Multi-Embedding and Path Approximation of Metric Spaces

Yair Bartal* Manor Mendel†

Abstract

Metric embeddings have become a frequent tool in the design of algorithms. The applicability is often dependent on how high the embedding's distortion is. For example embedding into ultrametrics (or arbitrary trees) requires linear distortion. Using probabilistic metric embeddings, the bound reduces to $O(\log n \log\log n)$. Yet, the lower bound is still logarithmic.

We make a step further in the direction of bypassing this difficulty. We define "multi-embeddings" of metric spaces where a point is mapped onto a set of points, while keeping the target metric being of polynomial size and preserving the distortion of paths. The distortion obtained with such multi-embeddings into ultrametrics is at most $O(\log \Delta \log\log \Delta)$ where Δ is the (normalized) diameter, and probabilistically $O(\log n \log\log\log n)$. In particular, for expander graphs, we are able to obtain *constant* distortions embeddings into trees vs. the $\Omega(\log n)$ lower bound for all previous notions of embeddings.

We demonstrate the algorithmic application of the new embeddings by obtaining improvements for two well-known problems: *group Steiner tree* and *metrical task systems*.

1 Introduction

Finite metric spaces and their analysis play a significant role in the design of combinatorial algorithms. Many algorithmic techniques were introduced in recent years concerning and using metric spaces and their approximate embedding in other spaces, see the surveys [14, 16] for an overview of this topic.

An embedding of a metric space $M = (V_M, d_M)$ into a metric space $N = (V_N, d_N)$ is a mapping $f : V_M \mapsto V_N$. An embedding is dominating if for all $u, v \in V_M$, $d_M(u,v) \le d_N(f(u), f(v))$ and has distortion α if also for all $u, v \in V_M$, $d_N(f(u), f(v)) \le \alpha d_M(u,v)$. When such an embedding exists we say that N α-approximates M.

*Computer Science Department, The Hebrew University, Jerusalem. email: yair@cs.huji.ac.il.

†Computer Science Department, The Hebrew University, Jerusalem. email: mendelma@cs.huji.ac.il. Supported in part by the Landau Center.

The general framework for applying metric embeddings in optimization problems is to embed arbitrary metric spaces in some metric space from some "simple" family and then apply an algorithm for that space. As a result, the approximation ratio increases by a factor of the embedding's distortion.

Among others, embeddings into low dimensional normed spaces [7, 17] as well as probabilistic embeddings into trees [2] have many algorithmic applications. In both cases the distortions of the embeddings are logarithmic (or slightly higher in the latter). Yet, there is a logarithmic lower bound on the distortion of these embeddings as well, which provides limits to their applicability. In this paper we attempt to tackle this problem.

Tree metrics, and in particular, ultrametrics seem a natural choice as a target class of "simple" metric spaces. We also note that ultrametrics are isometric in ℓ_p and in fact, can be $O(1)$-embedded in $\ell_p^{O(\log n)}$ [5].

Unfortunately, standard embedding is not useful for tree metrics. Embedding arbitrary metric spaces into trees requires distortion linear in the size of the metric space [18]. Probabilistic metric approximation [1] provides a way to bypass this problem:

Definition 1.1. (Probabilistic Approximation) *A metric space $M = (V_M, d_M)$ is* α-probabilistically approximated *by a set of metric spaces $\mathcal{S}$ if there exist a distribution $\mathcal{D}$ over $\mathcal{S}$ and for every $N \in \mathcal{S}$, a dominating embedding $f_N : V_M \mapsto V_N$, such that for all $u, v \in V_M$, $\mathbb{E}_{N \in \mathcal{D}}[d_N(f_N(u), f_N(v))] \le \alpha d_M(u,v)$.*

The following is the main theorem on metric embeddings into ultrametrics.

Theorem 1.1. ([2]) *Any metric space on n points can be $O(\log n \log\log n)$-probabilistically approximated by a set of n-point ultrametrics.*

In fact, Theorem 1.1 can was originally formulated for a class of metric spaces defined by the following natural generalization of ultrametrics:

Definition 1.2. ([1]) *For $k \ge 1$, a* k-hierarchically well-separated tree *(k-HST) is a metric space defined on the leaves of a rooted tree T. To each vertex $u \in T$ there is associated a label $\Delta(u) \ge 0$ such that $\Delta(u) = 0$ iff u*

is a leaf of T. The labels are such that if a vertex v is a child of a vertex u then $\Delta(v) \leq \Delta(u)/k$. The distance between two leaves $x, y \in T$ is defined as $\Delta(\mathrm{lca}(x,y))$, where $\mathrm{lca}(x,y)$ is the least common ancestor of x and y in T.

The definition of ultrametric is the same as a 1-HST. Any k-HST is therefore, in particular, an ultrametric and any ultrametric can be k-approximated by a k-HST [2]. We can therefore restrict our attention to ultrametrics, while all results generalize to k-HSTs.

Theorem 1.1 found many algorithmic applications in approximation algorithms, online algorithms, and distributed algorithms. The bound in Theorem 1.1 is almost tight even for probabilistic embeddings into tree metrics for which there is an $\Omega(\log n)$ lower bound [1].

The main contribution of this paper is in offering a new type of metric embedding that makes it possible to bypass lower bounds for the standard and even probabilistic metric embeddings. There are two key observations that lead to this new type of embedding. The first is that in some applications it is natural to match a point onto a set of points in the target metric space. Motivated by two well-known applications of Theorem 1.1: the group Steiner problem, and the metrical task systems problem, we propose the following definition:

Definition 1.3. (Multi-Embedding) *A multi embedding of M in N is a partial surjective function f from N on M, i.e. each point $x \in M$ is embedded into a set $\mathsf{f}^{-1}(x)$. Points in $\mathsf{f}^{-1}(x)$ are called* representatives *of x in N.*

To avoid confusion we stress that the role of f^{-1} in Definition 1.3 replaces the role of f in the definition of embedding and probabilistic embedding.

The second observation is that we need not have the distance between every two appropriate representatives approximate the original distance, rather what we really need is that paths in the original space will be approximated well by some path in the target space.

A path in a metric space is an arbitrary finite sequence of points in the metric. The length of a path $p = \langle u_1, u_2, \ldots, u_m \rangle$ in a metric space $M = (V, d)$ is defined as $\ell(p) = \sum_{i=1}^{m-1} d(u_i, u_{i+1})$.

Definition 1.4. (Path Approximation) *A multi-embedding f of M in N, is called* dominating *if for any $u, v \in N$, $d_N(u,v) \geq d_M(\mathsf{f}(u), \mathsf{f}(v))$. An α-path approximation of M by N is a dominating multi-embedding of M in N, such that for any path $p = \langle u_1, u_2, \ldots, u_m \rangle$ in M, there exists a path $p' = \langle u'_1, u'_2, \ldots, u'_m \rangle$ in N such that $\mathsf{f}(u'_i) = u_i$ and $\ell(p') \leq \alpha \cdot \ell(p)$.*

Obviously, it is possible now to consider the combination of Def. 1.1 and Def. 1.4.

Definition 1.5. (Probabilistic Path Approx') *A metric space $M = (V_M, d_M)$ is α-probabilistically-path-approximated by a set of metric spaces $\mathcal{S}$ if for each $N \in \mathcal{S}$ there exists a dominating multi-embedding of M in N, $\mathsf{f}_N : N \to M$, and there exists a distribution a distribution $\mathcal{D}$ over $\mathcal{S}$ such that for any path $p = \langle u_1, u_2, \ldots, u_m \rangle$ in M, and any $N \in \mathcal{S}$, there exists a path $p^N = \langle u_1^N, u_2^N, \ldots, u_m^N \rangle$ such that $\mathsf{f}_N(u_j^N) = u_j$ and $E_{N \in \mathcal{D}}(\ell(p^N)) \leq \alpha \cdot \ell(p)$.*

A crucial parameter for multi-embeddings is the blow-up factor Γ which is the size of the target metric space. For probabilistic multi embeddings $\Gamma = \sup_{N \in \mathcal{S}}\{|V_N|\}$. In general, it will be desirable that $\Gamma = \mathrm{poly(n)}$. In fact, if $\Gamma = \infty$ then there is a trivial 1-path approximation of any metric space by trees (simply by taking a tree which includes all possible finite paths). This motivates a study of the tradeoff between Γ and the path approximation of arbitrary metric spaces by tree metrics. In Section 5 we study the path approximation of expander graphs and the hypercube by tree metrics. We show e.g. that Ramanujan graphs have 3-path-approximation by tree metrics with blow-up factor $\Gamma(n) = \mathrm{poly}(n)$. This is in sharp contrast to the status of expander graphs for previous notions of embeddings for which they are considered "worst case" examples with $\Omega(\log n)$ distortion [17]. These results directly imply nearly tight results on the approximation ratio for the group Steiner tree problem on expander graphs and the hypercube.

Most of the paper, however, concentrates on multi embeddings when the class of target metric spaces are ultrametrics. First, we observe that probabilistic approximation by ultrametrics directly implies bounds for path approximation by putting all trees in $\mathcal{S}$ (the set used in the probabilistic approximation) under a common new root. This results with an α-path approximation by an ultrametric with $\Gamma = |\mathcal{S}|n$. Using the bound of [8] on the number of ultrametrics needed in Theorem 1.1 we get an $O(\log n \log\log n)$ path approximation by an ultrametric with $\Gamma = \mathrm{poly(n)}$. Here we get better bounds using much simpler constructions.

An important parameter of the metric spaces appearing in practice is the ratio between the maximum distance (the diameter) and minimum distance in the metric space. It will be convenient for us to assume that the minimum distance is 1, and so this ratio becomes the diameter. This parameter has a significant influence on the approximation ratios possible by path approximations.

THEOREM 1.2. *Given any metric space $M = (V, d)$ on $|V| = n$ points and diameter Δ, M is α-path-approximated by an efficiently constructible ultrametric with $\Gamma = \mathrm{poly}(n)$ points, and α satisfies.*

$$\alpha = O(\min\{\log n \cdot \log\log n,\ \log \Delta \cdot \log\log \Delta\}).$$

The constructions of multi-embeddings are in a sense dual to Ramsey-type theorems for metric spaces [4, 5], where the goal is to find a large subspace which approximates some ultrametric, and some of the techniques in the proof of Theorem 1.2 are dual to those of [4, 5].

We note that our construction beats the probabilistic approximation based constructions in two respects:

- Metrics with small diameters are better approximated using Theorem 1.2. Expander graphs are examples where a lower bound of $\Omega(\Delta)$ exists on probabilistic approximation using trees [17].
- The algorithm (and its analysis) in Theorem 1.2 is much simpler and more efficient (in particular, it does not involve a linear program).

Using probabilistic path approximation, we obtain a slightly better distortion as function of n.

THEOREM 1.3. *Given any n-point metric space $M = (V, d)$, M is $O(\log n \log\log\log n)$ probabilistically-path-approximated by a set of ultrametrics with $\Gamma = \mathrm{poly}(n)$.*

Theorem 1.3 is proved in Section 4 by a careful synthesis of the techniques used in [2] and those used to prove Theorem 1.2. We also provide a simple example in which Theorem 1.3 is almost tight: Any α probabilistic path approximation by ultrametrics of a simple unweighted path of length n has $\alpha = \Omega(\log n)$. This lower bound also implies that Theorem 1.2 is tight up to the sub-logarithmic term.

Summarizing Theorems 1.2 and 1.3 we have

COROLLARY 1.1. *Given any n-point metric space $M = (V, d)$ with diameter Δ, M is λ-probabilistically-path-approximated by a set of ultrametrics with $\Gamma = \mathrm{poly}(n)$, where*

$$\lambda = O(\min\{\log n \cdot \log\log\log n,\ \log \Delta \cdot \log\log \Delta\}).$$

As mentioned above, path approximation is motivated by two intensively-studied algorithmic minimization problems: group Steiner problem and metrical task system problem. For both, the best previous solution use the probabilistic approximation in Theorem 1.1. In Section 2 we prove that in order to reduce these problems to other metric spaces it is sufficient to use (probabilistic) path approximation.

2 Applications

In this section we define the metrical task systems and the group Steiner problems and prove path approximation reduces the solution of these problems to the case of ultrametrics. The reductions are stated in terms of probabilistic path approximation. Essentially the same (in fact simpler) argument imply such reductions hold for (deterministic) path approximation as well.

The *metrical task system problem* [6] was introduced as a framework for many online minimization problems. A metrical task system on metric space $M = (S, d)$, $|S| = n$, is defined as follows. The system has a set of n possible internal states S. It receives a sequence of *tasks* $\sigma = \tau_1 \tau_2 \cdots \tau_m$. Each task τ is a vector $\tau : S \to \mathbb{R}^+ \cup \{\infty\}$ of nonnegative costs for serving it in each of the internal states of the system. The system may switch states (say from u to v), paying a cost equal to the distance $d(u, v)$ in M, and then pays the service cost $\tau(v)$ associated with the new state. The major limiting factor for the system is the requirement to process the sequence in online fashion, *i.e.*, serving each task without knowing the future tasks.

As customary in the analysis of online algorithms, the metrical task system is analyzed using the notion of *competitive ratio*. A randomized online algorithm A is called r-competitive if exists some c such that for any task sequence σ, $\mathbb{E}[\mathrm{cost}_A(\sigma)] \le r\mathrm{cost}_{\mathrm{OPT}}(\sigma) + c$, where $\mathrm{cost}_A(\sigma)$ is the random variable of the cost for servicing σ by A, and $\mathrm{cost}_{\mathrm{OPT}}(\sigma)$ is the optimal (offline) cost for serving σ. The current best upper bound is $O((\log n \log\log n)^2)$ [11] (an improvement of [3]). Both these papers actually solve the MTS problem for ultrametrics (k-HSTs in fact), and then reduce arbitrary metric spaces to ultrametrics using Theorem 1.1. We show that it is sufficient to use (probabilistic) path approximation.

PROPOSITION 2.1. *Assume that a metric space M is α-probabilistically-path approximated by a distribution over $\mathcal{S}$ with blow-up factor Γ. Assume also that each m-point metric space $M' \in \mathcal{S}$ has an $r(m)$-competitive MTS algorithm. Then there is an $\alpha r(\Gamma)$-competitive randomized algorithm for M.*

Proof. We construct a randomized online algorithm $\mathcal{A}$ for M as follows: First $\mathcal{A}$ chooses a metric space $N \in \mathcal{S}$ according to the distribution, and let $\mathcal{A}_N$ be an r-competitive online algorithm for N. We translate the tasks sequence σ into a task sequence σ^N for N task by task. A task τ for M is translated into a task τ^N for N such that $\tau^N(u') = \tau(\mathsf{f}_N(u'))$. $\mathcal{A}$ maintains the invariant that if $\mathcal{A}_N$ is in state v', then $\mathcal{A}$ is in state $\mathsf{f}_N(v')$.

It is easy to verify that $\text{cost}_{\mathcal{A}}(\sigma) \le \text{cost}_{\mathcal{A}_N}(\sigma^N)$, since the local costs are the same, and the distances in N are larger. Consider $\text{OPT}(\sigma)$, it defines a path p of servicing in M. Thus there exists p^N as in the statement of Def. 1.5. The path p^N implies a servicing for N. In this way, since $\mathsf{f}_N(p^N) = p$, the local costs in N are the same as the costs of p in M, and $\mathbb{E}_{N\in\mathcal{S}}[\ell(p^N)] \le \alpha\ell(p)$. Thus $\mathbb{E}_{N\in\mathcal{S}}[\text{cost}_{\text{OPT}_N}(\sigma^N)] \le \alpha \cdot \text{cost}_{\text{OPT}}(\sigma)$. Summarizing:

$$\mathbb{E}[\text{cost}_{\mathcal{A}}(\sigma)] = \mathbb{E}_{N\in\mathcal{S}}[\mathbb{E}[\text{cost}_{\mathcal{A}_N}(\sigma^N)]]$$
$$\le r(\Gamma)\ \mathbb{E}_{N\in\mathcal{S}}[\text{cost}_{\text{OPT}_N}(\sigma^N)]+c \le \alpha r(\Gamma)\text{cost}_{\text{OPT}}(\sigma)+c. \square$$

COROLLARY 2.1. *There is an $O(\lambda \log n \log\log n)$-competitive randomized MTS algorithm against oblivious adversaries, where λ is the path approximation given in Corollary 1.1.*

Proof. We apply Proposition 2.1 to Corollary 1.1 and so we obtain $\alpha = \lambda$ and $\Gamma = \text{poly(n)}$. Using the algorithm for ultrametrics from [11] with $r(m) = O(\log m \log\log m)$, we obtain the claim. $\square$

The group Steiner problem can be stated as follows: Given a graph $G = (V, E)$ on n vertices with a cost function $c : E \to \mathbb{R}^+$, and subsets of the vertices $g_1, \dots, g_k \subset V$ (called *groups*), the objective is to find a minimum cost subtree T of G that contains at least one vertex from each g_i, $i \in [k]$. This problem generalizes the hitting set problem, and thus hard to approximate within a factor of $(1-\varepsilon)\ln k$ [10]. The current best upper bound on the approximation factor is $O(\log^2 n \log k \log\log n)$ [12]. They gives $O(\log n \log k)$ approximation algorithm for tree metrics and reduces the general case using Theorem 1.1. A recent paper [13] shows that the LP used in [12] to approximate the solution on trees has indeed an integrality gap of $\Omega(\log n \log k)$. Again, we show that it is actually sufficient to use (probabilistic) path approximation for this problem.

We show how path approximation is applicable for the group Steiner problem. As a first step we observe that the problem can be easily cast in terms of metric spaces instead of graphs: Given a graph $G = (V, E)$ with costs c, let $M = (V, d)$ be the shortest path metric induced by G (and c) on V. A tree T in M can be transformed into a tree $\hat{T}$ in G such that the total weight in $\hat{T}$ is not larger than the total weight in T, and $\hat{T}$ contains all the vertices in T. This is done by replacing each edge in T by the shortest path between its endpoints in G, and taking a spanning tree of the resulting subgraph. Thus it is enough to solve the group Steiner tree problem on metric spaces.

PROPOSITION 2.2. *Given a group Steiner problem with k groups on a metric space M. Assume that M is α-probabilistically-path approximated by a distribution over $\mathcal{S}$ with blow-up factor Γ. Also assume that for any $N \in \mathcal{S}$ there is an $r(\Gamma)$ approximation ratio for the group Steiner problem with k groups defined on N. Then there is a randomized $2\alpha r(\Gamma)$-approximation algorithm for the group Steiner problem in M.*

Proof. We construct a randomized approximation algorithm $\mathcal{A}$ for the instance $\sigma = (M; g_1, \dots, g_k)$ as follows: First $\mathcal{A}$ chooses a metric space $N \in \mathcal{S}$ according to the distribution. Consider the following group Steiner instance: $\sigma^N = (N; \mathsf{f}_N^{-1}(g_1), \dots \mathsf{f}_N^{-1}(g_k))$. Let $\mathcal{A}_N$ be an $r(\Gamma)$-approximation algorithm for σ^N. Let $T^N = \mathcal{A}_N(\sigma^N)$ be the tree constructed by A_N. Denote by $\mathsf{f}_N(T_N)$ the image graph of T_N, i.e., if $T_N = (V_N, E_N)$, then $\mathsf{f}_N(T_N) = (\mathsf{f}_N(V_N), \{\mathsf{f}_N(u)\mathsf{f}_N(v)|\ uv \in E_N\})$. $\mathsf{f}_N(T_N)$ is a connected graph of weight at most T_N, and spanning at least one representative form each group g_j. $\mathcal{A}$ returns a spanning tree of $\mathsf{f}_N(T_N)$. It is a legal solution and it satisfies $\text{cost}_{\mathcal{A}}(\sigma) \le \text{cost}_{\mathcal{A}_N}(\sigma^N)$.

Consider the tree $\text{OPT}(\sigma)$, double each edge in $\text{OPT}(\sigma)$ and take an Euler tour p of this graph. This weight of p is twice the weight of $\text{OPT}(\sigma)$. Thus there exists a path in N, p^N, as in the statement of Def. 1.5, such that $\mathsf{f}_N(p^N) = p$. The path p^N is a connected graph and it spans at least one representative from each group $\mathsf{f}_N^{-1}(g_j)$. Since $\mathbb{E}_{N\in\mathcal{S}}[\ell(p^N)] \le \alpha\ell(p) \le 2\alpha\text{cost}_{\text{OPT}}(\sigma)$, we have $\mathbb{E}_{N\in\mathcal{S}}[\text{cost}_{\text{OPT}_N}(\sigma^N)] \le \mathbb{E}_{N\in\mathcal{S}}[\ell(p^N)] \le 2\alpha \cdot \text{cost}_{\text{OPT}}(\sigma)$. Summarizing:

$$\mathbb{E}[\text{cost}_{\mathcal{A}}(\sigma)] \le \mathbb{E}_{N\in\mathcal{S}}[\mathbb{E}[\text{cost}_{\mathcal{A}_N}(\sigma^N)]]$$
$$\le r(\Gamma)\ \mathbb{E}_{N\in\mathcal{S}}[\text{cost}_{\text{OPT}_N}(\sigma^N)] \le 2\alpha r(\Gamma)\text{cost}_{\text{OPT}}(\sigma). \quad \square$$

COROLLARY 2.2. *There is an polynomial time $O(\lambda \log n \log k)$ approximation algorithm for the group Steiner problem, where λ is the path approximation given in Corollary 1.1.*

Proof. We apply Proposition 2.2 to Corollary 1.1 and so we obtain $\alpha = \lambda$ and $\Gamma = \text{poly(n)}$. Using the algorithm for trees from [12] with $r(m) = O(\log m \log k)$, we obtain the claim. $\square$

3 Path Approximation

The following theorem is a restatement of Theorem 1.2 in a more general form.

THEOREM 3.1. *Given any metric space $M = (V, d)$ on $|V| = n$ points and diameter Δ, for any $t \in \mathbb{N}$, M is $O(t \min\{\log\Delta, \log n\})$-path-approximated by an efficiently constructible ultrametric with $\Gamma = n^\beta$ points, where $\beta = \min\{(\log n)^{1/t}, [t\log(4\Delta)]^{2/t}\}$.*

Proof. The construction of the multi-embedding is motivated by the construction of subspaces approximated ultrametric in [4, 5], but instead of deleting points, we *duplicate* them. We then prove that the construction indeed implies a path-approximation.

Let Δ be the diameter of M. Let x and $\bar{x}$ two points realizing the diameter of M, and assume without loss of generality that $|\{y \in M : d(x,y) < \Delta/4\}| \leq n/2$ (otherwise, switch the roles of x and $\bar{x}$). Define a series of sets $A_0 = \{x\}$, and for $i \in \{1, 2, \ldots, t\}$, $A_i = \{y \in M | \; d(x,y) < i\Delta/4t\}$, and "shells" $S_i = A_i \setminus A_{i-1}$. Let $|V| = n$ and let $\varepsilon_i = |A_i|/n$.

The algorithm for constructing the multi-embedding work as follows: Choose a shell S_i $i \in [t]$. Recursively, construct a multi-embedding of the sub space A_i into an ultrametric T_1 and a multi-embedding of the subspace $V \setminus A_{i-1}$ into an ultrametric T_2. To construct the multi-embedding for M, we construct an ultrametric T with root labelled with Δ, and two children, one is T_1 and the other is T_2. This is a multi-embedding since the points in S_i are essentially being "duplicated" at this stage. Note also that it is a dominating multi-embedding.

Next we prove an upper bound on the size of the resulting ultrametric T, assuming the shell was chosen carefully enough. The bound we prove is n^β, where $\beta = \min\{(\log n)^{1/t}, [t\log(4\Delta)]^{2/t}\}$.

We begin with the first bound. Let $\beta = (\log n)^{1/t}$. We claim that there must exists $i \in [t]$ such that $\varepsilon_{i-1} \geq \varepsilon_i^\beta$. Indeed we note that $n^{-1} \leq \varepsilon_0 \leq \varepsilon_{t+1} \leq 1/2$. Assume for the contrary that $\varepsilon_{i-1} < \varepsilon_i^\beta$ for all $i \in [t]$, then

$$\frac{1}{n} \leq \varepsilon_0 < \varepsilon_1^\beta < \cdots \varepsilon_t^{\beta^t} \leq (\frac{1}{2})^{\log n} = \frac{1}{n},$$

which is a contradiction. Therefore we can fix i such that $\varepsilon_{i-1} \geq \varepsilon_i^\beta$. Inductively, assume that the recursive process results in at most $(\varepsilon_i n)^\beta$ leaves in T_1 and at most $((1-\varepsilon_{i-1})n)^\beta$ leaves in T_2. So $|T| \leq (\varepsilon_i^\beta + (1-\varepsilon_{i-1})^\beta) n^\beta$. Since $\varepsilon_{i-1} \geq \varepsilon_i^\beta$, we have $\varepsilon_i^\beta + (1-\varepsilon_{i-1})^\beta \leq \varepsilon_{i-1} + (1-\varepsilon_{i-1}) = 1$ and we are done.

We next prove the second bound. Let $\beta = \beta(\Delta) = [t\log(4\Delta)]^{2/t}$. The proof is by induction on (the rounded value of) Δ. We claim that there is some $i \in [t]$ such that:

$$\varepsilon_{i-1} \geq \varepsilon_i^{\beta(\Delta/2)} n^{\beta(\Delta/2) - \beta(\Delta)}. \tag{3.1}$$

Indeed, assume for the contrary no such i exists. For simplicity set $a = \log(2\Delta) \geq 1$, so that $\beta(\Delta/2) = (ta)^{2/t}$ and $\beta(\Delta) = [t(a+1)]^{2/t}$. Denote $b = n^{(ta)^{2/t} - [t(a+1)]^{2/t}}$ and $c = (ta)^{2/t}$. Our assumption then becomes $\epsilon_{i-1} < \epsilon_i^c b$. Iterating this t times we get:

$$\frac{1}{n} = \epsilon_0 < \epsilon_t^{c^t} b^{1+c+c^2+\ldots+c^{t-1}} \leq \epsilon_t^{c^t} b^{c^{t-1}} \leq b^{c^{t-1}}.$$

So that:

$$n^{(ta)^{2-2/t}\left[[t(a+1)]^{2/t} - (ta)^{2/t}\right]} < n,$$

but an application of the mean value theorem gives a contradiction, since:

$$\begin{aligned}
&(ta)^{2-2/t}\left[[t(a+1)]^{2/t} - (ta)^{2/t}\right] \\
&\geq (ta)^{2-2/t}[2t^{-1+2/t}(a+1)^{-1+2/t}] \\
&= 2(ta)\left(\frac{a}{a+1}\right)^{1-2/t} \geq ta \geq 1.
\end{aligned}$$

Choose an index $i \in \{1, \ldots, t\}$ satisfying (3.1). Since $i \leq t$, $\Delta(A_i) \leq \Delta/2$. The choice of the index i gives the required lower bound on the cardinality of T since:

$$\begin{aligned}
|T| &\leq (\varepsilon_i n)^{\beta(\Delta/2)} + [(1-\varepsilon_{i-1})n]^{\beta(\Delta)} \\
&\leq \varepsilon_{i-1} n^{\beta(\Delta)} + (1-\varepsilon_{i-1}) n^{\beta(\Delta)} \leq n^{\beta(\Delta)}.
\end{aligned}$$

We note that the the running time of the algorithm above is $O(n^2)$ on each vertex in the tree and therefore $O(n^{\beta+2})$ for the whole tree . A slight variation on this algorithm has an $\tilde{O}(n^{\beta+1})$ running time.

We have proven the upper bound on the size of the augmented space. We are left to see that it is indeed a path-approximation of M. To prove it, we need the following properties from the multi-embedding into an ultrametric we have just constructed: let u be an internal vertex in the HST, and T its subtree. Let T_1 and T_2 be the subtrees rooted with the children of u. Then,

1. The tree structure defining the ultrametric is a binary tree.[1]
2. The multi-embedding is dominating.
3. Let x and y two points in M. Further assume that x has representatives in T_1 but not in T_2, and y has representatives in T_2 but not in T_1, then $d(x,y) \geq \Delta(T)/4t$.
4. $|\mathsf{f}(T_1)| \leq |\mathsf{f}(T)|/2$.
5. $\Delta(T_1) \leq \Delta(T)/2$.

[1]Actually, any ultrametric can be defined by a binary tree.

The multi-embedding constructed above has these properties. Let $p = \langle u_1, u_2, \ldots, u_m \rangle$ be a path in M whose length is $\ell(p)$. We construct a path $\bar{p}$ on the leaves of T whose length satisfies $\ell(\bar{p}) \leq 8t \log \min\{n, \Delta\} \cdot \ell(p)$. The proof proceeds by induction on the height of the tree defining the ultrametric.

We partition p to sub-paths as follows. Define a sequence of indices and a sequence of sub-trees of the root: Let $x_1 = 1$, and let $\hat{T}_1 \in \{T_1, T_2\}$ be the subtree of the root that includes the longest prefix of p. Assume inductively that we have already defined x_{i-1} and $\hat{T}_{i-1}$. Define x_i to be the minimum x such that u_{x_i} is the first point in p after $u_{x_{i-1}}$ with no representative in $\hat{T}_{i-1}$. Let $\hat{T}_i$ be the other subtree of the root. Assume this process is finished with $x_{m'}$, $\hat{T}_{m'}$. Next we define another sequence of indexes: $y_{m'} = m$, for $i < m'$ we define y_i to be the largest number, smaller than x_{i+1}, such that u_{y_i} does not have a representative in $\hat{T}_{i+1}$. By the construction of $\hat{T}_i$, we have that $x_i \leq y_i$ and u_{y_i} has a representative in $\hat{T}_i$. We have partitioned p into sub-paths $(\langle u_{x_i}, \ldots, u_{y_i} \rangle)_i$ and $(\langle u_{y_i}, \ldots, u_{x_{i+1}} \rangle)_i$. Informally, a sub-path $\langle u_{x_i}, \ldots, u_{y_i} \rangle$ will be realized in $\hat{T}_i$, while sub-path $\langle u_{y_i}, \ldots, u_{x_{i+1}} \rangle$ will be realized in T_1, which is smaller among $\hat{T}_i$ and $\hat{T}_{i+1}$.

More formally, let $L = \ell(p)$, $L_{i,j} = \ell(\langle u_i, u_{i+1}, \ldots, u_j \rangle)$, $n = |\mathsf{f}(T)|$, $n_1 = |\mathsf{f}(T_1)|$, $n_2 = |\mathsf{f}(T_2)|$, and $\Delta = \Delta(T)$, $\Delta_1 = \Delta(T_1)$, and $\Delta_2 = \Delta(T_2)$. We construct by induction on the tree T a path $\bar{p}$ in T whose length satisfies $\bar{L} = \ell(\bar{p}) \leq 8t \log \min\{\Delta, n\} \cdot L$.

By the induction hypothesis it is possible to construct for any i, a path in $\hat{T}_i$ of representatives of $\langle u_{x_i}, \ldots, u_{y_i} \rangle$ whose length is

$$\bar{L}_{x_i,y_i} \leq L_{x_i,y_i} \cdot 8t \log \min\{n, \Delta\}.$$

Next, we construct for any i a path of representatives of $\langle u_{y_i}, \ldots, u_{x_{i+1}} \rangle$. Note that $u_{y_i+1}, \ldots, u_{x_{i+1}-1}$ have representatives in both $\hat{T}_i$ and $\hat{T}_{i+1}$. As T_1 has smaller number of original points we construct inductively a path from a representative of u_{y_i+1} to a representative of $u_{x_{i+1}-1}$ in T_1. We then connect the representative of u_{y_i} with the representative of u_{y_i+1} and the representative of $u_{x_{i+1}-1}$ with the representative of $u_{x_{i+1}}$, each such edge is of length at most Δ. We have therefore constructed a path of of representatives of $\langle u_{y_i}, \ldots, u_{x_{i+1}} \rangle$ whose length is

$$\bar{L}_{y_i,x_{i+1}} \leq L_{y_i,x_{i+1}} 8t \log\{n_1, \Delta_1\} + 2\Delta$$

Since u_{y_i} does not have a representative in $\hat{T}_{i+1}$ and $u_{x_{i+1}}$ does not have representative in $\hat{T}_i$ and by property (3.) above, $d_M(u_{y_i}, u_{x_{i+1}}) \geq \Delta/4t$, and so $\Delta \leq 4tL_{y_i,x_{i+1}}$. To summarize,

$$\begin{aligned}
\bar{L}_{y_i,x_{i+1}} &\leq L_{y_i,x_{i+1}} 8t \log \min\{n_1, \Delta_1\} + 2\Delta \\
&\leq L_{y_i,x_{i+1}} 8t \big(\log \min\{n/2, \Delta/2\} + 1\big) \\
&= L_{y_i,x_{i+1}} 8t \log \min\{n, \Delta\}.
\end{aligned}$$

We conclude,

$$\begin{aligned}
\bar{L} &= \sum_{i=1}^{m} \bar{L}_{x_i,y_i} + \sum_{i=1}^{m-1} \bar{L}_{y_i,x_{i+1}} \\
&\leq 8t \log \min\{n, \Delta\} \cdot \Big(\sum_i L_{x_i,y_i} + \sum_i L_{y_i,x_{i+1}} \Big) \\
&= 8t \log \min\{n, \Delta\} L. \qquad \square
\end{aligned}$$

4 Probabilistic Path Approximation

The following theorem is a restatement of Theorem 1.3.

THEOREM 4.1. *Given any n-point metric space $M = (V, d)$, for any $\beta > 1$, M is $O(\beta \log n \log\log_\beta \log n)$ probabilistically-path-approximated by a set of ultrametrics with $\Gamma = n^\beta$. Moreover the distribution is efficiently sampled.*

Proof. The proof given here is a synthesis of the technique from [2] with some ideas from the deterministic path approximation, Theorem 1.2. The (many) multi-embeddings into ultrametrics are constructed in a similar manner to the construction of the multi-embedding in Theorem 1.2, where here there will be many candidates S_i on which we will recurse into A_i and $V \setminus A_{i-1}$, and one of these S_i will be picked randomly. Later we shall see how to construct the distribution over the S_i s such that for any path of length L, the expected length of carefully chosen path of duplicates is at most $Lf \log n$, where $f = \gamma\beta \log\log_\beta \log n$, for some constant $\gamma > 0$ to be determined later.

Let Δ be the diameter of M, and let $t = 2\lceil \log_\beta \log n \rceil$. As before. let $x \in M$ an endpoint on the diameter, such that $|\{y : d(x,y) < \Delta/2\}| \leq n/2$. Define a series of sets $A_i = \{y \in M | \; d(x,y) \leq \Delta i/2t\}$, for $i \in [t+1] = \{0, 1, 2, \ldots, t\}$. The idea here is to "randomize" the algorithm in Theorem 1.2. We will show there exists $I \subset [t]$ such that $|I| \geq t/2$ for which for any $i \in I$, $\varepsilon_{i-1} \geq \varepsilon_i^\beta$. The algorithm then find a probability distribution $(p_i)_{i \in I}$ over I, then an S_i, $i \in I$ is chosen with probability p_i. Now, similarly to the algorithm of Theorem 1.2, the algorithm recourses on the subspaces A_i, and $V \setminus A_{i-1}$ sampling a multi-embedding into ultrametrics for each of them and connecting the two trees (sampled from these distributions) under a common new root labelled with Δ.

Let $|V| = n$ and let $\varepsilon_i = |A_i|/n$. Denote by $\hat{n}_i$ the largest number of representatives created by the algorithm above, assuming that in the top level $i \in \{1,\ldots,t\}$ was chosen. First, we shall see the existence of a set $I \subseteq \{1,\ldots,t\}$, where $|I| \geq t/2$ such that $|\hat{n}_i| \leq n^\beta$ for $i \in I$, and furthermore $\varepsilon_{i-1} \geq \varepsilon_i^\beta$. This is done similarly to the proof of Theorem 1.2. Inductively, assume that our procedure creates at most $(\varepsilon_i n)^\beta$ representatives for A_i and at most $((1-\varepsilon_i)n)^\beta$ representatives for $V \setminus A_{i-1}$. Thus we have at most $(\varepsilon_i^\beta + (1-\varepsilon_{i-1})^\beta)n^\beta$ representatives for M. We should find I such that for $i \in I$, $\varepsilon_i^\beta + (1-\varepsilon_{i-1})^\beta \leq 1$.

We claim that there exists $I \subseteq \{1,\ldots,t\}$, $|I| \geq t/2$ such that for $i \in I$, $\varepsilon_{i-1} \geq \varepsilon_i^\beta$. This means that for $i \in I$, $\varepsilon_i^\beta + (1-\varepsilon_{i-1})^\beta \leq \varepsilon_{i-1} + (1-\varepsilon_{i-1}) = 1$ as needed. Assume for the contrary that such a large I does not exist, then $|\{1,\ldots,t\} \setminus I| > t/2$ and for $i \in \{1,\ldots,t\} \setminus I$, $\varepsilon_{i-1} < \varepsilon_i^\beta$. Since $\varepsilon_t \leq 1/2$, we conclude by induction on j that i_j — the jth largest index in $\{1,\ldots,t\} \setminus I$ — satisfies $\varepsilon_{i_j-1} \leq (1/2)^{\beta^{t-j}}$ and so

$$\varepsilon_0 < \varepsilon_{i_1-1} \leq (\tfrac{1}{2})^{\beta^{t/2}} \leq (\tfrac{1}{2})^{\beta^{\log_\beta \log n}} = \frac{1}{n}.$$

But this is a contradiction.

Given an instance of the distribution — a multi-embedding into a specific ultrametric — the construction of the path of representatives is done the same way as in the proof of Theorem 1.2.

Next we analyze the expected stretch of an original path $p = \langle u_1,\ldots,u_m\rangle$, given a that the shell S_i, $i \in I$ is chosen. Recall that in the proof Theorem 1.2, the sequence of points in P were partitioned between A_i and $V \setminus A_{i-1}$. We assign the the length of an edge in p whose two endpoints are in the same "side" to that side. We are left to partition an edge $e = uv$ for which assuming u was assigned to A_i and v to $V \setminus A_{i-1}$. In this case we partition the length of e proportionally between the two sets, i.e., We assign A_i the length $\frac{i\Delta/2t - d(x,u)}{d(x,v)-d(x,u)} d(u,v)$, and the rest to $V \setminus A_{i-1}$

Denote by $L = \ell(p)$, and by $a_i L$ the part of p assigned to A_i, *i.e.*, a_i is the fraction of the length of p assigned to A_i. Let i^- denotes the largest index $k \in I$ such that $k < i$. We also denote by $J = I \cup \{2t\}$ and we extend a_i to $i = 2t$ such that $a_{2t} = 1$. We would like to bound the total length added to the path of duplicates by the term 2Δ for crossing between representatives in A_i to representatives in $V \setminus A_{i-1}$, in terms of $(a_i - a_{i^-})L$. As explained in Theorem 1.2, for each such term of 2Δ added there is disjoint segment of p that "crosses" the shell S_i and thus of length at least $\Delta/2t$. The way we partition the edge lengths ensures that this segment (which may be composed of edges and parts of edges) *is not assigned to* A_{i^-}, and thus we bound the total length added by the 2Δ terms by at most $2(a_i - a_{i^-})2tL$. Let $\bar{L}^i$ denote the expected length of the path of representatives conditioned on $i \in I$ being chosen in the top level. Then by the induction hypothesis,

$$\frac{\bar{L}^i}{L} \leq a_i \ln(\varepsilon_i n) f + (1-a_i)\ln((1-\varepsilon_{i-1})n) f + 4t(a_i - a_{i^-}).$$

Denote $q_i = a_i - a_{i^-}$, and

$$F = \sum_{i\in I}\Big(p_i f \cdot (\sum_{\substack{j\in J,\\ j\leq i}} q_j)\ln\varepsilon_i\Big) + \sum_{i\in T}\Big(p_i f \cdot (\sum_{\substack{j\in J,\\ j> i}} q_j)\ln(1-\varepsilon_{i-1})\Big) + 4t\sum_{i\in I} p_i q_i.$$

Assuming the shell S_i being duplicated is chosen according to a distribution $(p_i)_{i\in I}$. Denote by $\bar{L}$ the expected length of the path of representatives. Then,

$$\frac{\bar{L}}{L} = E_i[\frac{\bar{L}^i}{L}] = f \ln n + F$$

Given $(\varepsilon_i)_i$, we would like to prove the existence of distribution $(p_i)_{i\in I}$ that is also efficiently computable such that for any distribution $(q_i)_{i\in J}$, $F \leq 0$, *i.e.*, $\min_{(p_i)_{i\in I}} \max_{(q_j)_j} F \leq 0$. Note that F can be described as the profit formula of a mixed zero-sum game of two players. Hence, from von-Neuman's min-max theorem,

$$\min_{(p_i)_{i\in I}} \max_{(q_j)_{j\in J}} F = \max_{(q_j)_{j\in J}} \min_{(p_i)_{i\in I}} F.$$

Therefore, in order to prove existence, it is sufficient to show that for any $(q_j)_{j\in J}$ there exists $i \in I$ ($p_i = 1$), such that

$$f \cdot (\sum_{\substack{j:\ j\in J,\\ j\leq i}} q_j)\ln\varepsilon_i + f \cdot (\sum_{\substack{j:\ j\in J,\\ j>i}} q_j)\ln(1-\varepsilon_{i-1}) + 4tq_i \leq 0$$

The inequality above is rephrased as

$$a_i - a_{i^-} \leq \frac{f}{4t}(a_i \ln\varepsilon_i^{-1} + (1-a_i)\ln(1-\varepsilon_{i-1})^{-1}) \tag{4.2}$$

Since $\varepsilon_{i-1} \geq \varepsilon_i^\beta$, it is sufficient to prove that $\exists i \in I$,

$$a_i - a_{i^-} \leq \frac{f}{4t}(a_i \ln\varepsilon_i^{-1} + (1-a_i)\ln(1-\varepsilon_i^\beta)^{-1}) \tag{4.3}$$

In order to prove (4.2) we begin with the initial conditions. Denote by i_l the elements in I ordered in increasing order. If $\varepsilon_{i_1} \leq 1/e^{4t}$ then

$$\frac{f}{4t} a_{i_1} \ln\varepsilon_{i_1}^{-1} \geq \frac{f}{4t} a_{i_1} \ln e^{4t} \geq f a_{i_1} \geq a_{i_1} - a_{i_1^-}.$$

Otherwise, if $a_{i_2} \le 1/(e^{4t}4t)$, then

$$\frac{f}{4t}(1-a_{i_2})\ln(1-\varepsilon_{i_2-1})^{-1} \ge \frac{f}{4te^{4t}}(1-\frac{1}{e^{4t}4t}) \ge a_{i_2}-a_{i_1}.$$

Let $\psi = e^{4t}4t$. We are left with the case of $a_{i_2} \ge 1/\psi$. For $l \in \{0,1,\ldots,\lceil \ln\psi \rceil\}$ we define intervals $I_l = \{i \in I : 2^{-(l+1)} < a_i \le 2^{-l}\}$. It follows that there is some interval I_l of size at least $\frac{|I|}{4(l+1)\ln\ln\psi}+1$, since otherwise the total number of elements in the intervals would be bounded from above by

$$\frac{|I|}{4\ln\ln\psi}\sum_{l=0}^{\ln\psi}\frac{1}{l+1} < |I|-1,$$

a contradiction. Therefore there exists an i in the interval I_l such that $a_i - a_{i^-} \le \frac{4\ln\ln\psi}{|I|}2^{-(l+1)}(l+1)$. On the other hand

$$\begin{aligned}(4.4)\quad & a_i \ln\varepsilon_i^{-1} + (1-a_i)\ln(1-\varepsilon_i^{\beta})^{-1} \\ & \ge 2^{-(l+1)}\ \ln\varepsilon_i^{-1} + (1-2^{-l})\ln(1-\varepsilon_i^{\beta})^{-1}\end{aligned}$$

If $l=0$, the right hand side of (4.4) is bounded below by $\ln 2/2$, since $\varepsilon_i \le 1/2$.

If $l>0$, minimizing the right hand side of (4.4) for ε_i, the minimum is achieved when

$$\varepsilon_i = \Big(\frac{2^{-l}}{2\beta-(2\beta-1)2^{-l}}\Big)^{1/\beta}.$$

Therefore the RHS of (4.4) is bounded below by

$$2^{-(l+1)}\ln\Big(\frac{2\beta-(2\beta-1)2^{-l}}{2^{-l}}\Big)^{1/\beta} \ge 2^{-(l+1)}\frac{l}{2\beta}.$$

We conclude that the RHS of (4.4) is at least $2^{-(l+1)}\frac{l+1}{4\beta}$ for $l \ge 0$. Therefore, in order to satisfies (4.3), f should satisfy $f \ge 4(\ln\ln\psi)\frac{4t}{|I|}4\beta$. Since $\frac{4t}{|I|} \le 8$, choosing $f = 64\beta\ln\ln\psi = O(\beta\log\log_\beta\log n)$ completes the proof of the existence of $(p_i)_{i\in I}$ such that for all $(q_j)_{j\in J}$, $F \le 0$.

Next, we show that it is also computationally feasible to find $(p_i)_i$. From a convexity argument, given a distribution $(p_i)_{i\in I}$, F is maximized by some $q_j = 1$, and therefore it is enough to find $(p_i)_{i\in I}$ such that

$$\begin{aligned}&\forall j\in J \quad f\sum_{\substack{i\in I,\\ i\ge j}} p_i\ln\varepsilon_i \;+\; f\sum_{\substack{i\in I,\\ i<j}} p_i\ln(1-\varepsilon_i) + 2tp_j \le 0\\ &\qquad\sum_{i\in I} p_i = 1,\\ &\forall i\in I \quad p_i \ge 0.\end{aligned}$$

This can be done efficiently using algorithms for linear programming. □

We end the discussion on probabilistic multi embedding into ultrametrics with the following impossibility result.

PROPOSITION 4.1. *Consider n equally spaced points on the line. Then any α-probabilistic-path-approximation of this metric by a set of ultrametrics must have $\alpha = \Omega(\log n)$.*

Proof. Let $M = \{v_1, v_2, \ldots, v_n\}$ be the metric space on n points such that $d_M(v_i,v_j) = |i-j|$. We prove that any dominating multi-embedding into an ultrametric T and any path of representatives of $\langle v_1, v_2, \ldots, v_n\rangle$ is of length of at least $f(n) = \frac{n}{2}\log n$. This also means that any convex combination of dominating multi-embeddings has the same expansion, *i.e.*, it is a lower bound on the probabilistic path approximation.

The proof proceeds by induction on n. For $n=1$ the claim is trivial. For $n>1$, let $\langle v_1', v_2', \ldots, v_n'\rangle$ be a path of duplicates in T. Let $u = \mathrm{lca}_T(v_1', v_n')$, and we have $\Delta(u) = d_T(v_1', v_n') \ge d_M(v_1, v_n) = n-1$. Let T_1 be the subtree of the child of u that contains v_1', T_1 does not contains v_n'. Let $i_1 < n$ be the maximal i such that $v_1', \ldots, v_{i_1}'$ are all in the sub space induced by T_1. As i_1+1 is not contained in T_1, we have that $d_T(v_{i_1}', v_{i_1+1}') \ge \Delta(u) \ge n-1$. By the induction hypothesis

$$\ell_T(\langle v_1', \ldots, v_{i_1}'\rangle) \ge f(i_1), \quad \ell_T(\langle v_{i_1+1}', \ldots, v_n'\rangle) \ge f(n-i_1).$$

Since f is a convex function, $(f(i_1)+f(n-i_1))/2 \ge f((i_1+(n-i_1))/2) = f(n/2)$. We conclude

$$\begin{aligned}&\ell_T(\langle v_1', \ldots, v_n'\rangle) \ge f(i_1) + (n-1) + f(n-i_1)\\ &\ge 2f(n/2)+n-1 = 2\tfrac{n}{4}\log\tfrac{n}{2} + (n-1) \ge \tfrac{n}{2}\log n. \quad \square\end{aligned}$$

5 Multi-Embedding into Trees

In this section we consider multi embeddings into *general tree metrics*. We have only preliminary results. Specifically, we only consider two important types of metric spaces: expander graphs and the discrete cube with the Hamming metric, for which we obtain better results. For both of them the preceding sections proved an upper bound of $O(\log\log n \log\log\log n)$ and a lower bound of $\Omega(\log\log n)$ on the path approximation into ultrametrics (the lower bound follows since both metrics contain a path of length $\Omega(\log n)$).

We begin with the observation that for $\Gamma = \infty$ it is easy to obtain 1-path approximation of any metric space by trees. This is achieved by defining an infinite tree (pseudo) metric as follows: joining all possible finite paths with a common root, where the first node in the path is connected with an edge weight of $\Delta/2$ to the root. Moreover, PROPOSITION 5.1. *Given a metric space M defined by an unweighted graph of max degree d and diameter Δ, then M is $(2+\frac{\Delta}{s})$ path-approximated by a tree metric of size nd^s.*

Proof. Similarly to the construction described above we take all paths of length s, and join these with a common root, where the first node in the path is connected with an edge weight of $\Delta/2$ to the root. Obviously, there are at most nd^s such paths. We claim that the resulting tree is $(2+\frac{\Delta}{s})$ path-approximation of M. To see this, first notice that our definitions of edge weights to the root promise that distances in the resulting tree dominate the original distances. Consider a path $p=\langle v_1,\dots v_\ell\rangle$ of length ℓ. We partition p into sub-paths of length s: $p=p_1p_2\cdots p_t$, where $t=\lceil \ell/s\rceil$, $p_j=\langle v_{(j-1)s+1},\dots,v_{js}\rangle$ for $j<t$, and $p_t=\langle v_{(t-1)s+1},\dots,v_\ell\rangle$. Now we map the sub-path p_j to the appropriate path in the tree. Note that the length of the image path is $2\ell+\lfloor\frac{\ell}{s}\rfloor\Delta$. $\square$

This simple fact is particularly interesting for its implication for expander graphs. Let G be an $(n,d,\gamma d)$ graph, i.e., a d-regular, n-vertex graph whose second eigenvalue in absolute value is at most γd. It is known [9] that such a graph has diameter at most $1+\log_{1/\gamma} n$, and so we obtain a 3 path approximation with $\Gamma = dn^{1+\log_{1/\gamma} d}$.

We also note that for the trees constructed in the proof of Proposition 5.1, it is particulary easy to obtain better approximation algorithm for the group Steiner tree problem.

LEMMA 5.1. *Consider a tree metric space $M=(V,d)$, where $V=P_1\cup P_2\cup\cdots\cup P_\ell$, $P_i=\langle v_0^i,v_2^i,\dots,v_s^i\rangle$ is an unweighted simple path of length s, and $d(v_0^i,v_0^j)=\Delta$ for $i\neq j$. Then an instance of the group Steiner tree problem with k groups defined on M has $(1+\frac{2s}{\Delta})\ln k$ approximation algorithm.*

Proof. Consider a group Steiner tree instance $g_1,\dots g_k$ defined on M. We first check whether there is a solution that is completely contained in one P_i. This can be checked in polynomial time by noting that if an optimal solution is contained in one P_i then it is an interval. Thus all is needed to be checked are $\ell\binom{s+1}{2}$ intervals.

Otherwise, if the optimal solution intersects $t>1$ of the paths $P_1,\dots P_\ell$. Define a Hitting Set instance whose ground set is $\{P_1,\dots,P_\ell\}$ and the subsets are $g'_1,\dots,g'_k$, where $g'_i=\{P_j;\ P_j\cap g_i\neq\emptyset\}$. It follows that the optimal cost of the hitting set problem is at least $t-1$. The Hitting Set problem has a polynomial time $\ln k$ approximation algorithm [15]. Let S be the approximate solution for the hitting set. We define a solution for the group Steiner tree by taking a natural path over $\cup\{P_i;\ P_i\in S\}$. Note that its length is at most $(\Delta+2s)|S|\le(\Delta+2s)(t-1)\ln k$. But the cost of the optimal group Steiner algorithm is at least $(t-1)\Delta$. $\square$

COROLLARY 5.1. *There exist constants $0<c<C$ such that the group Steiner tree problem on expander graphs has polynomial time $C\log k$ approximation algorithm, and it is hard to approximate within $c\log k$.*

Proof. The approximation algorithm follows from Proposition 5.1 and Lemma 5.1. The hardness result follows since an n-point expander graph contains a subset of $n^{\Omega(1)}$ points that is $O(1)$ approximated by an equilateral space. Group Steiner tree problem on Equilateral space is just a standard Hitting Set problem, which is hard to approximate within $(1-\varepsilon)\ln k$. Usage of points not in this subspace ("Steiner points") can improve the approximation factor by at most a factor of two. $\square$

We next examine multi-embedding of the h-dimensional hypercube with $n=2^h$ vertices. Using Proposition 5.1 with $s=h/\log h$, and using $d=h$ and $\Delta=h$, way we obtain $(\log\log n+2)$ path approximation with blow-up $\Gamma(n)=n^2$. Using Lemma 5.1 it also implies a polynomial time $O(\log k\log\log n)$ approximation algorithm to the group Steiner tree problem on the cube. Similar to the the case in expander graphs, it is hard to approximate instances of the group Steiner tree problem on the cube to within $c\log k$, for some constant $c>0$, since the cube contains a subset of $n^{\Omega(1)}$ points that is $O(1)$ approximated by an equilateral space (a code). See the discussion below for better results on multi-embeddings of the hypercube into trees.

6 Discussion

An important open problem is to determine bounds for path approximation by trees of polynomial size. As indicated by the case of expander graphs, such bounds may be better than those for ultrametrics. For example, we are able to improve the result on multi-embedding of the hypercube into trees of polynomial size obtaining an $O(\log^* n)$ path-approximation.

Results on multi-embedding into trees directly reflect on the approximation for the Group Steiner problem. In fact, as shown for expanders and hypercubes it is possible that for special classes of metric spaces, a combination of improved path approximation and a specialized GSP solution would yield nearly tight upper bounds. In fact, our approach to show the (near) tightness of the results in those cases stems from metric Ramsey-type considerations (i.e. the existence of large equilateral subspace). Such considerations are in fact more general [5] and may lead to more results of this flavor.

Further study of multi-embeddings in more general settings and their applications seem attractive directions for future research.

Acknowledgments. We thank Robi Krauthgamer and Assaf Naor for fruitful discussions.

References

[1] Y. Bartal. Probabilistic approximations of metric space and its algorithmic application. In *37th FOCS*, pages 183–193, 1996.

[2] Y. Bartal. On approximating arbitrary metrics by tree metrics. In *Proc' of 30th STOC*, pages 183–193, 1998.

[3] Y. Bartal, A. Blum, C. Burch, and A. Tomkins. A polylog(n)-competitive algorithm for metrical task systems. In *Proc' of 29th STOC*, pages 711–719, May 1997.

[4] Y. Bartal, B. Bollobás, and M. Mendel. A ramsey-type theorem for metric spaces and its application for metrical task systems and related problems. In *Proc' of 42nd FOCS*, 2001.

[5] Y. Bartal, N. Linial, M. Mendel, and A. Naor. On metric Ramsey-type phenomena. Manuscript, 2002.

[6] A. Borodin, N. Linial, and M. Saks. An optimal online algorithm for metrical task systems. *J. Assoc. Comput. Mach.*, 39(4):745–763, 1992.

[7] J. Bourgain. On lipschitz embedding of finite metric spaces in hilbert space. *Israel J. Math.*, 52:46–52, 1985.

[8] M. Charikar, C. Chekuri, A. Goel, S. Guha, and S. Plotkin. Approximating a finite metric by a small number of tree metrics. In *Proc' of 39th FOCS*, 1998.

[9] F. R. K. Chung. Diameters and eigenvalues. *J. Amer. Math. Soc.*, 2(2):187–196, 1989.

[10] U. Feige. A threshold of ln n for approximating set cover. *J. Assoc. Comput. Mach.*, 45(4):634–652, 1998.

[11] A. Fiat and M. Mendel. Better algorithms for unfair metrical task systems and applications. In *Proc' of 32nd STOC*, pages 725–734, 2000.

[12] N. Garg, G. Konjevod, and R. Ravi. A polylogarithmic approximation algorithm for the group steiner tree problem. *J. Algorithms*, 37(1):66–84, 2000.

[13] E. Halperin, G. Kortsarz, R. Krauthgamer, A. Srinivasan, and N. Wang. Integrality ratio for group steiner trees and directed steiner trees. In *Proc' of 14th SODA*, 2003.

[14] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *Proc' of 42nd FOCS*, pages 10–33, 2001.

[15] D.S. Johnson. Approximation algorithms for combinatorial problems. *Journal of Computer ans System Science*, 9:256–278, 1974.

[16] N. Linial. Finite metric spaces - combinatorics, geometry and algorithms. In *ICM*, 2002.

[17] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[18] Y. Rabinovich and R. Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *Discrete Comput. Geom.*, 19(1):79–94, 1998.