &\leq p_i(\omega(e)\frac{T}{\tilde{\Delta}} + 1)\frac{\Delta}{\omega(e)} \\ &\leq p_i(2T + 2T) = p_i \cdot 4T. \end{aligned}$$

By induction $P_e(H_j^1) \leq \lambda \log(\epsilon_j \phi) \log\log(\phi)$ and $P_e(H_j^2) \leq \lambda \log((1-\epsilon_j)\phi) \log\log(\phi)$.

Thus we get

$$\begin{aligned} P_e(H) &\leq p_i \cdot 4T + \sum_{j<i} p_j \cdot \lambda \log((1-\epsilon_j)\phi) \log\log(\phi) \\ &\quad + \sum_{j>i} p_j \cdot \lambda \log(\epsilon_j \phi) \log\log(\phi) \\ &\leq \lambda \log(\phi) \log\log(\phi) + p_i \cdot 4T \\ &\quad - \sum_{j>i} p_j \cdot \lambda \log(\frac{1}{\epsilon_j}) \log\log(\phi) \\ &\quad - \sum_{j>i} p_j \cdot \lambda \log(\frac{1}{1-\epsilon_j}) \log\log(\phi). \end{aligned}$$

It follows that the problem of finding the best ratio over all probability distributions can be rephrased as a linear programming problem:

$$\begin{aligned} &\min_p s \qquad (2) \\ &s.t \quad \forall i \in [1,T], \\ &\quad p_i \cdot T - \sum_{j>i} p_j \cdot \gamma \log(\frac{1}{\epsilon_j}) \log\log(\phi) \\ &- \sum_{j>i} p_j \cdot \gamma \log(\frac{1}{1-\epsilon_j}) \log\log(\phi) \leq s \\ &and \quad \sum_j p_j = 1, \quad \forall j, p_j \geq 0 \end{aligned}$$

Linear programming duality (or equivalently the minimax principle) implies that the solution to Program 2 is the same as that of Program 1. Thus since the solution to Program 1 is ≤ 0 it then follows that the probability distribution given by the solution to Program 2 would give the promised bound.

The remainder of the section is dedicated to proving the upper bound on the solution of Program 1.

Recall that we want to show that for every vector of q_js there exist i such that

$$\begin{aligned} &q_i \cdot T - \sum_{j<i} q_j \cdot \gamma \log(\frac{1}{\epsilon_i}) \log\log(\phi) \\ &- \sum_{j>i} q_j \cdot \gamma \log(\frac{1}{1-\epsilon_i}) \log\log(\phi) \leq 0. \end{aligned}$$

Define the summations sequence $a_i = \sum_{j \leq i} q_j$. Recall that $T = \phi$. Then the following lemma completes the proof.

Lemma 15 *Let $T \geq \phi \geq 8$. Let $0 = a_0 \leq a_1 \leq a_2 \leq \ldots \leq a_T = 1$ and $\frac{1}{\phi} \leq \epsilon_1, \epsilon_2, \ldots, \epsilon_T \leq 1 - \frac{1}{\phi}$. There exist a constant γ and an index i such that*

$$\begin{aligned} a_i - a_{i-1} &\leq \frac{\gamma}{T}(a_{i-1} \log(\frac{1}{\epsilon_i}) \\ &\quad + (1-a_i) \log(\frac{1}{1-\epsilon_i})) \log\log(\phi T). \end{aligned}$$

Proof. We will assume that at least half of the a_is have value at most $1/2$. The symmetric case follows from the same argument on the sequences $a'_i = 1 - a_{T-i}$ and $\epsilon'_i = 1 - \epsilon_{T-i}$. Let $\psi = \phi T/32$. If $a_1 \leq \frac{1}{\psi}$ then we can choose $i = 1$. The inequality follows from the assumption that $\epsilon_1 \geq \frac{1}{\phi}$. Otherwise define intervals I_l for $l \leq \log(\psi)$ as follows. Let $I_0 = \{i | a_i > \frac{1}{2}\}$. For $l = 1, \ldots, \log(\psi)$ let $m_l = \max_{i \notin \cup_{j<l} I_j} a_i$ and let $I_l = \{i | \frac{m_l}{2} < a_i \leq m_l\}$. Note that $m_l \leq \frac{1}{2^l}$. It follows that there must be some interval I_l of size at least $\frac{T}{4l \log\log(\psi)} + 1$ since otherwise the length of the sequence would be bounded by $\sum_{1 \leq l \leq \log(\psi)} (\frac{T}{4l \log\log(\psi)} + 1) < T/2$, a contradiction. Fix that value of l. We get that there exists an i in the interval I_l such that $a_i - a_{i-1} \leq \frac{4 \log\log(\psi)}{T} \cdot \frac{m_l}{2} l$. On the other hand we have

$$\begin{aligned} &a_{i-1} \log(\frac{1}{\epsilon_i}) + (1-a_i) \log(\frac{1}{1-\epsilon_i}) \\ &\geq \frac{m_l}{2} \log(\frac{1}{\epsilon_i}) + (1-m_l) \log(\frac{1}{1-\epsilon_i}). \end{aligned}$$

By minimizing the expression above as a function of ϵ_i we get that this is at least

$$\begin{aligned} &\frac{m_l}{2} \log(\frac{2}{m_l} - 1) + (1-m_l) \log(\frac{\frac{2}{m_l} - 1}{\frac{2}{m_l} - 2}) \\ &\geq \frac{m_l}{2} \log(\frac{2}{m_l}) \geq \frac{m_l}{2} l \end{aligned}$$

using that $m_l \leq 1/2^l$. It follows that we can choose an appropriate constant γ to satisfy the inequality 3. This completes the proof of the lemma. ■

5 Conclusions and Open Problems

The metric space approximation theorems presented here have already found a large number of different applications in the areas of on-line, distributed and approximation algorithms. It would be interesting to explore further applications. Finding the best approximation factor for HSTs remains a challenging problem.

References

[AA97] B. Awerbuch and Y. Azar Buy-at-Bulk Network Design. In *Proc. 38th IEEE Symp. on Foundations of Computer Science* , October 1997.

[AAB96] B. Awerbuch, Y. Azar, and Y. Bartal. On-line Generalized Steiner Problem. In *Proc. of the 7th Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 68–74, January 1996.

[ABF93a] B. Awerbuch, Y. Bartal, and A. Fiat. Competitive Distributed File Allocation. In *Proc. 25th ACM Symp. on Theory of Computing*, pages 164–173, May 1993.

[ABF93b] B. Awerbuch, Y. Bartal, and A. Fiat. Heat & Dump: Randomized competitive distributed paging. In *Proc. 34rd IEEE Symp. on Foundations of Computer Science* , pages 22–31,November 1993.

[ABF96] B. Awerbuch, Y. Bartal, and A. Fiat. Distributed Paging for General Networks. In *Proc. of the 7th Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 574–583, January 1996.

[ABFNPT96] R. Agarwala, V. Bafna, M. Farach, B. Narayanan, M. Paterson, and M. Thorup. On the Approximability of Numerical Taxonomy. In *Proc. of the 7th Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 365–372, January 1996.

[ABLP89] B. Awerbuch, A. Bar-Noy, N. Linial, and D. Peleg. Compact Distributed Data Structures for Adaptive Network Routing. In *Proc. 21st ACM Symp. on Theory of Computing*, pages 230–240, May 1989.

[ADDJS90] I. Althöfer, G. Das, D. Dobkin, D. Joseph, and J. Soares. On Sparse Spanners of Weighted Graphs. In *Discrete and Computational Geometry*.

[AKPW91] N. Alon, R.M. Karp, D. Peleg, and D.B. West. Graph-Theoretic Game, and its Applications to the k-Server Problem. To appear in *SIAM J. on Computing*.

[AP90] B. Awerbuch and D. Peleg. Sparse partitions. In *Proc. 31st IEEE Symp. on Foundations of Computer Science*, pages 503–513, 1990.

[AP91] Baruch Awerbuch and David Peleg. Concurrent online tracking of mobile users. In *Proceedings of the Annual ACM SIGCOMM Symposium on Communication Architectures and Protocols, Zurich, Switzerland*, September 1991.

[AS98] Baruch Awerbuch and Yuval Shavitt. Topology Aggregation for Directed Graphs. To appear in *Proc. of the 3rd Ann. IEEE Symp. on Computers and Communications*, June 1998.

[Bart96] Y. Bartal. Probabilistic Approximation of Metric Spaces and its Algorithmic Applications. In *Proc. of the 37th Ann. IEEE Symp. on Foundations of Computer Science*, pages 184–193, October 1996.

[Bour85] J. Bourgain. On Lipshitz Embedding of Finite Metric Spaces in Hilbert Space. In *Israel J. Math.* 52, pages 46–52, 1985.

[BBKTW90] S. Ben-David, A. Borodin, R.M. Karp, G. Tardos, and A. Wigderson. On the Power of Randomization in Online Algorithms. In *Proc. of the 22nd Ann. ACM Symp. on Theory of Computing*, pages 379–386, May 1990.

[BBBT97] Y. Bartal, A. Blum, C. Burch, and A. Tomkins. A Polylogarithmic Competitive Algorithm for Metrical Task Systems. In *Proc. of the 29th Ann. ACM Symp. on Theory of Computing*, May 1997.

[BFR92] Y. Bartal, A. Fiat, and Y. Rabani. Competitive algorithms for distributed data management. In *Proc. 24th ACM Symp. on Theory of Computing*, pages 39–50, 1992. Also accepted for publication in the STOC 92 special issue *Journal of Computer and Systems Sciences*, 1993.

[BG92] J-P. Barthélemy and A. Guénoche. *Trees and Proximity Representations*. Wiley, New York, 1991.

[BLS87] A. Borodin, N. Linial, and M. Saks. An Optimal On-Line Algorithm for Metrical Task Systems. In *Proc. of the 19th Ann. ACM Symp on Theory of Computing*, pages 373–382, May 1987.

[BR92] Y. Bartal and A. Rosén. The Distributed k-Server Problem — A Competitive Distributed Translator for k-Server Algorithms. In *Proc. of the 33rd Ann. IEEE Symp. on Foundations of Computer Science*, pages 344-353, October 1992.

[Cai92] L. Cai. *Tree Spanners: Spanning trees that Approximate Distances*. Ph.D Thesis. Tech. Report 260/92, Univ. of Toronto, May 1992.

[CC95] L. Cai and D.G. Corneil. Tree Spanners. In *SIAM J. on Discrete Math.*, 8(3), pages 359–387, 1995.

[CCGG98] M. Charikar, C. Chekuri, A. Goel and S. Guha. Rounding via Trees: Deterministic Approximation Algorithms for Group Steiner Trees and k-Median. These proceedings.

[CCGGP98] M. Charikar, C. Chekuri, A. Goel, S. Guha, and S. Plotkin. Approximating an Arbitrary Metric by $O(n \log n)$ Trees. Personal communication.

[CDNS92] B. Chandra, G. Das, G. Narasimhan, and J. Soares. New Sparseness Results on Graph Spanners. Manuscript, 1996.

[CL91] M. Chrobak and L. Larmore. An Optimal On-Line k-Server Algorithm for Trees. *SIAM J. of Computing*, 20:144-148, 1991.

[ENRS95] G. Even, J. Naor, S. Rao, and B. Scheiber. Divide-and-Conquer Approximation Algorithms via Spreading Metrics. In *Proc. of the 36th Ann. IEEE Symp. on Foundations of Computer Science*, pages 62–71, October 1995.

[GKR98] N. Garg, G. Konjevod, and R. Ravi. A Polylogarithmic Approximation Algorithm for the Group Steiner Tree Problem. To appear in *Proc. of the 9th Ann. ACM-SIAM Symp. on Discrete Algorithms*, , January 1998.

[GW85] R.L. Graham and P.M. Winkler. On Isometric Embeddings of Graphs. In *Trans. of the Amer. Math. Soc.* 288, pages 527–536, 1985.

[Hu74] T.C. Hu. Optimum Communication Spanning Trees. In *SIAM J. Computing*, vol. 3, pages 188–195, 1974.

[JL84] W.B. Johnson and J. Lindenstrauss. Extensions of of Lipschitz Mappings into a Hilbert Space. In *Contemporary Mathematics* 26, pages 189–206, 1984.

[Karp89] R.M. Karp. A $2k$-competitive Algorithm for the Circle. Manuscript, August 1989.

[KH79] O. Kariv and S.L. Hakimi. An Algorithmic Approach to Network Location Problems, Part II: p-Medians. In *SIAM J. Appl. Math.*, 37:539–560, 1979.

[KP94] E. Koutsoupias, and C. Papadimitriou. On the k-Server Conjecture. In *Proc. 26th Annual ACM Symp. on Theory of Computing*, pages 507–511, 1994.

[KRS98] G. Konjevod, R. Ravi, and F.S. Salman. On Approximating Planar Metrics by Tree Metrics. Manuscript.

[LLR94] N. Linial, E. London, and Y. Rabinovich. The Geometry of Graphs and Some of its Algorithmic Applications. In *Proc. of the 35th Ann. IEEE Symp. on Foundations of Computer Science*, pages 577–591, October 1994.

[LS91] N. Linial and M. Saks. Decomposing Graphs into Regions of Small Diameter. In *Proc. of the 2nd Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 320–330, January 1991.

[MMS88] M.S. Manasse, L.A. McGeoch, and D.D. Sleator. Competitive Algorithms for On-Line Problems. In *Proc. of the 20th Ann. ACM Symp. on Theory of Computing*, pages 322–333, May 1988.

[Neum59] Zur Theorie der Gesellschaftsspiele. *Math. Ann.*, 100:295-320, 1928. Translation: On the Theory of Games of Strategy. Contribution to the Theory of Games, vol. 4, ed. by A.W. Tucker and R.D. Luce, Princeton Univ. Press 1959.

[PS89] D. Peleg and A.A. Schäffer. Graph Spanners. In *Journal of Graph Theory*, 13(1):99–106, 1989.

[PU89] D. Peleg and J. Ullman. An Optimal Synchronizer for the Hypercube. In *SIAM J. Computing*, 18, pages 740-747, 1989.

[RR95] Y. Rabinovich and R. Raz. Lower Bounds on the Distortion of Embedding Finite Metric Spaces in Graphs. To appear in *Discrete and Computational Geometry*.

[Seym95] P.D. Seymour. Packing Directed Circuits Fractionally. In *Combinatorica*, 15(2):281–288, 1995.

[ST85] D. Sleator and R.E. Tarjan. Amortized Efficiency of List Update and Paging Rules. *Comm. of ACM* , 28(2):202–208, 1985.

[Tami96] A. Tamir. An $O(pn^2)$ Algorithm for the p-Median and Related Problems on Tree Graphs. In *Operations Research Letters*, 19:59–64, 1996.

[WLBCRT98] B.Y. Wu, G. Lancia, V. Bafna, K. Chao, R. Ravi and C.Y. Tang. A Polynomial Time Approximation Scheme for Minimum Routing Cost Spanning Trees. To appear in *Proc. of the 9th Ann. ACM-SIAM Symp. on Discrete Algorithms*, , January 1998.