Probabilistic Metric Embedding via Metric Labeling

Kamesh Munagala
Department of Computer Science, Duke University, Durham, NC, USA

Govind S. Sankar
Department of Computer Science, Duke University, Durham, NC, USA

Erin Taylor
Department of Computer Science, Duke University, Durham, NC, USA

Abstract

We consider probabilistic embedding of metric spaces into ultra-metrics (or equivalently to a constant factor, into hierarchically separated trees) to minimize the expected distortion of any pairwise distance. Such embeddings have been widely used in network design and online algorithms. Our main result is a polynomial time algorithm that approximates the optimal distortion on any instance to within a constant factor. We achieve this via a novel LP formulation that reduces this problem to a probabilistic version of uniform metric labeling.

2012 ACM Subject Classification Theory of computation → Random projections and metric embeddings

Keywords and phrases Metric Embedding, Approximation Algorithms, Ultrametrics

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2023.2

Category APPROX

Funding This work is supported by NSF grant CCF-2113798.

1 Introduction

Embedding a finite metric space into simpler spaces such as trees, ultrametrics, and Euclidean spaces (called "target metrics") has a wide range of applications, and has been widely studied. In such an embedding, the distance between any pair of points should be at least as large as in the original space, while being at most a factor of α larger, where α is termed the *distortion* of the embedding. The goal is to design an embedding into a given target metric whose distortion is as small as possible. We will denote by n the number of points in the metric space.

A lot of attention has focused on probabilistic embeddings, which construct a distribution over metrics from the target space, for instance, a distribution over trees or ultrametrics. The goal is now to bound the *expected distortion* for any pair of points relative to their distance in the original metric space. Probabilistic embeddings typically allow for much lower values of distortion. Indeed, when the target metric is a tree or an ultrametric, deterministic embeddings have distortion $\Omega(n)$ [2], while probabilistic embeddings have distortion $\alpha = O(\log n)$ [20].

In this paper, we consider the problem of embedding metrics into a distribution over ultrametrics, defined in Section 2. To within a constant factor on distortion, these metrics embed into *hierarchically separated trees* (HSTs), and such embeddings have found myriad uses in network design, data analysis and online algorithms. This is because most network design problems such as Steiner tree or facility location are NP-Hard in general metric spaces, but amenable to polynomial time algorithms on trees. If the objective function is a linear combination of distances, the solution on the distribution over HSTs yields a α approximation

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2023).
Editors: Nicole Megow and Adam D. Smith; Article No. 2; pp. 2:1–2:10
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

algorithm for the input metric space, where α is the distortion of the embedding. Indeed, recent breakthroughs in developing competitive algorithms for the celebrated randomized k-server problem proceed via probabilistic embedding into HSTs [16, 4]. Other applications include analysis of hierarchical clusterings [10, 31, 17], and approximation algorithms for group Steiner trees [21], buy-at-bulk network design [3], and metric labeling [27].

In this context, it is known that in the worst case over input metric spaces, the distortion of embedding into a distribution over ultrametrics is $\alpha = O(\log n)$ [20], and this bound is tight [25, 5]. However, these algorithms typically use fixed ball-growing procedures oblivious to the actual metric. It is conceivable that any given metric can be embedded with much lower distortion than such procedures imply. For instance, an entirely different algorithm can embed doubling metrics into ultrametrics with distortion $O(\log \Delta)$, where Δ is the spread[1] of the point set; this can be significantly better if $\Delta = o(n)$ or independent of n [26]. Further, many real-world social graphs have small diameter, often independent of the network size n [32, 30] and again, it is conceivable that specifically tailored algorithms can embed these better than what the worst case bounds imply.

In this paper, we therefore ask: *Can we achieve the best possible (in terms of distortion) probabilistic embedding of a given discrete metric space into ultrametrics in polynomial time?*

1.1 Result

Our main result is positive:

▶ **Theorem 1.** *Given any n-point metric space that can be embedded into a distribution over ultrametrics with optimal distortion α, there is an algorithm with expected polynomial running time[2] that can find an embedding of distortion $16 \cdot \alpha$.*

Prior to our work, the best published approximation factor was $O(\log n)$ [20], which is also the tight existence result. In contrast, we provide a nearly tight *computational* result, which also yields improved instance-dependent approximation factors for problems like metric labeling and buy-at-bulk network design, where the factor of $O(\log n)$ in the approximation ratio improves to $O(\alpha)$.

1.2 Technique

Our algorithm for proving Theorem 1 proceeds via constructing an LP relaxation. In this LP relaxation, the variables $\gamma^r_{jj'}$ represent the probability that j, j' are separated in the HST at radius (or level) r. This LP is our main non-trivial contribution. Note that though there are LP formulations that embed metrics into trees [11, 27], we do not know how to solve them to a $o(\log n)$ approximation factor.

We then use the randomized rounding technique for uniform metric labeling in [27] in decreasing order of radius to construct a distribution over centers and assignments, while relaxing $\gamma^r_{jj'}$ by a factor of 2. In the uniform metric labeling problem, the goal is to assign one of k labels to the vertices of a graph with edge weights so that the total weight of the edges whose end-points have different labels is minimized. In the algorithm of [27], they encode the probability that two endpoints of an edge are separated as a random variable

---

[1] Spread is the ratio of the largest to smallest distance in the metric space.

[2] For any $\delta > 0$, the algorithm can easily be converted to a $16 + \delta$ approximation in deterministic polynomial time, with the polynomial depending on n and $\log \frac{\Delta}{\delta}$, where Δ is the ratio of largest to smallest distance in the metric space.

in their LP, and then minimize an expected cost over these random variables. Our main idea is that the event whether two nodes belong to the same subtree in the HST can be similarly treated as a random variable in our LP. The expected "cost" of the labeling is the contribution of these variables to the distortion, and these variables can be randomly rounded via the same ideas as for metric labeling.

Note that the algorithm for general metric labeling [27] – where the labels lie in a metric space and the weight of the edge is multiplied by the distance between the corresponding labels – proceeds via embedding the metric over labels into an ultrametric. We effectively reduce in the reverse direction and show that metric embedding reduces to (uniform) metric labeling!

1.3 Related Work

The main technique for probabilistically embedding metrics into ultrametrics is low-diameter decompositions. These involve decomposing the input metric space into small diameter components via sampling radii from a suitable distribution and randomly partitioning based on these radii. Starting with a result of [5], a sequence of results showed improving bounds for such embeddings [6, 11, 20], culminating in the optimal distortion bound of $O(\log n)$. Better results are known for special types of metrics [26]. All these decompositions proceed by deriving absolute bounds on the probability that a pair of nodes end up in different partitions; in contrast, we encode this probability as variables in an LP formulation.

Trees are more general spaces than ultrametrics, and the seminal work of [2] initiated the study of embedding metrics into trees. Though the worst case bound for embedding into a distribution over trees remains $\Theta(\log n)$, better results are known for embedding specific classes of metrics such as the shortest path metrics of k-outerplanar or small pathwidth graphs [13, 23, 29]. The work of [11] provides an LP formulation for computing the optimal distortion for embedding into a distribution over trees; however, their separation oracle – the minimum communication cost spanning tree problem – is unlikely to admit a $o(\log n)$ approximation. Our LP for ultrametrics, in contrast, is inspired by stochastic optimization and metric labeling and shows a constant approximation.

The above works provide worst case guarantees on the distortion of the embedding, while we provide an approximation result. Though other prior work has considered such approximation guarantees [19, 24, 1, 15], these works focus on *deterministic* embeddings, while ours consider approximations for probabilistic embeddings. The work of [1] provides a polynomial time algorithm for optimally embedding a metric into a single ultrametric, and the work of [15] shows an improved running time for obtaining such an embedding. However, the optimal distortion for embedding into a single ultrametric might be $\Omega(n)$, while it is $O(\log n)$ for embedding into a distribution. This result motivates the need for algorithms that approximate the optimal probabilistic embedding. We note that the algorithm of [1] does not extend to probabilistic embeddings.

Our LP is also similar in spirit to those in [28]. They give a 2-approximation to separating decompositions (see [28] for formal definitions) by writing an LP using variables similar to our variables γ for the separation probabilities. The key observation we make is that unlike in their setting where there is a fixed separation probability, we need to optimize over the separation probabilities at all levels of the HST simultaneously.

Stochastic Optimization. Our algorithms involve rounding a linear programming relaxation to the optimal solution. This LP is inspired by similar LPs approximate stochastic optimization, particularly those for stochastic knapsack [18], multi-armed bandits [22], Bayesian auctions [9, 12, 8], and stochastic matching [14]. The novel aspects of our work is the formulation of probabilistic embeddings as a stochastic optimization problem, and viewing the uniform metric labeling algorithm as a stochastic rounding procedure [27].

2 Terminology

▶ **Definition 2** (Metric space). *A metric space (N, d) is a finite set of n points N endowed with a distance function $d : N \times N \to \mathbb{R}^+ \cup \{0\}$. This distance function has the following properties:*

- $d(x, x) = 0$ *for all* $x \in N$;
- $d(x, y) = d(y, x)$ *for all* $x, y \in N$; *and*
- $d(x, z) \leq d(y, z) + d(x, y)$ *for all* $x, y, z \in N$.

▶ **Definition 3** (Embedding). *Given two metric spaces (N, d) and (T, d_T), an embedding from N to T is a function $f : N \to T$.*

With an abuse of notation, for $x, y \in N$, we use $d_T(x, y)$ to refer to $d_T(f(x), f(y))$.

Our goal is to *embed* a given n point metric space into a probability distribution over ultrametrics. An ultrametric and probabilistic embedding are defined below.

▶ **Definition 4** (Ultrametric). *A metric space (N, d) is an* ultrametric *if for all points $x, y, z \in N$, we have $d(x, z) \leq \max(d(x, y), d(y, z))$.*

▶ **Definition 5** (Probabilistic embedding). *Given a metric space (N, d), an embedding is a distribution over ultrametrics (N, d_T), where ultrametric T is chosen with probability p_T. Let F denote this distribution and S denote its support. The embedding should be* non-contractive, *meaning that*

$$\forall x, y \in N, \quad \forall T \in S, \; d_T(x, y) \geq d(x, y).$$

Further, this embedding has distortion α (where $\alpha \geq 1$) if

$$\forall x, y \in N, \mathsf{E}_{T \sim F}[d_T(x, y)] \leq \alpha \cdot d(x, y).$$

Hierarchically Separated Trees. It is convenient to consider a specific type of ultrametric termed *hierarchically separated trees.* These are defined as follows.

▶ **Definition 6** (exact c-HST). *A metric (N, d) is an exact c-HST (for $c > 1$) if the elements of N are the leaves of a rooted tree T, all of whose leaves are at the same level. Each internal node v of T is associated with a number δ_v. These numbers increase by a factor of exactly c as we move up the tree, so that $\delta_v = c \cdot \delta_u$ whenever u is a child of v. Given leaves $x, y \in N$, let z be their least common ancestor in T. Then $d(x, y) = \delta_z$.*

The diameter of a c-HST with root r is δ_r. Note that a c-HST with diameter D decomposes into c-HSTs with diameter D/c, where points in different parts are separated by distance exactly D.

The following result shows that it suffices to consider embedding into exact c-HSTs.

▶ **Lemma 7** ([7]). *Given a metric space and its embedding into a distribution over ultrametrics with distortion α, there is an embedding into a distribution over exact c-HSTs with distortion $\alpha \cdot c$.*

3 Linear Programming Relaxation

The LP relaxation is not obvious, and we will present it in some detail. We are given a metric space (N, d) whose smallest distance is 1 and largest distance is Δ. We will use exact c-HSTs for $c = 2$; using any other value of c only yields a worse approximation factor, and our presentation is simplified without the parameter c. By losing a factor of 2, we assume that the optimal solution embeds this metric into a distribution over exact 2-HSTs. Let this embedding have distortion $q \leq 2\alpha$ if the optimal distortion for probabilistically embedding into ultrametrics is α. Our goal therefore is to approximate q in polynomial time.

Given this optimal probabilistic embedding into 2-HSTs, let M denote the set of possible δ_z values in powers of 2. Note that $|M| = O(\log \Delta)$, since $1 \leq \delta_r \leq 2\Delta$ without loss of generality.[3]

Consider some 2-HST in the optimal embedding of (N, d) and some $r \in M$. For any sub-tree T' whose root z has $\delta_z = r$, all nodes in T' have distance in the 2-HST at most r from some node $i \in T'$. Since this embedding is non-contractive, we have $d(i, j) \leq r$ for all $j \in T'$. We arbitrarily pick one i in this subtree as the "representative" of this subtree and "assign" all other nodes in the subtree to i. Therefore, each node j is assigned to one representative i at each level $r \in M$ in the tree and this satisfies $d(i, j) \leq r$. Now we can define a graph $G_r(N, E_r)$ for each $r \in M$, where there is an edge $(i, j) \in E_r$ if and only if $d(i, j) \leq r$. Let $B_i(r) = \{j \in N, d(i, j) \leq r\}$.

For $(i, j) \in E_r$, define a variable x^r_{ij} as the probability that j's representative at level r is i. Similarly, let $z^r_{ijj'}$ be the probability that both j, j' have i as their level r representative, where we assume $(i, j) \in E_r$ and $(i, j') \in E_r$. Finally, let $\gamma^r_{jj'}$ be the probability that j and j' do not share a level r representative.

The LP relaxation with variables $x^r_{ij}, z^r_{ijj'}$, and $\gamma^r_{jj'}$, and is shown in Figure 1.

$$
\text{(LP1)} \quad
\begin{aligned}
& \text{Minimize } q && (1)\\
\forall j, j', \quad & \sum_{r \in M} r \cdot \gamma^r_{jj'} \leq q \cdot d(j, j') && (2)\\
\forall i, r, j, j' \in B_i(r) \quad & \min(x^r_{ij}, x^r_{ij'}) \geq z^r_{ijj'} && (3)\\
\forall j, j', r \quad & \sum_{i: j, j' \in B_i(r)} z^r_{ijj'} \geq 1 - \gamma^r_{jj'} && (4)\\
\forall j, r \quad & \sum_{i: j \in B_i(r)} x^r_{ij} = 1 && (5)\\
\forall j, j', r < d(j, j') \quad & \gamma^r_{jj'} = 1 && (6)\\
\forall i, j, j', r \quad & \gamma^r_{jj'}, x^r_{ij}, z^r_{ijj'} \geq 0. && (7)
\end{aligned}
$$

Figure 1 Linear program relaxation for embedding into 2-HSTs.

[3] To see this, consider any HST in the support of the optimal probabilistic embedding. We can contract all the internal nodes v of the HST with $\delta_v \geq 2\Delta$ into one node r with $\delta_r = 2\Delta$. Clearly, this preserves the non-contractivity property, since no pair of vertices are more than Δ far apart. Furthermore, this can only decrease the expected distortion of any edge.

▶ **Lemma 8.** *LP1 is feasible and is a 2-approximation to the distortion of the optimal probabilistic embedding into ultrametrics.*

Proof. Consider an optimal embedding of the input metric into 2-HSTs with distortion q. As mentioned before, since q is a factor 2 approximation to optimal distortion of embedding into ultrametrics, this means the objective is a 2-approximation. We only need to show that there is a feasible solution to the LP with objective at most q.

Now, consider any 2-HST in the support of the optimal embedding. We interpret the variables as described before. To interpret Equation (2), note that if (j, j') have least common level r, they are separated at levels $r/2, r/4, r/8$, and so on. Therefore, the contribution this 2-HST makes to the LHS of Equation (2) is $r/2 + r/4 + \cdots \leq r$. Taking expectation over all 2-HSTs in the optimal embedding, we have $\mathsf{E}[r] \leq q \cdot d(i, j)$, where q is the distortion of this embedding. This shows that the constraint holds.

Equation (6) captures that the embedding is non-contractive: If $d(j, j') < r$, then they are separated at level r. The remaining constraints are interpreted as follows. Equation (3) says j, j' are co-assigned to i implies they were both individually assigned to i; Equation (5) says each $j \in N$ has a representative at each level r; and Equation (4) says that the probabilities that j, j' are co-assigned and not co-assigned at level r sum to at least 1.

One feasible solution to the LP is to assign each node j to itself at all levels, so that $x^r_{jj} = 1$ for all $j \in N, r \in M$. Then $z^r_{ijj'} = 0$ and $\gamma^r_{jj'} = 1$ for all $j \neq j'$. This is feasible when $q = 2\Delta$. ◀

4 Rounding and HST Construction

We first present a procedure in Algorithm 1 that generates partitions separately for each level r. Essentially, our algorithm solves $|M|$ instances of metric labeling, one for each level r. This step adapts the rounding scheme in [27] for uniform metric labeling. Each "label" is a possible representative at level r. The expected cost of this metric labeling is then used to bound the expected distortion of the embedding.

To construct the HST itself, we inductively compute the tree in the following manner: At level 2Δ, every node is assigned the same representative. For every lower level r, the set of nodes assigned to the same representative at level $\frac{r}{2}$ belong to the same subtree. These nodes then assigned new representatives at level r based on the metric labeling.

Algorithm 1 Rounding to Create Partitions.

for $r \in M$ in decreasing order **do**
  $S \leftarrow N$; $P^r_i = \emptyset$ for all $i \in N$
  while $S \neq \emptyset$ **do**
    Choose a center $i \in N$ uniformly at random independent of past choices.
    Choose $\ell^r_i \in [0, 1]$ uniformly at random independent of past choices.
    For each $j \in S \cap B_i(r)$, if $x^r_{ij} \geq \ell^r_i$, assign j to P^r_i and remove j from S.
  end while
end for

We next combine these partitions into a 2-HST in Algorithm 2.

Analysis

We first consider Algorithm 1. The lemma below follows directly from the analysis of the rounding algorithm for uniform metric labeling in [27]. For completeness, we provide the proof here.

Algorithm 2 Combining Partitions and Constructing the 2-HST.

Place a root node w at the highest level with $\delta_w = 2\Delta$ and set $S_w = N$.
for $r \in M$ in decreasing order **do**
  for each node w at the previous (parent) level with set S_w **do**
    for each i with $P_i^r \cap S_w \neq \emptyset$ **do**
      Place a child node v with set $S_v = P_i^r \cap S_w$ and $\delta_v = 2r$
    end for
  end for
end for

▶ **Lemma 9** (Lemma 3.2 in [27]). *Consider some pair j, j' with $d(j, j') = R$. Consider some level $r \geq R$. For any $(j, j') \in E(G)$,*

$$\Pr[j, j' \text{ separated at level } r] \leq 2\gamma^r_{jj'}.$$

Proof. We assume $|C| = n \geq 2$; the result is trivial for $n = 1$. Since we fix a phase r, we will omit the superscript r from the proof below.

Suppose that both $j, j' \in S$ at some point in time. Then j is assigned to i with probability $\frac{x_{ij}}{n}$, where $\frac{1}{n}$ is the probability that i is chosen; conditioned on this, j is assigned to i if $\ell_i \leq x_{ij}$. Therefore, j is assigned to some center at this step with probability is $\sum_{i:j\in B_i} \frac{x_{ij}}{n} = \frac{1}{n}$.

Similarly, conditioned on both $j, j' \in S$, the probability with which they are assigned to center $i \in C$ is $\frac{\min(x_{ij}, x_{ij'})}{n} \geq \frac{z_{ijj'}}{n}$, so this pair is co-assigned with probability at least $\frac{\sum_i z_{ijj'}}{n} \geq \frac{1-\gamma_{jj'}}{n}$. This is therefore a lower bound on the probability with which both j and j' get assigned this step.

By the inclusion-exclusion principle on the pair (j, j'), conditioned on both $j, j' \in S$, the probability with which either j or j' gets assigned is:

$$\Pr[\text{Either } j \text{ or } j' \text{ assigned } | j, j' \in S] \leq \frac{1}{n} + \frac{1}{n} - \frac{1-\gamma_{jj'}}{n} = \frac{1+\gamma_{jj'}}{n}. \tag{8}$$

Since $n \geq 2$ and $\gamma_{jj'} \leq 1$, the RHS above is at most 1. Therefore, the probability that both $j, j' \in S$ at time t is at least $\left(1 - \frac{1+\gamma_{jj'}}{n}\right)^{t-1}$. Conditioned on this event, they are co-assigned at time step t with probability at least $\frac{1-\gamma_{jj'}}{n}$. Therefore, we have

$$\Pr[j, j' \text{ not separated }] \geq \sum_{t=1}^{\infty} \left(1 - \frac{1+\gamma_{jj'}}{n}\right)^{t-1} \cdot \frac{1-\gamma_{jj'}}{n} = \frac{1-\gamma_{jj'}}{1+\gamma_{jj'}}.$$

Noting that $\Pr[j, j' \text{ separated}] \leq 1 - \frac{1-\gamma_{jj'}}{1+\gamma_{jj'}} \leq 2\gamma_{jj'}$, this completes the proof. ◀

The following two lemmas will now complete the proof of Theorem 1.

▶ **Lemma 10.** *The construction in Algorithm 1 and Algorithm 2 is non-contractive.*

Proof. Suppose $d(j, j') = R$. Consider some level $r \in M \cap [R/4, R/2]$. For this value of r, there is no center i such that $j, j' \in B_i(r)$. Therefore, j, j' lie in different partitions at this level. Observing that Algorithm 2 sets $\delta_v = 2r$ for nodes v at level r, their common ancestor u must have $\delta_u \geq 4r \geq R$. Therefore, the embedding is non-contractive. ◀

▶ **Lemma 11.** *In the output of Algorithm 2, the expected distortion of any distance is at most $8q$. This implies a 16 approximation to the optimal embedding into a distribution over ultrametrics.*

Proof. Consider some pair j, j' with $d(j, j') = R$. Consider some level $r \geq R$. By Lemma 9,

$$\Pr[j, j' \text{ separated at level } r] \leq 2\gamma^r_{jj'}.$$

If $r < R$, note that $\gamma^r_{jj'} = 1$, so the above inequality trivially holds.

If r^* is the highest level at which j, j' are separated, the distance in the embedding is $4r^*$. As an upper bound, we simply add a distance of $4r$ for all levels r at which j, j' are cut. This yields:

$$\mathsf{E}[\text{ Distance in embedding between } (j, j')] \leq \sum_{r \in M} 4r \cdot 2\gamma^r_{jj'} \leq 8q \cdot d(j, j').$$

Since q itself is a 2 approximation to the optimal distortion of embedding into ultrametrics, this implies a 16 approximation to the distortion of embedding into a distribution over ultrametrics. ◀

Running Time. Since each j gets assigned with probability $1/n$ each step, the expected number of steps is $O(n \log n)$ per level, and there are $O(\log \Delta)$ levels. Suppose we stop the process after $c \cdot n \ln \frac{\Delta}{\delta}$ steps. Then for large constant c, the probability that at some level, all j have not been assigned is at most $\frac{\delta}{\Delta}$. In this event, we pretend the ultrametric distorts all distances to Δ. The expected distortion now becomes a $16 + \delta$ approximation. This completes the proof of Theorem 1.

5 Conclusion

We have in effect reduced probabilistic metric embeddings to metric labeling, performing the reverse of the reduction in [27] that reduces metric labeling to metric embedding. There are some open questions that arise from this work. First, is there an *exact* polynomial time algorithm for embedding into ultrametrics or even a PTAS? Second, can similar results be obtained for the more general problem of embedding into tree metrics?

References

1. Noga Alon, Mihai Bădoiu, Erik D. Demaine, Martin Farach-Colton, Mohammadtaghi Hajiaghayi, and Anastasios Sidiropoulos. Ordinal embeddings of minimum relaxation: General properties, trees, and ultrametrics. *ACM Trans. Algorithms*, 4(4), August 2008.
2. Noga Alon, Richard M. Karp, David Peleg, and Douglas West. A graph-theoretic game and its application to the k-server problem. *SIAM Journal on Computing*, 24(1):78–100, 1995.
3. B. Awerbuch and Y. Azar. Buy-at-bulk network design. In *Proceedings 38th Annual Symposium on Foundations of Computer Science*, pages 542–547, 1997. `doi:10.1109/SFCS.1997.646143`.
4. Nikhil Bansal, Niv Buchbinder, Aleksander Madry, and Joseph (Seffi) Naor. A polylogarithmic-competitive algorithm for the k-server problem. *J. ACM*, 62(5), November 2015.
5. Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of 37th Conference on Foundations of Computer Science*, pages 184–193, 1996. `doi:10.1109/SFCS.1996.548477`.
6. Yair Bartal. On approximating arbitrary metrices by tree metrics. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, pages 161–168, New York, NY, USA, 1998. Association for Computing Machinery. `doi:10.1145/276698.276725`.
7. Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric Ramsey-type phenomena. *Annals of Mathematics*, 162(2):643–709, 2005.

8 Sayan Bhattacharya, Gagan Goel, Sreenivas Gollapudi, and Kamesh Munagala. Budget constrained auctions with heterogeneous items. In Leonard J. Schulman, editor, *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 379–388. ACM, 2010.

9 Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. Optimal multi-dimensional mechanism design: Reducing revenue to welfare maximization. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 130–139. IEEE Computer Society, 2012.

10 Gunnar Carlsson and Facundo Mémoli. Characterization, stability and convergence of hierarchical clustering methods. *Journal of Machine Learning Research*, 11(47):1425–1470, 2010. URL: `http://jmlr.org/papers/v11/carlsson10a.html`.

11 Moses Charikar, Chandra Chekuri, Ashish Goel, Sudipto Guha, and Serge Plotkin. Approximating a finite metric by a small number of tree metrics. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, FOCS '98, page 379, USA, 1998. IEEE Computer Society.

12 Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. Multi-parameter mechanism design and sequential posted pricing. In Leonard J. Schulman, editor, *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 311–320. ACM, 2010.

13 Chandra Chekuri, Anupam Gupta, Ilan Newman, Yuri Rabinovich, and Alistair Sinclair. Embedding k-outerplanar graphs into l1. *SIAM Journal on Discrete Mathematics*, 20(1):119–136, 2006. `doi:10.1137/S0895480102417379`.

14 Ning Chen, Nicole Immorlica, Anna R Karlin, Mohammad Mahdian, and Atri Rudra. Approximating matches made in heaven. In *International Colloquium on Automata, Languages, and Programming*, pages 266–278. Springer, 2009.

15 Vincent Cohen-Addad, C. S. Karthik, and Guillaume Lagarde. On efficient low distortion ultrametric embedding. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020.

16 Aaron Coté, Adam Meyerson, and Laura Poplawski. Randomized k-server on hierarchical binary trees. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 227–234, New York, NY, USA, 2008. Association for Computing Machinery.

17 Sanjoy Dasgupta. A cost function for similarity-based hierarchical clustering. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 118–127. ACM, 2016. `doi:10.1145/2897518.2897527`.

18 Brian C Dean, Michel X Goemans, and Jan Vondrák. Approximating the stochastic knapsack problem: The benefit of adaptivity. *Mathematics of Operations Research*, 33(4):945–964, 2008.

19 Kedar Dhamdhere, Anupam Gupta, and R Ravi. Approximation algorithms for minimizing average distortion. *Theory of Computing Systems*, 39(1):93–111, 2006.

20 Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *Journal of Computer and System Sciences*, 69(3):485–497, 2004. Special Issue on STOC 2003. `doi:10.1016/j.jcss.2004.04.011`.

21 Naveen Garg, Goran Konjevod, and R. Ravi. A polylogarithmic approximation algorithm for the group steiner tree problem. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '98, pages 253–259, USA, 1998. Society for Industrial and Applied Mathematics.

22 Sudipto Guha and Kamesh Munagala. Approximation algorithms for budgeted learning problems. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 104–113, 2007.

23 Anupam Gupta, Ilan Newman, Yuri Rabinovich, and Alistair Sinclair. Cuts, trees and ℓ_1-embeddings of graphs. *Combinatorica*, 24(2):233–269, April 2004. `doi:10.1007/s00493-004-0015-x`.

24 Alexander Hall and Christos Papadimitriou. Approximating the distortion. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 111–122. Springer, 2005.

25 Makoto Imase and Bernard M. Waxman. Dynamic steiner tree problem. *SIAM Journal on Discrete Mathematics*, 4(3):369–384, 1991. doi:10.1137/0404033.

26 Piotr Indyk, Avner Magen, Anastasios Sidiropoulos, and Anastasios Zouzias. Online embeddings. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 246–259, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

27 Jon Kleinberg and Éva Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and markov random fields. *J. ACM*, 49(5):616–639, September 2002.

28 Robert Krauthgamer and Tim Roughgarden. Metric clustering via consistent labeling. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '08, pages 809–818, USA, 2008. Society for Industrial and Applied Mathematics.

29 James R. Lee and Anastasios Sidiropoulos. Pathwidth, trees, and random embeddings. *arXiv e-prints*, page arXiv:0910.1409, October 2009. doi:10.48550/arXiv.0910.1409.

30 Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. Kronecker graphs: An approach to modeling networks. *J. Mach. Learn. Res.*, 11:985–1042, March 2010.

31 Aurko Roy and Sebastian Pokutta. Hierarchical clustering via spreading metrics. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 2324–2332, Red Hook, NY, USA, 2016. Curran Associates Inc.

32 Duncan J. Watts and Steven H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393(6684):440–442, 1998.