Clan Embeddings into Trees, and Low Treewidth Graphs*

Arnold Filtser
Columbia University
New York City, NY, USA
arnold273@gmail.com

Hung Le
University of Massachusetts at Amherst
Amherst, Massachusetts, USA
hungle@cs.umass.edu

ABSTRACT

In low distortion metric embeddings, the goal is to embed a host "hard" metric space into a "simpler" target space while approximately preserving pairwise distances. A highly desirable target space is that of a tree metric. Unfortunately, such embedding will result in a huge distortion. A celebrated bypass to this problem is stochastic embedding with logarithmic expected distortion. Another bypass is Ramsey-type embedding, where the distortion guarantee applies only to a subset of the points. However, both these solutions fail to provide an embedding into a single tree with a worst-case distortion guarantee on all pairs. In this paper, we propose a novel third bypass called *clan embedding*. Here each point x is mapped to a subset of points $f(x)$, called a *clan*, with a special *chief* point $\chi(x) \in f(x)$. The clan embedding has multiplicative distortion t if for every pair (x, y) some copy $y' \in f(y)$ in the clan of y is close to the chief of x: $\min_{y' \in f(y)} d(y', \chi(x)) \le t \cdot d(x, y)$. Our first result is a clan embedding into a tree with multiplicative distortion $O(\frac{\log n}{\epsilon})$ such that each point has $1 + \epsilon$ copies (in expectation). In addition, we provide a "spanning" version of this theorem for graphs and use it to devise the first compact routing scheme with constant size routing tables.

We then focus on minor-free graphs of diameter prameterized by D, which were known to be stochastically embeddable into bounded treewidth graphs with expected additive distortion ϵD. We devise Ramsey-type embedding and clan embedding analogs of the stochastic embedding. We use these embeddings to construct the first (bicriteria quasi-polynomial time) approximation scheme for the metric ρ-dominating set and metric ρ-independent set problems in minor-free graphs.

CCS CONCEPTS

• **Theory of computation → Routing and network design problems**; **Random projections and metric embeddings**.

KEYWORDS

Metric embeddings, Clan Embedding, Ramsey Type Embedding, Minor-free Graphs, Treewidth, Metric ρ-dominating set, Metric ρ-isolated set, Compact Routhing Scheme.

*The reader is encourged to read the full version of the paper [49].

STOC '21, June 21–25, 2021, Virtual, Italy

ACM ISBN 978-1-4503-8053-9/21/06…$15.00
https://doi.org/10.1145/3406325.3451043

ACM Reference Format:
Arnold Filtser and Hung Le. 2021. Clan Embeddings into Trees, and Low Treewidth Graphs. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC '21), June 21–25, 2021, Virtual, Italy.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3406325.3451043

1 INTRODUCTION

Low distortion metric embeddings provide a powerful algorithmic toolkit, with applications ranging from approximation/sublinear/online/distributed algorithms [7, 30, 59, 63] to machine learning [52], biology [54], and vision [9]. Classically, we say that an embedding f from a metric space (X, d_X) to a metric space (Y, d_Y) has multiplicative distortion t, if for every pair of points $u, v \in X$ it holds that $d_X(u, v) \le d_Y(f(u), f(v)) \le t \cdot d_X(u, v)$. Typical applications of metric embeddings naturally have the following structures: take some instance of a problem in a "hard" metric space (X, d_X); embed X into a "simple" metric space (Y, d_Y) via a low-distortion metric embedding f; solve the problem in Y, and "pull-back" the solution in X. Thus, the objectives are low distortion and "simple" target space.

Simple target spaces that immediately come to mind are Euclidean space and tree metric, or — even better — an ultrametric. [1] In a celebrated result, Bourgain [28] showed that every n-point metric space embeds into Euclidean space with multiplicative distortion $O(\log n)$ (which is tight [63]). On the other hand, any embedding of the n-vertex cycle graph C_n into a tree metric will incur multiplicative distortion $\Omega(n)$ [68]. Karp [56] observed that deleting a random edge from C_n results in an embedding into a line with expected distortion 2 (see Figure 1(a)). This idea was developed by Bartal [14, 15] (improving over [6]), and culminating in the celebrated work of Fakcharoenphol, Rao, and Talwar [46] (see also [16]) who showed that every n-point metric space stochastically embeds into trees (actually ultrametrics) with expected multiplicative distortion $O(\log n)$. Specifically, there is a distribution $\mathcal{D}$, over dominating metric embeddings [2] into trees (ultrametrics), such that $\forall u, v \in X$, $\mathbb{E}_{(f,T)\sim\mathcal{D}} d_T(f(u), f(v)) \le O(\log n) \cdot d_X(u, v)$. The $O(\log n)$ multiplicative distortion is known to be optimal [14]. Stochastic embeddings into trees are widely successful and have found numerous applications (see e.g. [55]).

In many applications of metric embeddings, a worst-case distortion guarantee is required. A different type of compromise (compared to expected distortion) is provided by *Ramsey-type* embeddings. The classical Ramsey problem for metric spaces was introduced by Bourgain *et al.* [29], and is concerned with finding "nice" structures in arbitrary metric spaces. Following [19, 21], Mendel

[1]Ultrametric is a metric space satisfying a strong form of the triangle inequality: $d(x, z) \le \max\{d(x, y), d(y, z)\}$ (for all x, y, z). Ultrametrics embed isometrically into both Euclidean space [62], and tree metric. See Definition 3.1.
[2]Metric embedding $f : X \to Y$ is dominating if $\forall u, v \in X$, $d_X(u, v) \le d_Y(f(u), f(v))$.

and Naor [65] showed that for every integer parameter $k \geq 1$, every n-point metric (X, d) has a subset $M \subseteq X$ of size at least $n^{1-1/k}$ that embeds into a tree (ultrametric) with multiplicative distortion $O(k)$ (see [1, 25, 66] for improvements). In fact, the embedding has multiplicative distortion $O(k)$ for any pair in $M \times X$. We say that the vertices in M are *satisfied* (see Figure 1(b) for an illustration). As a corollary, every n-point metric space (X, d_X) admits a collection $\mathcal{T}$ of $k \cdot n^{1/k}$ dominating trees over X and a mapping $\mathbf{home} : X \to \mathcal{T}$, such that for every $x, y \in X$, it holds that $d_{\mathbf{home}(x)}(x, y) \leq O(k) \cdot d_X(x, y)$. These are called Ramsey trees, and they have found applications to online algorithms [19], approximate distance oracles [35, 65], and routing [1].

A new type of embedding: clan embedding. Recall that our initial goal was to embed a general metric space into a "simple" target space, specifically a tree metric. A drawback of both the stochastic embedding and the Ramsey-type embedding is that the embeddings are actually into a collection of trees rather than into a single one; thus the target space is not as simple as one might desire. Each embedding type makes a different type of compromise: the distortion guaranteed in stochastic embedding is only in expectation, while in the Ramsey-type embedding, only a subset of the vertices enjoys a bounded distortion guarantee. In this paper, we propose a novel type of compromise, which we call *clan embedding*. Here we will have a single embedding with a worst-case guarantee on all vertex pairs. The caveat is that each vertex might be mapped to multiple copies. This violates the classical paradigm of having a one-to-one relationship between the source and target spaces. However, we obtain a map into a single tree with a worst-case guarantee; this is beneficial and opens a new array of possibilities.

A *one-to-many* embedding $f : X \to 2^Y$ maps each point x into a subset $f(x) \subseteq Y$ called the *clan* of x. Each vertex $x' \in f(x)$ is called a *copy* of x (see Definition 3.2). Clan embedding is a pair (f, χ), where f is a one-to-many embedding, and $\chi : X \to Y$ maps each vertex x to a special vertex $\chi(x) \in f(x)$ called the *chief*. Clan embeddings are *dominating*, that is, for every $x, y \in X$, the distance between every two copies is at least the original distance: $\min_{x' \in f(x), y' \in f(y)} d_Y(x', y') \geq d_X(x, y)$. (f, χ) has multiplicative distortion t, if for every $x, y \in X$, some vertex in the clan of x is close to the chief of y: $\min_{x' \in f(x)} d_Y(x', \chi(y)) \leq t \cdot d_X(x, y)$ (see Definition 3.3). See Figure 1(c) for an illustration.

Clan embeddings into trees. One can easily construct an isometric clan embedding into a tree by allowing n copies for each vertex. On the other hand, with a single copy per vertex, the clan embedding becomes a classic embedding, which requires a multiplicative distortion of $\Omega(n)$. Our goal is to construct a low distortion clan embedding, while keeping the number of copies each vertex has as small as possible. To this end, we construct a distribution over clan embeddings, where all the embeddings in the support have a worst-case distortion guarantee; however, the expected number of copies each vertex has is bounded by a constant close to 1.

Theorem 1.1 (Clan embedding into ultrametric). *Given an n-point metric space (X, d_X) and parameter $\epsilon \in (0, 1]$, there is a uniform distribution $\mathcal{D}$ over $O(n \log n/\epsilon^2)$ clan embeddings (f, χ) into ulrametrics with multiplicative distortion $O(\frac{\log n}{\epsilon})$ such that for every point $x \in X$, $\mathbb{E}_{f \sim \mathcal{D}}[|f(x)|] \leq 1 + \epsilon$.*

In addition, for every $k \in \mathbb{N}$, there is a uniform distribution $\mathcal{D}$ over $O(n^{1+\frac{2}{k}} \log n)$ clan embeddings (f, χ) into ulrametrics with multiplicative distortion $16k$ such that for every point $x \in X$,

$$\mathbb{E}_{f \sim \mathcal{D}}[|f(x)|] = O(n^{\frac{1}{k}}).$$

We fist show that there exists a distribution $\mathcal{D}$ of clan embeddings that has distortion and expected clan size via the minimax theorem. We then use the multiplicative weights update (MWU) method to explicitly construct a uniform distribution $\mathcal{D}$ of polynomial support as specified by Theorem 1.1.

Our clan embedding into ultrametric is asymptotically tight (up to a constant factor in the distortion), and cannot be improved even if we embed into a general tree (rather than to the much more restricted structure of an ultrametric). Additionally, our lower bound implies that the ultra-sparse spanner construction of Elkin and Neiman [44] is asymptotically tight. (Elkin and Neiman [44] constructed a spanner with stretch $O(\frac{\log n}{\epsilon})$ and $(1 + \epsilon)n$ edges.)

Theorem 1.2 (Lower bound for clan embedding into a tree). *For every fixed $\epsilon \in (0, 1)$ and large enough n, there is an n-point metric space (X, d_X) such that for every clan embedding (f, χ) of X into a tree with multiplicative distortion $O(\frac{\log n}{\epsilon})$, it holds that $\sum_{x \in X} |f(x)| \geq (1 + \epsilon)n$.*
Furthermore, for every $k \in \mathbb{N}$, there is an n-point metric space (X, d_X) such that for every clan embedding (f, χ) of X into a tree with multiplicative distortion $O(k)$, it holds that $\sum_{x \in X} |f(x)| \geq \Omega(n^{1+\frac{1}{k}})$.

Often, we are given a weighted graph $G = (V, E, w)$, and the goal is to embed the shortest path metric of the graph d_G into a tree T. However, if, for example, one is required to construct a network while using only pre-existing edges from E, it is desirable that the tree T will be a subgraph of G, called a spanning tree. Abraham and Neiman [4] (improving over [43]) constructed a stochastic embedding of general graphs into spanning trees with expected distortion $O(\log n \log \log n)$ (losing a $\log \log n$ factor compared to general trees [46]). Later, Abraham *et al.* [1] constructed Ramsey spanning trees, showing that for every $k \in \mathbb{N}$, every graph can be embedded into a spanning tree with a subset M of at least $n^{1-\frac{1}{k}}$ satisfied vertices which suffers a distortion at most $O(k \log \log n)$ w.r.t. any other vertex (again losing a $\log \log n$ factor compared to general trees). Here we provide a "spanning" analog of Theorem 1.1. Similar to [1, 4], we also lose a $\log \log n$ factor compared to general trees. In particular, by Theorem 1.2, our spanning clan embedding is optimal up to second-order terms. As an application, we construct the first compact routing scheme with routing tables of constant size in expectation; see Section 1.1.1. We say that a clan embedding (f, χ) of a graph G into a graph H is *spanning* if $f(V(G)) = V(H)$ (i.e., every vertex in H is an image of a vertex in G) and for every edge $\{v', u'\} \in E(H)$ where $v' \in f(v), u' \in f(u)$, it holds that $\{v, u\} \in E(G)$ (see Definitions 3.2 and 3.3).

Theorem 1.3 (Spanning clan embedding into trees). *Given an n-vertex weighted graph $G = (V, E, w)$ and parameter $\epsilon \in (0, 1]$, there is a distribution $\mathcal{D}$ over spanning clan embeddings (f, χ) into trees with multiplicative distortion $O(\frac{\log n \log \log n}{\epsilon})$ such that for every vertex $v \in V$, $\mathbb{E}_{f \sim \mathcal{D}}[|f(v)|] \leq 1 + \epsilon$.*

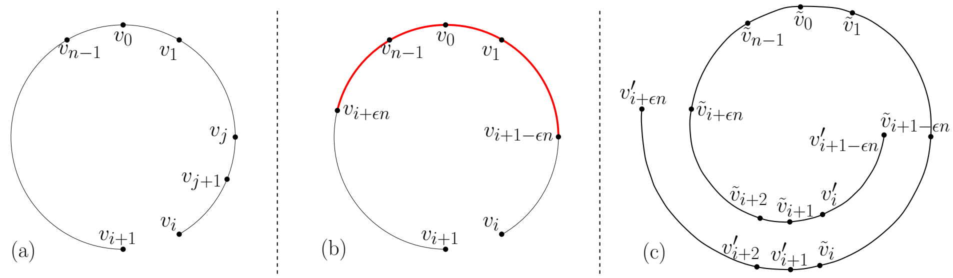


Figure 1: Three different types of embeddings of the cycle graph C_n into a tree.
(a) On the left illustrated a stochastic embedding that is created by deleting an edge $\{v_i, v_{i+1}\}$ uniformly at random. The expected multiplicative distortion of a pair of neighboring vertices v_j, v_{j+1} is $\mathbb{E}[d_T(v_j, v_{j+1})] = \frac{n-1}{n} \cdot 1 + \frac{1}{n} \cdot (n-1) = \frac{2n-2}{n} < 2$. By the triangle inequality and linearity of expectation, the expected multiplicative distortion is ≤ 2.
(b) In the middle illustrated a Ramsey type embedding: an arbitrary edge $\{v_i, v_{i+1}\}$ is deleted. The vertices in the subset M (on the thick red line), which constitutes a $(1-2\epsilon)$ fraction of the vertex set, are satisfied. That is, they suffer from a multiplicative distortion at most $\frac{1}{\epsilon}$ w.r.t. any other vertex.
(c) On the right illustrated a clan embedding, where i is chosen uniformly at random. The chief of a vertex v_j denoted $\tilde{v}_j$. Each vertex $v_j \in \{v_{i+1-\epsilon n}, \dots, v_{i+\epsilon n}\}$ has additional copy v'_j; thus the probability that a vertex has two copies is 2ϵ, implying that $\mathbb{E}[|f(v_a)|] = 1 + 2\epsilon$. The distortion is $\min\{d(\tilde{v}_a, \tilde{v}_b), d(v'_a, \tilde{v}_b)\} \le \frac{1}{\epsilon} \cdot d_{C_n}(v_a, v_b)$

In addition, for every $k \in \mathbb{N}$, there is a distribution $\mathcal{D}$ over spanning clan embeddings (f, χ) into trees with multiplicative distortion $O(k \log\log n)$, where for every vertex $v \in V$, $\mathbb{E}_{f\sim\mathcal{D}}[|f(v)|] = O(n^{\frac{1}{k}})$.

Clan embedding of minor-free graphs into bounded treewidth graphs. As [28] and [46] are tight, a natural question arises: by embedding from a simpler space (than general n-point metric space) into a richer space (than trees), could the distortion be reduced? The family of low-treewidth graphs is an excellent candidate for a target space: it is a much more expressive target space than trees, while many hard problems remain tractable. Unfortunately, by the work of Chakrabarti *et al.* [32] (see also [31]), there are n-vertex planar graphs such that every (stochastic) embedding into $o(\sqrt{n})$-treewidth graphs must incur expected multiplicative distortion $\Omega(\log n)$. Bypassing this roadblock, Fox-Epstein *et al.* [51] (improving over [42]), showed how to embed planar metrics into bounded treewidth graphs while incurring only a small *additive* distortion. Specifically, given a planar graph G and a parameter ϵ, they constructed a deterministic dominating embedding f into a graph H of treewidth $\mathrm{poly}(\frac{1}{\epsilon})$, such that $\forall u, v \in G$, $d_H(f(u), f(v)) \le d_G(u, v) + \epsilon D$, where D is the diameter of G. While ϵD looks like a crude additive bound, it suffices to obtain approximation schemes for several classic problems: k-center, vehicle routing, metric ρ-dominating set, and metric ρ-independent set.

Following the success in planar graphs, Cohen-Addad *et al.* [36] wanted to generalize to minor-free graphs. Unfortunately, they showed that already obtaining additive distortion $\frac{1}{20}D$ for K_6-minor-free graphs requires the host graph to have treewidth $\Omega(\sqrt{n})$. Inspired by the case of trees, [36] bypass this barrier by constructing a stochastic embedding from K_r-minor-free n-vertex graphs into a distribution over treewidth-$O_r(\frac{\log n}{\epsilon^2})$ graphs with expected additive distortion ϵD, [3] that is $\forall u, v \in G, \mathbb{E}_{(f,H)\sim\mathcal{D}}[d_H(f(u), f(v))] \le d_G(u, v) + \epsilon D$. Similar to the case in planar graphs, Cohen-Addad *et al.* [36] used their embedding to construct an approximation scheme for the capacitated vehicle routing problem in K_r-minor-free graphs. However, due to the stochastic nature of the embedding, it was not strong enough to imply any results for the metric ρ-dominating/independent problems in minor-free graphs, which, prior to our work, remain wide open.

In this paper, similar to the case of trees, we construct Ramsey-type and clan embedding analogs to the stochastic embedding of [36]. Our Ramsey-type embedding bypasses the lower bound of $\Omega(\sqrt{n})$ from [36] while guaranteeing a worst-case distortion (for a large random subset of vertices). As an application, we obtain a bicriteria quasi-polynomial time approximation scheme (QPTAS) [5] for the metric ρ-independent set problem in minor-free graphs (see Section 1.1.2).

Theorem 1.4 (Ramsey-type embedding of minor-free graphs). *Given an n-vertex K_r-minor-free graph $G = (V, E, w)$ with diameter D and parameters $\epsilon \in (0, \frac{1}{4})$, $\delta \in (0, 1)$, there is a distribution over dominating embeddings $g : G \to H$ into graphs of treewidth $O_r(\frac{\log^2 n}{\epsilon\delta})$, such that there is a subset $M \subseteq V$ of vertices for which the following claims hold:*

(1) For every $u \in V$, $\Pr[u \in M] \ge 1 - \delta$.
(2) For every $u \in M$ and $v \in V$,

$$d_H(g(u), g(v)) \le d_G(u, v) + \epsilon D.$$

By setting $\delta = \frac{1}{2}$ and repeating $\log n$ times, a straightforward corollary is the following.

Corollary 1.5. *Given a K_r-minor-free n-vertex graph G with diameter D and parameter $\epsilon \in (0, \frac{1}{4})$, there are $\log n$ dominating embeddings $g_1, \dots, g_{\log n}$ into graphs of treewidth $O_r(\frac{\log^2 n}{\epsilon})$, such that for every vertex v, there is some embedding g_{i_v}, such that*

$$\forall u \in V, \qquad d_{H_{i_v}}(g_{i_v}(u), g_{i_v}(v)) \le d_G(u, v) + \epsilon D\,.$$

[3] O_r hides some function depending only on r. That is, there is some function $\chi : \mathbb{N} \to \mathbb{N}$ such that $O_r(x) \le \chi(r) \cdot x$.

While Ramsey-type embedding is sufficient for the metric ρ-independent set problem (as we can restrict our search to independent sets in M), we cannot use it for the metric ρ-dominating set problem (as every good solution might contain vertices outside M). To resolve this issue, we construct a clan embedding of minor-free graphs into bounded treewidth graphs. As we have a worst-case distortion guarantee for all vertex pairs, we obtain a QPTAS [5] for the metric ρ-dominating set problem in minor-free graphs (see Section 1.1.2).

Theorem 1.6 (Clan embedding for minor-free graphs). *Given a K_r-minor-free n-vertex graph $G = (V, E, w)$ of diameter D and parameters $\epsilon \in (0, \frac{1}{4})$, $\delta \in (0, 1)$, there is a distribution $\mathcal{D}$ over clan embeddings (f, χ) with additive distortion ϵD into graphs of treewidth $O_r(\frac{\log^2 n}{\delta\epsilon})$ such that for every $v \in V$, $\mathbb{E}[|f(v)|] \le 1 + \delta$.*

1.1 Applications

1.1.1 Compact Routing Scheme. A *routing scheme* in a network is a mechanism that allows packets to be delivered from any node to any other node. The network is represented as a weighted undirected graph, and each node can forward incoming data by using local information stored at the node, called a *routing table*, and the (short) packet's *header*. The routing scheme has two main phases: in the preprocessing phase, each node is assigned a routing table and a short *label*; in the routing phase, when a node receives a packet, it should make a local decision, based on its own routing table and the packet's header (which may contain the label of the destination, or a part of it), of where to send the packet. The *stretch* of a routing scheme is the worst-case ratio between the length of a path on which a packet is routed to the shortest possible path.

Compact routing schemes were extensively studied [1, 10, 11, 34, 38, 41, 67, 72], starting with Peleg and Upfal [67]. Using $\tilde{O}(n^{\frac{1}{k}})$ table size, Awerbuch et al. [10] obtained stretch $O(k^2 9^k)$, which was improved later to $O(k^2)$ by Awerbuch and Peleg [11]. In their celebrated compact routing scheme, Thorup and Zwick [72] obtained stretch $4k - 5$ while using $O(k \cdot n^{1/k})$ size tables and labels of size $O(k \log n)$. [4] The stretch was improved to roughly $3.68k$ by Chechik [34], using a scheme similar to [72] (while keeping all other parameters intact). Recently, Abraham *et al.* [1] devise a compact routing scheme (using Ramsey spanning trees) with labels of size $O(\log n)$, tables of size $O(k \cdot n^{1/k})$, and stretch $O(k \log \log n)$.

In all previous works, the guarantees on the table size are worst case. That is, the table size of every node in the network is bounded by a certain parameter. Here our guarantee is only in expectation. Note that such an expected guarantee makes a lot of sense for a central planner constructing a routing scheme for a network where the goal is to minimize the total amount of resources rather than the maximal amount of resources in a single spot. Even though previous works analyzed worst-case guarantees, if one tries to analyze their expected bounds per vertex, the guarantees will not be improved. Our contribution is the following:

Theorem 1.7 (Compact routing scheme). *Given a weighted graph $G = (V, E, w)$ on n vertices and integer parameter $k > 1$, there is a compact routing scheme with stretch $O(k \log \log n)$ that has (worst-case) labels (and headers) of size $O(\log n)$, and the expected size of the routing table of each vertex is $O(n^{1/k})$.*

[4] Unless stated otherwise, we measure space in machine words, each word is $\Theta(\log n)$ bits.

Table 1: The table compares various routing schemes for n-vertex graphs. In rows 1-4, we compare different schemes in their full generality, here k is an integer parameter. In rows 5,6,8,10, we fix $k = \log n$, while in rows 7 and 9, we fix $k = \frac{\log n}{\log \log n}$. Note that our result in line 9 is superior to all previous results: it has reduced label size compared to lines 5-6, reduced table size compared to line 7, and reduced stretch compared to line 8. Our result in line 10 is the first to obtain a constant table size.
The sizes of the table and label are measured in words, each word is $O(\log n)$ bits. The header size is asymptotically equal to the label size in all the compared routing schemes. The main caveat is that, while in all previous results the table size is analyzed w.r.t. a worst-case guarantee, we only provide bounds in expectation (marked by (*)). The label size (as well as the stretch) is a worst-case guarantee in our work as well.

Routing s.		Stretch	Label	Table
1.	[72]	$4k-5$	$O(k\log n)$	$O(kn^{1/k})$
2.	[34]	$3.68k$	$O(k\log n)$	$O(kn^{1/k})$
3.	[1]	$O(k\log\log n)$	$O(\log n)$	$O(kn^{1/k})$
4.	Theorem 1.7	$O(k\log\log n)$	$O(\log n)$	$O(n^{1/k})^{(*)}$
5.	[72]	$O(\log n)$	$O(\log^2 n)$	$O(\log n)$
6.	[34]	$O(\log n)$	$O(\log^2 n)$	$O(\log n)$
7.	[1]	$O(\log n)$	$O(\log n)$	$O(\log^2 n)$
8.	[1]	$\widetilde{O}(\log n)$	$O(\log n)$	$O(\log n)$
9.	Theorem 1.7	$O(\log n)$	$O(\log n)$	$O(\log n)^{(*)}$
10.	Theorem 1.7	$\widetilde{O}(\log n)$	$O(\log n)$	$\boldsymbol{O(1)}^{(*)}$

See Table 1 for comparison of our and previous results. We mainly focus on the very compact regime where all the parameters are at most poly-logarithmic. A key result in [72] is a stretch 1 routing scheme for the special case of a tree, where a routing table has constant size, and logarithmic label size. All the previous works are based on constructing a collection of trees. Specifically, in [34, 72], there are n trees, where each vertex belongs to $O(\log n)$ trees, and for each pair of nodes, there is a tree that guarantees a small stretch. Routing is then done in that tree. This is the reason for their large label size of $\log^2 n$ (as a label consists of $\log n$ labels in different trees). [1] constructs $\log n$ (Ramsey spanning) trees in total, where each vertex v has a home tree T_v, such that v enjoys a small stretch w.r.t. any other vertex in T_v. The label then consists of the name of T_v and the label of v in T_v. However, the routing table is still somewhat large as one needs to store the routing information in $\log n$ different trees.

In contrast, our construction is based on the spanning clan embedding (f, χ) of Theorem 1.3 into a single tree T, where the clan of each vertex consists of $O(1)$ copies (in expectation). The label of each vertex v is simply the label of $\chi(v)$ in T. The routing table of v contains the routing tables of all the corresponding copies in $f(v)$.

1.1.2 Metric Baker Problems in Minor-free Graphs. Baker [13] introduced a "layering" technique in order to construct efficient polynomial approximation schemes (EPTAS) [5] for many "local" problems in planar graphs such as minimum-measure *dominating set* and maximum-measure *independent set*. The key observation is that planar graphs have the "bounded local treewidth" property. Baker showed that for some problems solvable on bounded treewidth graphs, one can construct efficient approximation schemes for graphs possessing the bounded local treewidth property. This approach was generalized by Demaine *et al.* [40] to minor-free graphs.

Eisenstat *et al.* [42] proposed metric generalizations of Baker problems: minimum measure ρ*-dominating set*, and maximum measure ρ*-independent set*. Given a metric space (X, d_X), a subset $S \subseteq X$ of points is a ρ-independent set if for every $x, y \in S$, $d_X(x, y) > \rho$. Similarly, a ρ-dominating set is a subset $S \subseteq X$ such that for every $x \in X$, there exists $y \in S$, such that $d_X(x, y) \le \rho$. Given a measure $\mu : X \to \mathbb{R}_+$, the goal of the metric ρ-dominating (resp. independent) set problem is to find a ρ-dominating (resp. independent) set of minimum (resp. maximum) measure. It is often the case that metric Baker problems are much easier under the uniform measure. Sometimes, in addition, we are given a set of terminals $\mathcal{K} \subseteq X$, and required only that the terminals will be dominated ($\forall x \in \mathcal{K}$, $\exists y \in S$ s.t. $d_X(x, y) \ge \rho$). Note that the metric generalization of Becker problems in structured graphs (e.g. planar) is considerably harder than the non-metric problems. This is because the graph describing dominance/independence relations no longer posses the original structure (e.g. planarity).

An approximation scheme for the ρ-dominating (resp. independent) set problem returns a ρ-dominating (resp. independent) set S such that for every ρ-dominating (resp. independent) set S' it holds that $\mu(S) \le (1+\epsilon)\mu(S')$ (resp. $\mu(S) \ge (1-\epsilon)\mu(S')$). A bicriteria approximation scheme for the ρ-dominating (resp. independent) set problem returns a $(1+\epsilon)\rho$-dominating (resp. $(1-\epsilon)\rho$-independent) set S such that for every ρ-dominating (resp. independent) set S' it holds that $\mu(S) \le (1+\epsilon)\mu(S')$ (resp. $\mu(S) \ge (1-\epsilon)\mu(S')$).

For unweighted graphs with treewidth tw, Borradaile and Le [26] provided an exact algorithm for the ρ-dominating set problem with $O((2\rho+1)^{\mathrm{tw}+1} n)$ running time (see also [39]). For general treewidth tw graphs, using dynamic programming technique, Katsikarelis *et al.* [57] designed a fixed parameter tractable (FPT) approximation algorithm for the metric ρ-dominating set problem with $(\mathrm{tw}/\epsilon)^{O(\mathrm{tw})} \cdot \mathrm{poly}(n)$ runtime that returns a $(1+\epsilon)\rho$-dominating set S, such that for every ρ-dominating set S' it holds that $\mu(S) \le \mu(S')$. A similar result was also obtained for the metric ρ-independent set problem [58]. In particular, for the very basic case of bounded treewidth graphs, no true approximation scheme (even with quasi-polynomial time) is known for these problems. Additional evidence was provided by Marx and Pilipczuk [64] (see also [51]), who showed that the existence of EPTAS [5] for either ρ-dominating/independent set problem in planar graphs would refute the exponential-time hypothesis (ETH). Given this evidence, it is natural to settle for bicriteria approximation.

For unweighted planar graphs and constant ρ, there are linear time approximation schemes (not bicriteria) for the metric ρ-independent/dominating set problems [39, 45]. In weighted planar graphs, under the uniform measure, Marx and Pilipczuk [64] gave exact $n^{O(\sqrt{k})}$ time solution to both metric ρ-dominating/isolated set problems, provided that the solution is guaranteed to be of size at most k. Using their embedding of planar graphs into $\epsilon^{-O(1)} \log n$-treewidth graphs with additive distortion ϵD, Eisenstat *et al.* [42] provided a bicriteria PTAS [5] for both metric ρ-independent set and ρ-dominating set problems in planar graphs. Later, by constructing an improved embedding into $\epsilon^{-O(1)}$-treewidth graphs, Fox-Epstein *et al.* [51] obtained a bicriteria EPTAS.[5]

Finally, we turn to the most challenging case of minor-free graphs. For the restricted uniform measure case, using local search (similarly to [37]), we construct PTAS for both metric ρ-dominating and ρ-independent set problems. (See the full version for details). However, the local search approach seems to be hopeless for general measures. Alternately, one can try the metric embedding approach (for which bicriteria approximation is unavoidable). Unfortunately, unlike the classic embeddings in [42, 51], Cohen-Addad *et al.* [36] provided a stochastic embedding with an *expected distortion* guarantee. Such a stochastic guarantee is not strong enough to construct approximation schemes for the metric ρ-independent/dominating set problems. Using our clan and Ramsey-type embeddings, we are able to provide the first bicriteria QPTAS [5] for these problems. See Table 2 for a summary of previous and current results.

Theorem 1.8 (Metric ρ-independent set). *There is a bicriteria quasi-polynomial time approximation scheme (QPTAS) for the metric ρ-independent set problem in K_r-minor-free graphs.*
Specifically, given a weighted n-vertex K_r-minor-free graph $G = (V, E, w)$, measure $\mu : V \to \mathbb{R}_+$ and parameters $\epsilon \in (0, \frac{1}{4})$, $\rho > 0$, in $2^{\tilde{O}_r(\frac{\log^2 n}{\epsilon^2})}$ time, one can find a $(1-\epsilon)\rho$-independent set $S \subseteq V$ such that for every ρ-independent set $\tilde{S}$, $\mu(S) \ge (1-\epsilon)\mu(\tilde{S})$.

Theorem 1.9 (Metric ρ-dominating set). *There is a bicriteria quasi-polynomial time approximation scheme (QPTAS) for the metric ρ-dominating set problem in K_r-minor-free graphs.*
Specifically, given a weighted n-vertex K_r-minor-free graph $G = (V, E, w)$, measure $\mu : V \to \mathbb{R}_+$, a subset of terminals $\mathcal{K} \subseteq V$, and parameters $\epsilon \in (0, \frac{1}{4})$, $\rho > 0$, in $2^{\tilde{O}_r(\frac{\log^2 n}{\epsilon^2})}$ time, one can find a $(1+\epsilon)\rho$-dominating set $S \subseteq V$ for $\mathcal{K}$ such that for every ρ-dominating set $\tilde{S}$ of $\mathcal{K}$, $\mu(S) \le (1+\epsilon)\mu(\tilde{S})$.

1.2 Related Work

Path-distortion A closely related notion to clan embeddings is multi-embedding studied by Bartal and Mendel [23]. A multi-embedding is a dominating one-to-many embedding. The distortion guarantee, however, is very different. We say that a multi-embedding $f : X \to 2^Y$ between metric spaces (X, d_X), (Y, d_Y) has *path distortion* t, if for every "path" in X, i.e., a sequence of points $x_0, x_1, \dots, x_q$, there are copies $x'_i \in f(x_i)$ such that $\sum_{i=0}^{q-1} d_Y(x'_i, x'_{i+1}) \le t \cdot \sum_{i=0}^{q-1} d_X(x_i, x_{i+1})$. For n point metric space (X, d) with aspect ratio

[5] A polynomial time approximation scheme (PTAS) is an algorithm that for any fixed $\epsilon \in (0, 1)$, provides a $(1+\epsilon)$-approximation in polynomial time. A PTAS is an *efficient* polynomial time approximation scheme (EPTAS) if running time is of the form $n^{O(1)} \cdot f(\epsilon)$ for some function $f(.)$ depending on ϵ only. A quasi-polynomial time approximation scheme (QPTAS) has running time $2^{\mathrm{polylog}(n)}$ for every fixed ϵ.

Table 2: The table compares different approximation schemes for metric Becker problems on weighted graphs. All compared results apply to both metric ρ-dominating/independent set problems. All the results (other than in line 5) apply to the general measure case.

Reference		Family	Result	Technique
1.	[64]	planar	No EPTAS under ETH	
2.	[57, 58]	treewidth	FPT with approx $(1+\epsilon)\rho$	Dynamic programming
3.	[42]	planar	Bicriteria PTAS	Deterministic embedding
4.	[51]	planar	Bicriteria EPTAS	Deterministic embedding
5.	Theorems 16&17 in the full version	minor-free	PTAS (uniform measure)	Local search
6.	Theorems 1.8&1.9	minor-free	Bicriteria QPTAS	Clan/Ramsey type embedding

Φ [6], and parameter $k \geq 1$, Bartal and Mendel [23] constructed a multi-embedding into ultrametric with $O(n^{1+\frac{1}{k}})$ vertices and distortion $O(k \cdot \min\{\log n \cdot \log\log n, \log\Phi \cdot \log\log\Phi\})$. Formally, path distortion and multiplicative distortion of clan embedding are incomparable, as clan embedding guarantees small distortion with respect to a single chief vertex (which is crucial to our applications), while the multi-embedding [23] distortion guarantee is w.r.t. arbitrary copies, but preserve entire "paths". Interestingly, a small modification to our clan embedding provides the path distortion guarantee as well! See the full version for details [49]. Specifically, we obtain embedding into ultrametric with $O(n^{1+\frac{1}{k}})$ (resp. $(1+\epsilon)n$) vertices and distortion $O(k \cdot \min\{\log n, \log\Phi\})$ (resp. $O(\frac{\log n}{\epsilon} \cdot \min\{\log n, \log\Phi\})$), shaving a $\log\log$ factor compared with [23]. In a private communication, Bartal told us that he obtained the exact same path distortion guarantees more than a decade ago; Bartal's manuscript is made public recently [18].

In a concurrent paper, Haeupler *et al.* [53] studied a closely related notion of tree embeddings with copies. They construct a one-to-many embedding of a graph G into a tree T where every vertex has at most $O(\log n)$ copies, and such that every connect subgraph H of G has a connected copy H' in T, of weight at most $O(\log^2 n) \cdot w(H)$. Using the path distortion gurantee in our embedding (or [18]), one will obtain an embedding such that every connect subgraph H of G has a connected copy H' in T, of weight at most $O(\log n) \cdot w(H)$, however the bound on the maximal number of copies will be only polynomial.

Tree covers. The constructions of Ramsey trees are asymptotically tight [19]. Furthermore, as was shown by Bartal *et al.* [20] that they cannot be substantially improved even for planar graphs with a constant doubling dimension. [7] Therefore [20] suggested studying a weaker gurantee provided by tree covers. Here the goal is to construct a small collection of dominating embeddings into trees such that every pair of vertices has a small distortion in some tree in the collection. For n-vertex minor-free graph [20] constructed $1+\epsilon$ tree covers of size $O_r(\frac{\log^2 n}{\epsilon^2})$ (or a $O(1)$-tree cover $O(1)$ size). For metrics with doubling dimension d, [20] constructed $1+\epsilon$-tree covers of size $(\frac{1}{\epsilon})^{O(d)}$. Recently, the authors [50] showed that for doubling metrics, we can replace the trees by ultrametrics.

[6]The aspect ratio of a metric space (X,d) is the ratio between the maximal and minimal distances $\frac{\max_{x,y} d(x,y)}{\min_{x\neq y} d(x,y)}$.

[7]Specifically, for every $\alpha > 0$, [20] constructed planar graph with constant doubling dimension, such that for every tree embedding, the subset of vertices enjoying distortion $\leq \alpha$ is of size at most $n^{1-\Omega(\frac{1}{\alpha\log\alpha})}$, which is almost as bad as general graphs.

Minor free graphs. Different types of embedding were studied for minor-free graphs. K_r-minor-free graphs embed into ℓ_p space with multiplicative distortion $O_r(\log^{\min\{\frac{1}{2},\frac{1}{p}\}} n)$ [2, 3, 60, 69]. In particular, they embed into ℓ_∞ of dimension $O_r(\log^2 n)$ with a constant multiplicative distortion. They also admit spanners with multiplicative distortion $1+\epsilon$ and $\tilde{O}_r(\epsilon^{-3})$ lightness [27]. On the other hand, there are other graph families that embed well into bounded treewidth graphs. Talwar [71] showed that graphs with doubling dimension d and aspect ratio Φ [6], stochastically embed into graphs with treewidth $\epsilon^{-O(d\log d)} \cdot \log^d \Phi$ with expected distortion $1+\epsilon$. Similar embeddings are known for graphs with highway dimension h [47] (into treewidth $(\log\Phi)^{-O(\log^2 \frac{h}{\epsilon})}$ graphs), and graphs with correlation dimension k [33] (into treewidth $\tilde{O}_{k,\epsilon}(\sqrt{n})$ graphs).

1.3 Organization

In Section 3, we review basic notation in this paper. The construction of clan embedding into ultrametric is given in Section 4, and the lower bound for clan embeddings into trees is given in Section 5. Missing details of claimed results can be found in the full version of the paper [49].

2 PAPER OVERVIEW

Clan embedding into ultrametric. The main task is to prove a "distributional" version of Theorem 1.1. Specifically, given a parameter k, and a measure $\mu : X \to \mathbb{R}_{\geq 1}$, we construct a clan embedding with distortion $16k$ such that $\sum_{x\in X} \mu(x) \cdot |f(x)| \leq \mu(X)^{1+\frac{1}{k}}$, where $\mu(X) = \sum_{x\in X} \mu(x)$ (Lemma 4.1). We show that the distributioal version implies Theorem 1.1 by using the minimax theorem.

The algorithm to construct the distributional version is a deterministic recursive ball growing algorithm, which is somewhat similar to previous deterministic algorithms constructing Ramsey trees [1, 17]. Let D be the diameter of the metric space. We grow a ball $B(v,R)$ around a point v and partition the space into two clusters: the interior $B(v, R+\frac{D}{16k})$ and exterior $X \setminus B(v, R-\frac{D}{16k})$ of the ball, while points at distance $\frac{D}{16k}$ from the boundary of the ball belong to both clusters. We then recursively create a clan embedding into ultrametrics for each of the two clusters. These two embeddings are later combined into a single ultrametric where the root has label D. The $16k$ distortion guarantee follows from the wide "belt" around the boundary of the ball belonging to both clusters. Note that the images of vertices in this "belt" contain copies in the clan embeddings of both clusters, while "non-belt" points have copies in a single embedding only. However, the two clusters have cardinality smaller than $|X|$. The key is to carve the partition while guaranteeing that the relative measure of points belonging to both clusters will be small compared to the reduction in cardinality.

Spanning clan embedding into trees. In Theorem 1.3, the spanning version, we try to imitate the approach of Theorem 1.1. However, we cannot simply carve balls and continue recursively. The reason is that the diameter of a cluster could grow unboundedly after deleting some vertices. In particular, there is no clear upper bound on the distance between separated points.

To imitate the ball growing approach nonetheless, we use the petal-decomposition framework that was previously applied to create stochastic embedding into spanning trees [4], and Ramsey spanning trees [1]. The petal decomposition framework enables one to iteratively construct a spanning tree for a given graph. In each level, the current cluster is partitioned into smaller diameter pieces (called *petals*), which have properties resembling balls. The algorithm continues recursively on the petals. Later, the petals are connected back to create a spanning tree. The key property is that while creating a petal, we have a certain degree of freedom to chose its "radius", which enables us to use the ball growing approach from above. Crucially, the framework guarantees that for every choice of radii (within the sepecified limits), the diameter of the resulting tree will be only constant times larger than that of the original graph. However, the petal decomposition framework does not provide us with the freedom to choose the center of the petal. This makes the task of controlling the number of copies more subtle.

Lower bound for clan embedding into a tree. We provide here a proof sketch for the first assertion in Theorem 1.2. We begin by constructing an n-vertex graph $G=(V,E)$ with $(1+\epsilon)n$ edges and girth $g=\Omega(\frac{\log n}{\epsilon})$; the girth is the length of the shortest cycle. Consider an arbitrary clan embedding of G into a tree T with distortion $\frac{g}{c}=O(\frac{\log n}{\epsilon})$ (for some constant c) and κ copies overall. We create a new graph H by merging all the copies of each vertex into a single vertex. There is a naturally defined classic embedding from G to H with distortion $\le\frac{g}{c}$. The Euler characteristic of the graph G equals $\chi(G)=|E|-|V|+1=\epsilon n+1$, while the Euler characteristic of H is at most $\chi(H)\le\kappa-n$. However, Rabinovich and Raz [68] showed that, if an embedding from a girth-g graph G has distortion $\le\frac{g}{c}$, the host graph must have the Euler characteristic at least as large as that of G. Thus, we conclude that $\kappa\ge(1+\epsilon)n+1$ as required.

Ramsey type embedding for minor-free graphs. The structure theorem of Robertson and Seymour [70] stated that every minor-free graph can be decomposed into a collection of graphs embedded on the surface of constant genus (with some vortices and apices), glued together into a tree structure by taking clique-sums. The stochastic embedding of minor free graphs into a distribution over bounded treewidth graphs by Cohen-Addad *et al.* [36] was constructed according to the layers of the structure theorem. First, they constructed an embedding for a planar graph with a single vortex. Then, they generalized it to planar graphs with multiple vortices, subsequently to graphs embedded on the surface of constant genus with multiple vortices, and to surface embeddable graphs with multiple vortices and apices. Finally, they incorporated cliques-sums and generalized to minor-free graphs. Most crucially, for this paper, the only step requiring randomness was the incorporation of apices. Specifically, [36] constructed a deterministic embedding for graphs embedded on the surface of constant genus with multiple vortices. This is the starting point of our embeddings.

Our first step is to incorporate apices, however, instead of guaranteeing that the distance of each pair is distorted by ϵD in expectation, we will show that each vertex with probability $1-\delta$ enjoys a small distortion w.r.t. any other vertex. We begin by deleting all the apices Ψ and obtaining a surface embeddable graph with multiple vortices $G'=G[V\setminus\Psi]$. However, the diameter of the resulting graph is essentially unbounded. Pick an arbitrary vertex r, and partition G' into layers of width $O(\frac{D}{\delta})$ w.r.t. distances from r with a random shift [8]. It follows that every vertex v is $2D$-padded (that is, the ball $B(v,2D)$ is fully contained in a single layer) with probability $1-\delta$. The set M of *satisfied* vertices defined to be the set of all D-padded vertices. We then use the deterministic embedding from [36] on every layer with distortion parameter $\epsilon'=\Theta(\epsilon\delta)$ to incur additive distortion ϵD. Finally, we combine all these embeddings together into a single embedding, which also contains the apices.

The next step is to incorporate clique-sums. This is done recursively w.r.t. the clique-sum decomposition tree $\mathbb{T}$. In each step, we pick a central piece $\tilde{G}\in\mathbb{T}$ such that $\mathbb{T}\setminus\tilde{G}$ breaks into connected components $\mathbb{T}_1,\mathbb{T}_2,\dots$, where each $\mathbb{T}_i$ contains at most $|\mathbb{T}|/2$ pieces. We construct a Ramsey-type embedding for $\tilde{G}$ using the lemma above and obtain a set $\tilde{M}$ of satisfied vertices. Recursively, we construct a Ramsey-type embedding for each $\mathbb{T}_i$ and obtain a set M_i of satisfied vertices. We ensure that all these embeddings are clique-preserving. Thus even though eventually we will obtain a one-to-one embedding, during the process, we keep them *one-to-many and clique-preserving*. This provides us with a natural way to combine all the embeddings of $\tilde{G},\mathbb{T}_1,\mathbb{T}_2,\dots$ into a single embedding into a graph of bounded treewidth (by identifying vertices of respective clique copies). All the vertices in $\tilde{M}$ will be satisfied. A vertex $v\in\mathbb{T}_i$ will be satisfied if $v\in M_i$ and all the vertices in the clique Q_i, used in the clique sum of $\tilde{G}$ with $\mathbb{T}_i$, are satisfied $Q_i\subseteq\tilde{M}$. Analyzing the entire process, we show that each vertex is satisfied with probability at least $(1-\delta)^{\log n}$. The theorem follows by setting the parameter $\delta'=\Theta(\frac{\delta}{\log n})$.

Clan embedding for minor-free graphs. The construction here follows similar lines to our Ramsey-type embedding. However, we cannot simply "give-up" on vertices, as we required to provide a worst-case distortion guarantee on all vertex pairs. Similarly to the Ramsey-type case, we build on the deterministic embedding of surface embeddable graphs with vortices from [36], and generalize it to a clan embedding of graphs including the apices. However, there is one crucial difference in creating the "layering" (with the random shift). In the Ramsey-type embedding, vertices near the boundary between two layers simply failed and did not join M. Here, instead, the layers will somewhat overlap such that copies of vertices near boundary areas will be split into two unrelated

[8] Alternatively, one could use here a strong padded decomposition [48] (as in [36]) into clusters of diameter $O_r(\frac{D}{\delta})$ such that each radius-D ball is fully contained in a single cluster with probability $1-\delta$. However, this approach will not work for our clan embedding, as there is no bound on the number of copies we will need for failed vertices. We use the layering approach for the Theorem 1.4 as well to keep the proofs of Theorems 1.4 and 1.6 similar.

sets. In particular, cliques that lie near boundary areas will have two separated clique copies w.r.t. each corresponding layer (at most two). Even though that actually each vertex will have an essentially unbounded number of copies (due to the clique-preservation requirement), the copies of each vertex will be divided to either one or two sets, such that in the final embedding, it will be enough to pick an arbitrary single copy from each set. The copies of a vertex will split into two sets only if it is in the area of the boundary, the probability of which is bounded by δ.

The generalization to clique-sums also follows similar lines to the Ramsey-type embedding. We create a clan embedding for $\tilde{G}$ into treewidth graph $\tilde{H}$ as above, and recursively clan embeddings $H_1, H_2, \dots$ for $\mathbb{T}_1, \mathbb{T}_2, \dots$. For each $\mathbb{T}_i$, we will make the vertices of the clique Q_i, used for the clique-sum between $\tilde{G}$ and $\mathbb{T}_i$, into apices, thereby ensuring that H_i will succeed on Q_i. In particular, every vertex $v \in Q_i$ will have a single copy in H_i. When combining H_i with $\tilde{H}$, there are two cases. If the embedding $\tilde{H}$ was successful w.r.t. Q_i we will simply identify between the two clique copies and done. Otherwise, $\tilde{H}$ will contain two vertex-disjoint clique copies $\tilde{Q}_i^1, \tilde{Q}_i^2$ of Q_i. We will create two disjoint copies of the embedding H_i: H_i^1, H_i^2, and identify the two copies of Q_i in H_i^1, H_i^2 with $\tilde{Q}_i^1, \tilde{Q}_i^2$, respectively. It follows that for a vertex $v \in \mathbb{T}_i$, with probability at least $1-\delta$, the number of copies it will have is the same as in H_i, while with probability at most δ it will be doubled. Analyzing the entire process (and picking a single copy from each relevant set as above), we show that each vertex is expected to have at most $(1+\delta)^{\log n}$ copies. The theorem follows by using the parameter $\delta' = \Theta(\frac{\delta}{\log n})$.

3 PRELIMINARIES

$\tilde{O}$ notation hides poly-logarithmic factors, that is $\tilde{O}(g) = O(g) \cdot \mathrm{polylog}(g)$, while O_r notation hides factors in r, e.g. $O_r(m) = O(m) \cdot f(r)$ for some function f of r. All logarithms are at base 2 (unless specified otherwise).

We consider connected undirected graphs $G = (V, E)$ with edge weights $w_G : E \to \mathbb{R}_{\geq 0}$. A graph is called unweighted if all its edges have unit weight. Additionally, we denote G's vertex set and edge set by $V(G)$ and $E(G)$, respectively. Often, we will abuse notation and write G instead of $V(G)$. d_G denotes the shortest path metric in G, i.e., $d_G(u,v)$ is the shortest distance between u to v in G. Note that every metric space can be represented as the shortest path metric of a weighted complete graph. We will use the notions of metric spaces, and weighted graphs interchangeably. When the graph is clear from the context, we might use w to refer to w_G, and d to refer to d_G. $G[S]$ denotes the induced subgraph by S. The diameter of S, denoted by $\mathrm{diam}(S)$, is $\max_{u,v\in S} d_{G[S]}(u,v)$. [9]

An ultrametric (X, d) is a metric space satisfying a strong form of the triangle inequality, that is, for all $x, y, z \in X$, $d(x,z) \leq \max\{d(x,y), d(y,z)\}$. The following definition is known to be an equivalent one (see [22]).

Definition 3.1. An ultrametric is a metric space (X, d) whose elements are the leaves of a rooted labeled tree T. Each $z \in T$ is associated with a label $\ell(z) \geq 0$ such that if $x \in T$ is a descendant of z then $\ell(x) \leq \ell(z)$ and $\ell(x) = 0$ iff x is a leaf. The distance between leaves $x, y \in X$ is defined as $d_T(x,y) = \ell(\mathrm{lca}(x,y))$ where $\mathrm{lca}(x,y)$ is the least common ancestor of x and y in T.

[9] This is often called *strong* diameter. A related notion is the *weak* diameter of a cluster S, defined to be $\max_{u,v\in S} d_G(u,v)$. Note that for a metric space, weak and strong diameters are equivalent.

Metric Embeddings. Classically, a metric embedding is defined as a function $f : X \to Y$ between the points of two metric spaces (X, d_X) and (Y, d_Y). A metric embedding f is said to be *dominating* if for every pair of points $x, y \in X$, it holds that $d_X(x,y) \leq d_Y(f(x), f(y))$. The distortion of a dominating embedding f is $\max_{x\neq y\in X} \frac{d_Y(f(x),f(y))}{d_X(x,y)}$. Here we will study a more permitting generalization of metric embedding introduced by Cohen-Addad *et al.* [36], which is called *one-to-many* embedding .

Definition 3.2 (One-to-many embedding). A *one-to-many embedding* is a function $f : X \to 2^Y$ from the points of a metric space (X, d_X) into non-empty subsets of points of a metric space (Y, d_Y), where the subsets $\{f(x)\}_{x\in X}$ are disjoint. $f^{-1}(x')$ denotes the unique point $x \in X$ such that $x' \in f(x)$. If no such point exists, $f^{-1}(x') = \emptyset$. A point $x' \in f(x)$ is called a *copy* of x, while $f(x)$ is called the *clan* of x. For a subset $A \subseteq X$ of vertices, denote $f(A) = \cup_{x\in A} f(x)$.

We say that f is *dominating* if for every pair of points $x, y \in X$, it holds that $d_X(x,y) \leq \min_{x'\in f(x), y'\in f(y)} d_Y(x', y')$. We say that f has multiplicative distortion t, if it is dominating and $\forall x, y \in X$, it holds that $\max_{x'\in f(x), y'\in f(y)} d_Y(x', y') \leq t \cdot d_X(x,y)$. Similarly, f has additive distortion ϵD if f is dominating and $\forall x, y \in X$, $\max_{x'\in f(x), y'\in f(y)} d_Y(x', y') \leq d_X(x,y) + \epsilon D$.

A stochastic one-to-many embedding is a distribution $\mathcal{D}$ over dominating one-to-many embeddings. We say that a stochastic one-to-many embedding has expected multiplicative distortion t if $\forall x, y \in X$, $\mathbb{E}[\max_{x'\in f(x), y'\in f(y)} d_Y(x', y')] \leq t \cdot d_X(u,v)$. Similarly, f has expected additive distortion ϵD, if $\forall x, y \in X$, $\mathbb{E}[\max_{x'\in f(x), y'\in f(y)} d_Y(x', y')] \leq d_X(x,y) + \epsilon D$.

For a one-to-many embedding f between weighted graphs $G = (V, E, w)$ and $H = (V', E', w')$, we say that f is spanning if $V' = f(V)$ (i.e. f is "onto"), and for every edge $(u,v) \in E'$, it holds that $\left(f^{-1}(u), f^{-1}(v)\right) \in E$ and $w'(u,v) = w\left(f^{-1}(u), f^{-1}(v)\right)$.

This paper is mainly devoted to the new notion of clan embeddings.

Definition 3.3 (Clan embedding). A clan embedding from metric space (X, d_X) into a metric space (Y, d_Y) is a pair (f, χ) where $f : X \to 2^Y$ is a dominating one-to-many embedding, and $\chi : X \to Y$ is a classic embedding. For every $x \in X$, we have that $\chi(x) \in f(x)$; here $f(x)$ called the clan of x, while $\chi(x)$ is referred to as the chief of the clan of x (or simply the chief of x).

We say that clan embedding f has multiplicative distortion t if for every $x, y \in X$, $\min_{y'\in f(y)} d_Y(y', \chi(x)) \leq t \cdot d_X(x,y)$. Similarly, f has additive distortion ϵD if for every $x, y \in X$, $\min_{y'\in f(y)} d_Y(y', \chi(x)) \leq d_X(x,y) + \epsilon D$.

A clan embedding (f, χ) is said to be spanning if f is a spanning one-to-many embedding.

4 CLAN EMBEDDING INTO AN ULTRAMETRIC

This section is devoted to proving Theorem 1.1, restated bellow.

Theorem 1.1 (Clan embedding into ultrametric). *Given an n-point metric space (X, d_X) and parameter $\epsilon \in (0,1]$, there is a uniform distribution $\mathcal{D}$ over $O(n\log n/\epsilon^2)$ clan embeddings (f,χ) into ulrametrics with multiplicative distortion $O(\frac{\log n}{\epsilon})$ such that for every point $x \in X$, $\mathbb{E}_{f\sim\mathcal{D}}[|f(x)|] \le 1+\epsilon$.*

In addition, for every $k \in \mathbb{N}$, there is a uniform distribution $\mathcal{D}$ over $O(n^{1+\frac{2}{k}}\log n)$ clan embeddings (f,χ) into ulrametrics with multiplicative distortion $16k$ such that for every point $x \in X$,

$$\mathbb{E}_{f\sim\mathcal{D}}[|f(x)|] = O(n^{\frac{1}{k}}).$$

First, we will prove a "distributional" version of Theorem 1.1. That is, we will receive a distribution μ over the points, and deterministically construct a single clan embedding (f,χ) such that $\sum_{x\in X}\mu(x)|f(x)|$ will be bounded. Later, we will use the minimax theorem to conclude Theorem 1.1. We begin with some definitions: a *measure* over a finite set X, is simply a function $\mu : X \to \mathbb{R}_{\ge 0}$. The measure of a subset $A \subseteq X$, is $\mu(A) = \sum_{x\in A}\mu(x)$. Given some function $f : X \to \mathbb{R}$, it's expectation w.r.t. μ is $\mathbb{E}_{x\sim\mu}[f] = \sum_{x\in X}\mu(x)\cdot f(x)$. We say that μ is a *probability measure* if $\mu(X) = 1$. We say that μ is a (≥ 1)*-measure* if for every $x \in X$, $\mu(x) \ge 1$.

Lemma 4.1. *Given an n-point metric space (X, d_X), (≥ 1)-measure $\mu : X \to \mathbb{R}_{\ge 1}$, and integer parameter $k \ge 1$, there is a clan embedding (f,χ) into an ultrametric with multiplicative distortion $16k$ such that $\mathbb{E}_{x\sim\mu}[|f(x)|] \le \mu(X)^{1+\frac{1}{k}}$.*

Proof. Our proof is inspired by Bartal's lecture notes [17], who provided a deterministic construction of Ramsey trees. Specifically, Claim 1 bellow is due to [17]. Lemma 4.1 could also be proved using the techniques of Abraham *et al.* [1]; however the proof based on [17] we present here is shorter. For a subset $A \subseteq X$, denote by $B_A(x,r) := B_X(x,r)\cap A$ the ball in the metric space (X,d_X) restricted to A. Set $\mu^*(A) := \max_{x\in A}\mu\left(B_A(x,\frac{\mathrm{diam}(A)}{4})\right)$. Note that μ^* is monotone: i.e. $A' \subseteq A$ implies $\mu^*(A') \le \mu^*(A)$, and $\forall A$, $\mu^*(A) \le \mu(A)$. The following claim is crucial for our construction; the proof is deferred to the full version [49]. See Figure 2 for an illustration of the claim.

Claim 1. *There is a point $v \in X$ and radius $R \in (0, \frac{\mathrm{diam}(X)}{2}]$, such that the sets $P = B_X(v, R + \frac{1}{8k}\cdot\mathrm{diam}(X))$, $Q = B_X(v,R)$, and $\bar{Q} = X \setminus Q$ satisfy $\mu(P) \le \mu(Q)\cdot\left(\frac{\mu^*(X)}{\mu^*(P)}\right)^{\frac{1}{k}}$.*

The construction of the embedding is by induction on n, the number of points in the metric space. We assume that for a metric space X with strictly less than n points, and arbitrary (≥ 1)-measure μ, we can construct a clan embedding (f,χ) with distortion $16k$, such that $\mathbb{E}_{x\sim\mu}[|f(x)|] \le \mu(X)\mu^*(X)^{\frac{1}{k}} \le \mu(X)^{1+\frac{1}{k}}$. Find sets $P, Q, \bar{Q} \subseteq X$ using Claim 1. Let μ_P (resp. $\mu_{\bar{Q}}$) be the (≥ 1)-measure μ restricted to P (resp. $\bar{Q}$). Using the induction hypothesis, construct clan embeddings (f_P,χ_P) for P, and $(f_{\bar{Q}},\chi_{\bar{Q}})$ for $\bar{Q}$ into ultra-metrics $U_P, U_{\bar{Q}}$ respectively. Construct a new ultrametric U by combining U_P and $U_{\bar{Q}}$ by adding a new root node r_U with label $\mathrm{diam}(X)$ and making roots of U_P and $U_{\bar{Q}}$ children of r_U. For every $x \in X$ set $f(x) = f_P(x)\cup f_{\bar{Q}}(x)$. If $d_X(v,x) \le R + \frac{1}{16k}\cdot\mathrm{diam}(X)$ set $\chi(x) = \chi_P(x)$, otherwise set $\chi(x) = \chi_{\bar{Q}}(x)$. This finishes the construction; see Figure 2 for an illustration.

Next, we argue that the clan embedding (f,χ) has multiplicative distortion $16k$. Consider a pair of points $x, y \in X$. We will show that $\min_{y'\in f(y)} d_U(y',\chi(x)) \le 16k\cdot d_X(x,y)$. Suppose first that $d_X(v,x) \le R + \frac{1}{16k}\cdot\mathrm{diam}(X)$. If $y \in P$, then by the induction hypothesis

$$\begin{aligned}\min_{y'\in f(y)} d_U(y',\chi(x)) &\le \min_{y'\in f_P(y)} d_{U_P}(y',\chi_P(x))\\ &\le 16k\cdot d_P(x,y) = 16k\cdot d_X(x,y)\ .\end{aligned}$$

Else, $y \notin P$, then $d_X(v,y) > R + \frac{1}{8k}\cdot\mathrm{diam}(X)$. Using the triangle inequality $d_X(x,y) \ge d_X(v,y) - d_X(v,x) \ge \frac{\mathrm{diam}(X)}{16}$. Note that the label of r_U is $\mathrm{diam}(X)$, implying that $\min_{y'\in f(y)} d_U(y',\chi(x)) \le \mathrm{diam}(X) \le 16\cdot d_X(x,y)$. The case where $d_X(v,x) > R + \frac{1}{16k}\cdot\mathrm{diam}(X)$ is symmetric (using $\bar{Q}$ instead of P).

Next, we bound the weighted number of leafs in the ultrametric. Note that the process is deterministic and there is no probability involved. Using the induction hypothesis, it holds that

$$\begin{aligned}\mathbb{E}_{x\sim\mu}[|f(x)|] &= \sum_{x\in X}\mu(x)\cdot\left(|f_P(x)| + |f_{\bar{Q}}(x)|\right)\\ &= \mathbb{E}_{x\sim\mu_P}[|f_P(x)|] + \mathbb{E}_{x\sim\mu_{\bar{Q}}}[|f_{\bar{Q}}(x)|]\\ &\le \mu_P(P)\mu_P^*(P)^{\frac{1}{k}} + \mu_{\bar{Q}}(\bar{Q})\mu_{\bar{Q}}^*(\bar{Q})^{\frac{1}{k}}\\ &\le \mu(P)\mu^*(P)^{\frac{1}{k}} + \mu(\bar{Q})\mu^*(\bar{Q})^{\frac{1}{k}}\\ &\overset{(*)}{\le} \mu(Q)\mu^*(X)^{\frac{1}{k}} + \mu(\bar{Q})\mu^*(X)^{\frac{1}{k}} = \mu(X)\mu^*(X)^{\frac{1}{k}}\ ,\end{aligned}$$

where in the inequality $(*)$ is due to Claim 1 and the fact that $\mu^*(\bar{Q}) \le \mu^*(X)$. □

Next, we translate the language of (≥ 1)-measures used in Lemma 4.1 to probability measures:

Lemma 4.2. *Given an n-point metric space (X, d_X), and probability measure $\mu : X \to \mathbb{R}_{\ge 0}$, we can construct the two following clan embeddings (f,χ) into ultrametrics:*

1. *For every parameter $k \ge 1$, multiplicative distortion $16k$ such that $\mathbb{E}_{x\sim\mu}[|f(x)|] \le O(n^{\frac{1}{k}})$.*
2. *For every $\epsilon \in (0,1]$, multiplicative distortion $O(\frac{\log n}{\epsilon})$ such that*
$$\mathbb{E}_{x\sim\mu}[|f(x)|] \le 1+\epsilon.$$

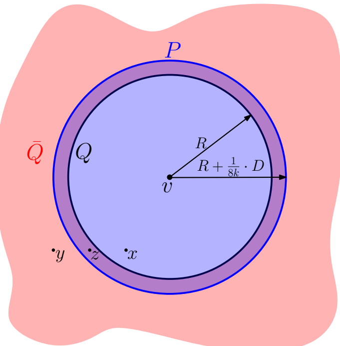
Proof. We define the following probability measure $\widetilde{\mu}$: $\forall x \in X$, $\widetilde{\mu}(x) = \frac{1}{2n} + \frac{1}{2}\mu(x)$. Set the following (≥ 1)-measure $\widetilde{\mu}_{\ge 1}(x) = 2n\cdot\tilde{\mu}(x)$. Note that $\widetilde{\mu}_{\ge 1}(X) = 2n$. We execute Lemma 4.1 w.r.t. the (≥ 1)-measure $\widetilde{\mu}_{\ge 1}$, and parameter $\frac{1}{\delta} \in \mathbb{N}$ to be determined later. It holds that

$$\begin{aligned}\widetilde{\mu}_{\ge 1}(X)\cdot\mathbb{E}_{x\sim\widetilde{\mu}}[|f(x)|] &= \mathbb{E}_{x\sim\widetilde{\mu}_{\ge 1}}[|f(x)|]\\ &\le \widetilde{\mu}_{\ge 1}(X)^{1+\delta} = \widetilde{\mu}_{\ge 1}(X)\cdot(2n)^{\delta}\ ,\end{aligned}$$

implying

$$\begin{aligned}(2n)^{\delta} \ge \mathbb{E}_{x\sim\widetilde{\mu}}[|f(x)|] &= \frac{1}{2}\cdot\mathbb{E}_{x\sim\mu}[|f(x)|] + \frac{\sum_{x\in X}|f(x)|}{2n}\\ &\ge \frac{1}{2}\cdot\mathbb{E}_{x\sim\mu}[|f(x)|] + \frac{1}{2}\ .\end{aligned}$$

(1) Set $\delta = \frac{1}{k}$, then we have multiplicative distortion $\frac{16}{\delta} = 16k$, and $\mathbb{E}_{x\sim\mu}[|f(x)|] \le 2\cdot(2n)^{\delta} = O(n^{\frac{1}{k}})$.



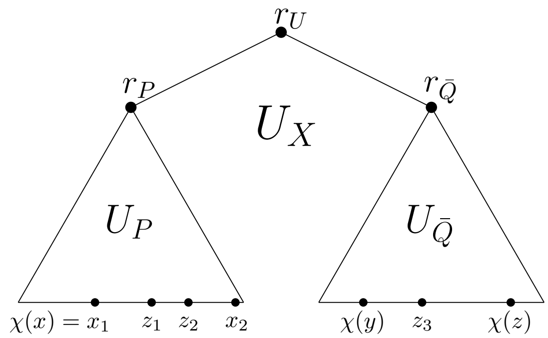


Figure 2: On the left illustrated the clusters $P, Q, \bar{Q}$ from Claim 1. On the right we illustrate the clan embedding of the metric space (X, d_X) into ultrametric U. r_U is the root of U, and its children are the roots of the ultrametrics $U_P, U_{\bar{Q}}$ which were constructed recursively. The point $x \in P \cap Q$ has $f(x) = f_P(x)$ and $\chi(x) = \chi_P(x)$ (where $|f(x)| = 2$). The point y is in $\bar{Q} \setminus P$ and thus $f(y) = f_{\bar{Q}}(y)$ and $\chi(y) = \chi_{\bar{Q}}(y)$ (there is a single copy of y). The point z belongs to $P \cap \bar{Q}$, where $d_X(v,z) > R + \frac{1}{16}\cdot \mathrm{diam}(X)$, hence $f(z) = f_P(z) \cup f_{\bar{Q}}(z)$ and $\chi(z) = \chi_{\bar{Q}}(z)$. Note that $|f_P(z)| = |f_{\bar{Q}}(z)| = 2$, and hence $|f(z)| = 4$.

(2) Choose $\delta \in (0,1]$ such that $\frac{1}{\delta} = \left\lceil \frac{\ln(2n)}{\ln(1+\epsilon/2)} \right\rceil$, note that $\delta \le \frac{\ln(1+\epsilon/2)}{\ln(2n)}$. Then we have multiplicative distortion $O(\frac{1}{\delta}) = O(\frac{\log n}{\epsilon})$, and $\mathbb{E}_{x\sim\mu}[|f(x)|] \le 2\cdot(2n)^{\delta} - 1 \le 1+\epsilon$.

□

Remark 1. *Lemma 4.2, note that for the clan embedding (f,χ) returned by Lemma 4.2 for input k, it holds that $|f(X)| \le \tilde{\mu}_{\ge 1}(X)^{1+\frac{1}{k}} = (2n)^{1+\frac{1}{k}}$. In particular, every $x \in X$ has at most $(2n)^{1+\frac{1}{k}}$ copies. Similarly, for input ϵ, $|f(X)| \le \tilde{\mu}_{\ge 1}(X)^{1+\delta} \le (2n)^{1+\frac{\ln(1+\epsilon/2)}{\ln 2n}} = 2n\cdot(1+\frac{\epsilon}{2})$. As for every $y \in X$, $f(y) \ne \emptyset$, it follows that for every $x \in X$, its number of copies is bounded by $|f(x)| = |f(X)| - |f(X \setminus \{x\})| \le 2n\cdot(1+\frac{\epsilon}{2}) - (n-1) = (1+\epsilon)n + 1$.*

Using the minimax theorem, as shown bellow, we show that there exists a distribution $\mathcal{D}$ of clan embeddings with distortion and expected clan size as specified by Theorem 1.1. Afterwards, using the multiplicative weights update (MWU) method, we explicitly construct such distributions efficiently, and with small support size.

Proof of Theorem 1.1 (exsistential agrument). Let μ be an arbitrary probability measure over the vertices, and $\mathcal{D}$ be any distribution over clan embeddings (f,χ) of (X, d_X) intro trees with multiplicative distortion $O(\frac{\log n}{\epsilon})$. Using Lemma 4.2 and the minimax theorem we have that

$$\min_{\mathcal{D}} \max_{\mu} \mathbb{E}_{(f,\chi)\sim\mathcal{D}, x\sim\mu}[|f(x)|] = \max_{\mu} \min_{(f,\chi)} \mathbb{E}_{x\sim\mu}[|f(x)|] \le 1+\epsilon\,.$$

Let $\mathcal{D}$ be the distribution from above, denote by μ_z the probability measure where $\mu_z(z) = 1$ (and $\mu_z(y) = 0$ for $y \ne z$). Then for every $x \in X$

$$\begin{aligned}\mathbb{E}_{(f,\chi)\sim\mathcal{D}}[|f(z)|] &= \mathbb{E}_{(f,\chi)\sim\mathcal{D}, x\sim\mu_z}[|f(x)|]\\ &\le \max_{\mu} \mathbb{E}_{(f,\chi)\sim\mathcal{D}, x\sim\mu}[|f(x)|] \le 1+\epsilon\,.\end{aligned}$$

The second claim of Theorem 1.1 could be proven using exactly the same argument. □

Constructive Proof of Theorem 1.1. Our construction relies on the multiplicative weights update method (MWU) [10] and the notion of a (ρ,α,β)-bounded Oracle.

Definition 4.3 ((ρ,α,β)-bounded Oracle). Given a probability measure μ over the metric points, a (ρ,α,β)-bounded Oracle returns a clan embedding (f,χ) with multiplicative distortion β such that:

(1) $\mathbb{E}_{x\sim\mu}[|f(x)|] \le \alpha$.
(2) $\max_{x\in V} |f(x)| \le \rho$.

In Lemma 4.4 below, we show that one can construct a uniform distribution $\mathcal{D}$ by making a polynomial number of oracle calls.

Lemma 4.4. *Given a (ρ,α,β)-bounded Oracle, and parameter $\epsilon \in (0,\frac{1}{2})$ one can construct a uniform distribution $\mathcal{D}$ over $O(\frac{\rho\alpha\log(n)}{\epsilon^2})$ clan embeddings with multiplicative distortion β such that:*

$$\text{For every } x \in X,\ \mathbb{E}_{(f,\chi)\sim\mathcal{D}}[|f(x)|] \le \alpha+\epsilon$$

Furthermore, the construction only makes $O(\frac{\rho\alpha\log(n)}{\epsilon^2})$ queries to the (ρ,α,β)-bounded Oracle.

Proof. Let O be a (ρ,α,β)-bounded Oracle and $O(\mu)$ be the clan embedding returned by the oracle given a probability measure μ. We follow the standard set up of MWU: we have n "experts" where the i-th expert is associated with the i-th point $x_i \in X$. The construction happens in T rounds. At the beginning of round t, we have a weight vector $\mathbf{w}^t = (w_1^t, \ldots, w_n^t)^\intercal$; at the first round, $\mathbf{w}^1 = (1,1,\ldots,1)^\intercal$.

The weight vector $\mathbf{w}^t$ induces a probability measure

$$\mu^t = (\frac{w_1^t}{W^t}, \ldots, \frac{w_n^t}{W^t}),$$

where $W^t = \sum_{i=1}^n w_i^t$. We construct a clan embedding $(f^t,\chi^t) = O(\mu^t)$ by making an oracle call to O with μ_t as input. Let $g_i^t =$

[10] For an excellent introduction of the MWU method and its historical account, see the survey by Arora, Hazan and Kale [8].

$\frac{|f^t(x_i)|}{\rho}$, and $\mathrm{g}^t = (g_1^t, \dots, g_n^t)^\intercal$ be the "penalty" vector for the set of n points (or experts). We then update:

$$w_i^{t+1} = (1+\delta)^{g_i^t} w_i^t \qquad \forall x_i \in X, \tag{1}$$

for some small parameter δ chosen later.

The penalty for each additional copy of each point is proportional to the number of copies it has in the clan embeddings constructed in previous steps. This is because in the next round, we will increase the measure of points with a large number of copies. Hence the oracle will be "motivated" to reduce the number of copies of these points in the next outputted clan embedding.

After T rounds, we have a collection

$$\mathcal{D}_T = \{(f^1, \chi^1), \dots, (f^T, \chi^T)\}$$

of T clan embeddings. The distribution $\mathcal{D}$ is constructed by sampling an embedding from $\mathcal{D}_T$ uniformly at random. Note that the distortion bound follows directly from the fact that the distortion of every clan embedding returned by the oracle is β. Our goal is to show that, by setting $T = O(\frac{\rho \log(n)}{\epsilon^2})$, we have:

$$\frac{1}{T} \cdot \sum_{t=1}^{T} |f^t(x_i)| \;\le\; \alpha + \epsilon \qquad \forall x_i \in V \tag{2}$$

To that end, we first observe that:

$$\begin{aligned} W^{t+1} &= \sum_{i=1}^{n} w_i^{t+1} = \sum_{i=1}^{n} (1+\delta)^{g_i^t} w_i^t \overset{(*)}{\le} \sum_{i=1}^{n} (1+\delta g_i^t) w_i^t \\ &= (1 + \sum_{i=1}^{n} \delta g_i^t \mu_i^t) W^t \le e^{\delta \langle \mathrm{g}^t, \mu^t \rangle} W^t \end{aligned}$$

where inequality $(*)$ follows from that $(1+x)^r \le (1+rx)$ for any $x \ge 0$ and $r \in [0,1]$. Thus, we have:

$$W^{T+1} \le e^{\delta \sum_{t=1}^{T} \langle \mathrm{g}^t, \mu^t \rangle} W^1 = e^{\delta \sum_{t=1}^{T} \langle \mathrm{g}^t, \mu^t \rangle} n \tag{3}$$

Observe that $W^{T+1} \ge w_i^{T+1} = (1+\delta)^{\sum_{t=1}^{T} g_i^t} w_i^1 = (1+\delta)^{\sum_{t=1}^{T} g_i^t}$ and that:

$$\begin{aligned} \sum_{t=1}^{T} \langle \mathrm{g}^t, \mu^t \rangle &= \sum_{t=1}^{T} \sum_{x \in X} \frac{|f^t(x)|}{\rho} \cdot \mu^t(v_i) \\ &= \frac{1}{\rho} \cdot \sum_{t=1}^{T} \mathbb{E}_{x \sim \mu^t}[|f^t(x)|] \le \frac{T\alpha}{\rho}. \end{aligned}$$

Thus, by equation (3), it holds that:

$$(1+\delta)^{\sum_{t=1}^{T} g_i^t} \le e^{\frac{\delta T \alpha}{\rho}} n. \tag{4}$$

Taking the natural logarithm from both sides we obtain that $\frac{\delta T \alpha}{\rho} + \ln n \ge \sum_{t=1}^{T} g_i^t \cdot \ln(1+\delta) = \frac{\ln(1+\delta)}{\rho} \cdot \sum_{t=1}^{T} |f^t(x_i)|$, and thus

$$\begin{aligned} \frac{1}{T} \cdot \sum_{t=1}^{T} |f^t(x_i)| &\le \frac{\rho}{T \cdot \ln(1+\delta)} \cdot \left(\frac{\delta T \alpha}{\rho} + \ln n \right) \\ &= \frac{\delta \alpha}{\ln(1+\delta)} + \frac{\rho \cdot \ln n}{T \cdot \ln(1+\delta)} \\ &\le \alpha(1 + \frac{\delta}{2}) + \frac{2\rho \cdot \ln n}{T \cdot \delta}, \end{aligned}$$

where the last inequality follows as $\frac{\delta}{\ln(1+\delta)} \le (1+\frac{\delta}{2})$ and $\ln(1+\delta) \ge \frac{\delta}{2}$ for $\delta \in (0, \frac{1}{2})$. By choosing $T = \frac{4\rho\alpha \ln n}{\epsilon^2} = O(\frac{\rho\alpha \log n}{\epsilon^2})$ and $\delta = \sqrt{\frac{4\rho \ln n}{T\alpha}} = \sqrt{\frac{\epsilon^2}{\alpha^2}} = \frac{\epsilon}{\alpha} < \frac{1}{2}$, we obtain that

$$\frac{1}{T} \cdot \sum_{t=1}^{T} |f^t(x_i)| \le \alpha + \frac{\delta \cdot \alpha}{2} + \frac{\epsilon^2}{2\alpha \cdot \delta} = \alpha + \epsilon,$$

satisfying equation (2), which completes our proof. ◻

Observe that Lemma 4.2, combined with Remark 1, provides an $(O(n), 1+\frac{\epsilon}{2}, O(\frac{\log n}{\epsilon}))$-bounded Oracle (when we apply Lemma 4.2 with parameter $\frac{\epsilon}{2}$). Using Lemma 4.4 with parameter $\frac{\epsilon}{2}$ provides us with an efficiently computable distribution over clan embeddings with support size $O(\frac{n \log n}{\epsilon^2})$, distortion $O(\frac{\log n}{\epsilon})$, and such that for every $x \in X$, $\mathbb{E}_{(f,\chi)\sim\mathcal{D}}[|f(x)|] \le 1+\epsilon$.

Similarly, by applying Lemma 4.2 with parameter k, we get an $(O(n^{1+\frac{1}{k}}), O(n^{\frac{1}{k}}), 16k)$-bounded Oracle. Thus Lemma 4.4 will produce an efficiently computable distribution over clan embeddings with support size $O(n^{1+\frac{2}{k}} \log n)$, distortion $16k$, and such that for every $x \in X$, $\mathbb{E}_{(f,\chi)\sim\mathcal{D}}[|f(x)|] = O(n^{\frac{1}{k}})$. Theorem 1.1 now follows.

5 LOWER BOUND FOR CLAN EMBEDDINGS INTO TREES

This section is devoted to proving Theorem 1.2, restated below.

Theorem 1.2 (Lower bound for clan embedding into a tree). *For every fixed $\epsilon \in (0,1)$ and large enough n, there is an n-point metric space (X, d_X) such that for every clan embedding (f, χ) of X into a tree with multiplicative distortion $O(\frac{\log n}{\epsilon})$, it holds that $\sum_{x\in X} |f(x)| \ge (1+\epsilon)n$.*
Furthermore, for every $k \in \mathbb{N}$, there is an n-point metric space (X, d_X) such that for every clan embedding (f, χ) of X into a tree with multiplicative distortion $O(k)$, it holds that $\sum_{x\in X} |f(x)| \ge \Omega(n^{1+\frac{1}{k}})$.

The girth of an unweighted graph G is the length of the shortest cycle in G. The Erdős' girth conjecture states that for any g and n, there exists an n-vertex graph with girth g and $\Omega(n^{1+\frac{2}{g-2}})$ edges. The conjecture is known to holds for $g = 4, 6, 8, 12$ (see [24, 73]). However, the best known lower bound for general k is due to Lazebnik *et al.* [61].

Theorem 5.1 ([61]). *For every even g, and n, there exists an unweighted graph with girth g and $\Omega(n^{1+\frac{4}{3}\cdot\frac{1}{g-2}})$ edges.*

From the upper bound perspective, the (generalized) Moore's bound [5, 12] states that every n vertex graph with girth g has at most $n^{1+\frac{2}{g-2}}$ edges for $g \le 2\log n$, and at most

$$n\left(1 + (1+o(1))\frac{\ln(m-n+1)}{g}\right)$$

edges for larger g; here m is the number of edges.

We will be able to use Theorem 5.1 to prove the second assertion in Theorem 1.2. That is, any clan embedding into a tree with distortion $O(k)$ must have $\sum_{x\in X} |f(x)| \ge \Omega(n^{1+\frac{1}{k}})$. However, the first assertion requires a much tighter lower bound of $(1+\epsilon)n$ on the number of edges. Therefore, the asymptotic nature of Theorem 5.1 is unfortunately not strong enough for our needs. We begin by showing that for large enough n and $\epsilon \in (0,1)$, there exists an n-vertex graph with $(1+\epsilon)n$ edges and girth $\Omega(\frac{\log n}{\epsilon})$. We are not aware of this very basic fact to previously appear in the literature. Note that Lemma 5.2 matches Moore's upper bound (up to a constant dependency on the girth g).

Lemma 5.2. *For every fixed $\epsilon \in (0,1)$ and large enough n, there exists a graph with at least $(1+\epsilon)n$ edges and girth $\Omega(\frac{\log n}{\epsilon})$.*

Remark 2 (Ultra sparse spanners)**.** *Given a graph $G = (V, E, w)$, a t-spanner is a subgraph H of G such that for every pair of vertices $u, v \in V$, $d_H(u,v) \le t \cdot d_G(u,v)$. For every fixed $\epsilon \in (0,1)$, Elkin and Neiman [44] constructed ultra-sparse spanners with $(1+\epsilon)n$ edges and stretch $O(\frac{\log n}{\epsilon})$. Even though they noted that the sparsity of their spanner matches the Moore's bound, it remained open whether one can construct better spanners. As the only $(g-2)$-spanner of a graph with girth g is the graph itself, Lemma 5.2 implies that the ultra sparse spanner from [44] is tight (up to a constant in the stretch).*

For the case of girth $\Omega(\log n)$, the first step is to replace the asymptotic notation in the lower bound on the number of edges from Theorem 5.1 with explicit bound.

Claim 2. *For every $n \in N$, there exist an n-vertex graph with $2n$ edges and girth $\Omega(\log n)$.*

Proof. Set $p = \frac{4n}{\binom{n}{2}} = \frac{8}{n-1}$. Consider a graph $G = (V, E)$ sampled according to $G(n,p)$ (that is, each edge sampled to G i.i.d. with probability p.). It holds that $\mathbb{E}[|E|] = \binom{n}{2} \cdot p = 4n$. By Chernoff bound,

$$\Pr\left[|E| < 3n\right] \le e^{-\frac{1}{32}\mathbb{E}[E]} = e^{-\frac{n}{8}} .$$

On the other hand, for $t \ge 3$, denote by C_t the set of cycles of length exactly t. Then,

$$\begin{aligned}\mathbb{E}\left[|C_t|\right] &\le n(n-1)\cdots(n-t+1)\cdot p^t\\ &= \frac{n(n-1)\cdots(n-t+1)}{(n-1)^t}\cdot 4^t < 4^t .\end{aligned}$$

Denote by C the set of all cycles of length smaller than $\frac{1}{3}\log n$. Then

$$\mathbb{E}\left[|C|\right] = \sum_{t=3}^{\frac{1}{3}\log n-1}\mathbb{E}\left[|C_t|\right] \le \sum_{t=3}^{\frac{1}{3}\log n-1} 4^t < 4^{\frac{1}{3}\log n} = n^{\frac{2}{3}} .$$

By Markov inequality, $\Pr\left[|C| \ge n\right] \le \frac{\mathbb{E}[|C|]}{n} < n^{-\frac{1}{3}} < \frac{1}{2}$. By union bound, there exists a graph G with at least $3n$ edges, and at most n cycles of length less than $\frac{1}{3}\log n$. Let G' be the graph obtained by deleting an arbitrary single edge from each cycle. Continue deleting edges until G' has exactly $2n$ edges. We conclude that G' has $2n$ edges and girth at least $\frac{1}{3}\log n$ as required. □

Proof of Lemma 5.2. Fix $\delta = \frac{1-\epsilon}{2\epsilon}$. Set $n' = \epsilon n = \frac{n}{1+2\delta}$. We ignore issues of integrality during the proof. Such issues could be easily fixed as we don't state an explicit bound on the girth. Using Claim 2, construct a graph G' with n' vertices, $2n'$ edges, and girth $\Omega(\log n')$.

Let G be the graph obtained from G' by replacing each edge with a path of length $\delta + 1$. Then:

$$\begin{aligned}|V(G)| &= |V(G')| + \delta\cdot|E(G')| = n' + \delta\cdot 2n' = n'(1+2\delta) = n\\ |E(G)| &= (\delta+1)\cdot|E(G')| = (\delta+1)\cdot 2n'\\ &= n\cdot\frac{2(1+\delta)}{1+2\delta} = (1+\epsilon)n ,\end{aligned}$$

where the last equality follows by the definition of δ. Note that the girth of G is at least $\Omega((1+\delta)\log n') = \Omega(\frac{\log \epsilon n}{\epsilon}) = \Omega(\frac{\log n}{\epsilon})$, for n large enough. □

The *Euler characteristic* of a graph G is defined as $\chi(G) := |E(G)| - |V(G)| + 1$. Our lower bound is based on the following theorem by Rabinovich and Raz [68].

Theorem 5.3 ([68]). *Consider an unweighted graph G with girth g, and consider a (classic) embedding $f: G \to H$ of G into a weighted graph H, such that $\chi(H) < \chi(G)$. Then f has multiplicative distortion at least $\frac{g}{4} - \frac{3}{2}$.*

Next, we transfer the language of classic embeddings into graphs used in Theorem 5.3 to that of clan embeddings into trees.

Lemma 5.4. *Consider an unweighted, n-vertex graph $G = (V, E)$ with girth g, and let (f, χ) be a clan embedding of G into a tree T with multiplicative distortion $t < \frac{g}{4} - \frac{3}{2}$. Then necessarily $\sum_{v\in V}|f(v)| \ge n + \chi(G)$.*

Proof. Let H be the graph obtained from T by merging all the copies of each vertex. Specifically, arbitrarily order the vertices in V: $v_1, v_2, \dots, v_n$. Iteratively construct a series of graphs $H_0 = T, H_1, H_2, \dots, H_n$ with one-to-many embeddings $f_i : G \to H_i$. In the i'th iteration, we create H_i, f_i out of H_{i-1}, f_{i-1} by replacing all the vertices in $f_{i-1}(v_i)$ by a single vertex $\tilde{v}_i$. For a vertex $u \in H_{i-1}$, we add an edge from u to $\tilde{v}_i$ if there was an edge from u to some vertex in $f_{i-1}(v)$. If an edge $\{u, \tilde{v}_i\}$ is added, its weight is defined to be $\min_{v'\in f_{i-1}(v)} w_{H_{i-1}}(v', u)$. Set $H = H_n$, and $\tilde{f} = f_n$. Clearly, distances in H can only decrease compared to T. This is because for every $u, v \in V$, $d_H(\tilde{u}, \tilde{v}) \le \min_{u'\in f(u),\, v'\in f(v)} d_T(u', v') \le \min_{u'\in f(u)} d_T(u', \chi(v)) \le t \cdot d_G(u,v)$. On the other hand, by induction (and the triangle inequality), since f is a dominating embedding, one can show that $\tilde{f}$ is also dominating. That is $\forall u, v \in V$, $d_H(\tilde{u}, \tilde{v}) \ge d_G(u,v)$.

We conclude that $\tilde{f}$ is a classic embedding of G with a multiplicative distortion at most $t < \frac{g}{4} - \frac{3}{2}$. By Theorem 5.3, it follows that $\chi(H) \ge \chi(G)$. For every i, it holds that

$$\begin{aligned}\chi(H_i) &= |E(H_i)| - |V(H_i)| - 1\\ &\le |E(H_{i-1})| - (|V(H_{i-1})| - |f(v_i)| + 1) - 1\\ &= \chi(H_{i-1}) + |f(v_i)| - 1\end{aligned}$$

As the Euler characteristic of a tree equals 0, we obtain

$$\chi(G) \le \chi(H) = \chi(H_n) \le \sum_i (|f(v_i)| - 1) + \chi(T) = \sum_{v\in V}|f(v)| - n ,$$

as desired. □

We are now ready to prove Theorem 1.2.

Proof of Theorem 1.2. For the first assertion, using Lemma 5.2, let G be an unweighted graph with girth $g = \Omega(\frac{\log n}{\epsilon})$ and $(1+\epsilon)n$ edges. Consider a clan embedding of G into a tree with distortion smaller than $\frac{g}{4} - \frac{3}{2} = \Omega(\frac{\log n}{\epsilon})$. By Lemma 5.4, it holds that

$$\sum_{v\in V} |f(v)| \geq n + \chi(G) = |E(G)| + 1 > (1+\epsilon)n\ .$$

The second assertion follows similar lines. Set $g = 2 \cdot \left\lfloor \frac{\frac{4}{3}k+2}{2} \right\rfloor$. Note that g is largest even number up to $\frac{4}{3}k+2$. Using Theorem 5.1, let G be an unweighted graph with girth g and $\Omega(n^{1+\frac{4}{3}\cdot\frac{1}{g-2}}) \geq \Omega(n^{1+\frac{1}{k}})$ edges. Consider a clan embedding of G into a tree with distortion smaller than $\frac{g}{4} - \frac{3}{2} = \Omega(k)$. By Lemma 5.4, it holds that

$$\sum_{v\in V} |f(v)| \geq n + \chi(G) = |E(G)| + 1 = \Omega(n^{1+\frac{1}{k}})\ .$$

□

ACKNOWLEDGMENTS

The authors are grateful to Philip Klein for suggesting the metric ρ-dominating/independent set problems, which eventually led to this project. We thank Vincent Cohen-Addad for useful conversations. The first author would like to thank Alexandr Andoni for helpful discussions. The second author would like to thank Michael Lampis for discussing dynamic programming algorithms for metric independent set/dominating set on bounded treewidth graphs. Arnold Filtser was supported by the Simons Foundation; Hung Le was supported by the start-up grant of UMass Amherst.

REFERENCES

[1] Ittai Abraham, Shiri Chechik, Michael Elkin, Arnold Filtser, and Ofer Neiman. 2020. Ramsey Spanning Trees and Their Applications. *ACM Trans. Algorithms* 16, 2 (2020), 19:1–19:21. https://doi.org/10.1145/3371039 preliminary version published in SODA 2018.

[2] Ittai Abraham, Arnold Filtser, Anupam Gupta, and Ofer Neiman. 2018. Metric embedding via shortest path decompositions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018.* 952–963. https://doi.org/10.1145/3188745.3188808 full version: https://arxiv.org/abs/1708.04073.

[3] Ittai Abraham, Cyril Gavoille, Anupam Gupta, Ofer Neiman, and Kunal Talwar. 2019. Cops, Robbers, and Threatening Skeletons: Padded Decomposition for Minor-Free Graphs. *SIAM J. Comput.* 48, 3 (2019), 1120–1145. https://doi.org/10.1137/17M1112406 preliminary version published in STOC 2014.

[4] Ittai Abraham and Ofer Neiman. 2019. Using Petal-Decompositions to Build a Low Stretch Spanning Tree. *SIAM J. Comput.* 48, 2 (2019), 227–248. https://doi.org/10.1137/17M1115575 preliminary version published in STOC 2012.

[5] Noga Alon, Shlomo Hoory, and Nathan Linial. 2002. The Moore Bound for Irregular Graphs. *Graphs Comb.* 18, 1 (2002), 53–57. https://doi.org/10.1007/s003730200002

[6] Noga Alon, Richard M. Karp, David Peleg, and Douglas B. West. 1995. A Graph-Theoretic Game and Its Application to the k-Server Problem. *SIAM J. Comput.* 24, 1 (1995), 78–100. https://doi.org/10.1137/S0097539792224474 preliminary version published in On-Line Algorithms 1991.

[7] Noga Alon, Yossi Matias, and Mario Szegedy. 1999. The Space Complexity of Approximating the Frequency Moments. *J. Comput. Syst. Sci.* 58, 1 (1999), 137–147. https://doi.org/10.1006/jcss.1997.1545 preliminary version published in STOC 1996.

[8] S. Arora, E. Hazan, and S. Kale. 2012. The Multiplicative Weights Update Method: a Meta-Algorithm and Applications. *Theory of Computing* 8, 1 (2012), 121–164. https://doi.org/10.4086/toc.2012.v008a006

[9] Vassilis Athitsos and Stan Sclaroff. 2003. Database Indexing Methods for 3D Hand Pose Estimation. In *Gesture-Based Communication in Human-Computer Interaction, 5th International Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers.* 288–299. https://doi.org/10.1007/978-3-540-24598-8_27

[10] Baruch Awerbuch, Amotz Bar-Noy, Nathan Linial, and David Peleg. 1990. Improved Routing Strategies with Succinct Tables. *J. Algorithms* 11, 3 (1990), 307–341. https://doi.org/10.1016/0196-6774(90)90017-9

[11] Baruch Awerbuch and David Peleg. 1992. Routing with Polynomial Communication-Space Trade-Off. *SIAM J. Discret. Math.* 5, 2 (1992), 151–162. https://doi.org/10.1137/0405013

[12] Ajesh Babu and Jaikumar Radhakrishnan. 2010. An entropy based proof of the Moore bound for irregular graphs. *CoRR* abs/1011.1058 (2010). arXiv:1011.1058 http://arxiv.org/abs/1011.1058

[13] B. S. Baker. 1994. Approximation Algorithms for NP-complete Problems on Planar Graphs. *J. ACM* 41, 1 (1994), 153–180. https://doi.org/10.1145/174644.174650 preliminary version published in FOCS 1983.

[14] Yair Bartal. 1996. Probabilistic Approximations of Metric Spaces and Its Algorithmic Applications. In *37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October, 1996.* 184–193. https://doi.org/10.1109/SFCS.1996.548477

[15] Yair Bartal. 1998. On Approximating Arbitrary Metrices by Tree Metrics. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998.* 161–168. https://doi.org/10.1145/276698.276725

[16] Yair Bartal. 2004. Graph Decomposition Lemmas and Their Role in Metric Embedding Methods. In *Algorithms - ESA 2004, 12th Annual European Symposium, Bergen, Norway, September 14-17, 2004, Proceedings.* 89–97. https://doi.org/10.1007/978-3-540-30140-0_10

[17] Yair Bartal. 2011. Lecture notes in Metric Embedding Theory and Its Algorithmic Applications. URL: http://moodle.cs.huji.ac.il/cs10/file.php/67720/GM_Lecture6.pdf.

[18] Yair Bartal. 2021. Advances in Metric Ramsey Theory and its Applications. *CoRR* abs/2104.03484 (2021). arXiv:2104.03484 https://arxiv.org/abs/2104.03484

[19] Yair Bartal, Béla Bollobás, and Manor Mendel. 2006. Ramsey-type theorems for metric spaces with applications to online problems. *J. Comput. Syst. Sci.* 72, 5 (2006), 890–921. https://doi.org/10.1016/j.jcss.2005.05.008 Special Issue on FOCS 2001.

[20] Yair Bartal, Nova Fandina, and Ofer Neiman. 2019. Covering Metric Spaces by Few Trees. In *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece.* 20:1–20:16. https://doi.org/10.4230/LIPIcs.ICALP.2019.20

[21] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. 2005. On Metric Ramsey-Type Dichotomies. *Journal of the London Mathematical Society* 71, 2 (2005), 289–303. https://doi.org/10.1112/S0024610704006155

[22] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. 2005. Some Low Distortion Metric Ramsey Problems. *Discret. Comput. Geom.* 33, 1 (2005), 27–41. https://doi.org/10.1007/s00454-004-1100-z

[23] Yair Bartal and Manor Mendel. 2004. Multiembedding of Metric Spaces. *SIAM J. Comput.* 34, 1 (2004), 248–259. https://doi.org/10.1137/S0097539703433122 preliminary version published in SODA 2003.

[24] Clark T. Benson. 1966. Minimal Regular Graphs of Girths Eight and Twelve. *Canadian Journal of Mathematics* 18 (1966), 1091â€"1094. https://doi.org/10.4153/CJM-1966-109-8

[25] Guy E. Blelloch, Yan Gu, and Yihan Sun. 2016. A New Efficient Construction on Probabilistic Tree Embeddings. *CoRR* abs/1605.04651 (2016). arXiv:1605.04651 http://arxiv.org/abs/1605.04651 https://arxiv.org/abs/1605.04651.

[26] Glencora Borradaile and Hung Le. 2016. Optimal Dynamic Program for r-Domination Problems over Tree Decompositions. In *11th International Symposium on Parameterized and Exact Computation, IPEC 2016, August 24-26, 2016, Aarhus, Denmark.* 8:1–8:23. https://doi.org/10.4230/LIPIcs.IPEC.2016.8

[27] G. Borradaile, H. Le, and C. Wulff-Nilsen. 2017. Minor-free graphs have light spanners. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS '17).* 767–778. https://doi.org/10.1109/FOCS.2017.76

[28] J. Bourgain. 1985. On lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics* 52, 1-2 (1985), 46–52. https://doi.org/10.1007/BF02776078

[29] J. Bourgain, T. Figiel, and V. Milman. 1986. On Hilbertian subsets of finite metric spaces. *Israel J. Math.* 55, 2 (1986), 147–152. https://doi.org/10.1007/BF02801990

[30] Sébastien Bubeck, Michael B. Cohen, Yin Tat Lee, James R. Lee, and Aleksander Madry. 2018. k-server via multiscale entropic regularization. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018.* 3–16. https://doi.org/10.1145/3188745.3188798

[31] Douglas E. Carroll and Ashish Goel. 2004. Lower Bounds for Embedding into Distributions over Excluded Minor Graph Families. In *Algorithms - ESA 2004, 12th Annual European Symposium, Bergen, Norway, September 14-17, 2004, Proceedings.* 146–156. https://doi.org/10.1007/978-3-540-30140-0_15

[32] Amit Chakrabarti, Alexander Jaffe, James R. Lee, and Justin Vincent. 2008. Embeddings of Topological Graphs: Lossy Invariants, Linearization, and 2-Sums. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA.* 761–770. https://doi.org/10.1109/FOCS.2008.79

[33] T.-H. Hubert Chan and Anupam Gupta. 2012. Approximating TSP on Metrics with Bounded Global Growth. *SIAM J. Comput.* 41, 3 (2012), 587–617. https://doi.org/10.1137/090749396 preliminary version published in SODA 2008.

[34] Shiri Chechik. 2013. Compact routing schemes with improved stretch. In *ACM Symposium on Principles of Distributed Computing, PODC '13, Montreal, QC, Canada, July 22-24, 2013*. 33–41. https://doi.org/10.1145/2484239.2484268

[35] Shiri Chechik. 2015. Approximate Distance Oracles with Improved Bounds. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*. 1–10. https://doi.org/10.1145/2746539.2746562

[36] Vincent Cohen-Addad, Arnold Filtser, Philip N. Klein, and Hung Le. 2020. On Light Spanners, Low-treewidth Embeddings and Efficient Traversing in Minor-free Graphs. *CoRR* abs/2009.05039 (2020). arXiv:2009.05039 https://arxiv.org/abs/2009.05039 To appear in FOCS 2020,https://arxiv.org/abs/2009.05039.

[37] Vincent Cohen-Addad, Philip N. Klein, and Claire Mathieu. 2019. Local Search Yields Approximation Schemes for k-Means and k-Median in Euclidean and Minor-Free Metrics. *SIAM J. Comput.* 48, 2 (2019), 644–667. https://doi.org/10.1137/17M112717X preliminary version published in FOCS 2016.

[38] Lenore Cowen. 2001. Compact Routing with Minimum Stretch. *J. Algorithms* 38, 1 (2001), 170–183. https://doi.org/10.1006/jagm.2000.1134 preliminary version published in SODA 1999.

[39] Erik D. Demaine, Fedor V. Fomin, Mohammad Taghi Hajiaghayi, and Dimitrios M. Thilikos. 2005. Fixed-parameter algorithms for (k, r)-center in planar graphs and map graphs. *ACM Trans. Algorithms* 1, 1 (2005), 33–47. https://doi.org/10.1145/1077464.1077468 preliminary version published in ICALP 2003.

[40] E. D. Demaine, M. Hajiaghayi, and K. Kawarabayashi. 2005. Algorithmic Graph Minor Theory: Decomposition, Approximation, and Coloring. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*. 637–646. https://doi.org/10.1109/SFCS.2005.14

[41] Tamar Eilam, Cyril Gavoille, and David Peleg. 2003. Compact routing schemes with low stretch factor. *J. Algorithms* 46, 2 (2003), 97–114. https://doi.org/10.1016/S0196-6774(03)00002-6 preliminary version published in PODC 1998.

[42] David Eisenstat, Philip N. Klein, and Claire Mathieu. 2014. Approximating k-center in planar graphs. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*. 617–627. https://doi.org/10.1137/1.9781611973402.47

[43] Michael Elkin, Yuval Emek, Daniel A. Spielman, and Shang-Hua Teng. 2008. Lower-Stretch Spanning Trees. *SIAM J. Comput.* 38, 2 (2008), 608–628. https://doi.org/10.1137/050641661 preliminary version published in STOC 2005.

[44] Michael Elkin and Ofer Neiman. 2019. Efficient Algorithms for Constructing Very Sparse Spanners and Emulators. *ACM Trans. Algorithms* 15, 1 (2019), 4:1–4:29. https://doi.org/10.1145/3274651 preliminary version published in SODA 2017.

[45] Hiroshi Eto, Takehiro Ito, Zhilong Liu, and Eiji Miyano. 2016. Approximability of the Distance Independent Set Problem on Regular Graphs and Planar Graphs. In *Combinatorial Optimization and Applications - 10th International Conference, COCOA 2016, Hong Kong, China, December 16-18, 2016, Proceedings*. 270–284. https://doi.org/10.1007/978-3-319-48749-6_20

[46] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. 2004. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. Syst. Sci.* 69, 3 (Nov. 2004), 485–497. https://doi.org/10.1016/j.jcss.2004.04.011 preliminary version published in STOC 2003.

[47] Andreas Emil Feldmann, Wai Shing Fung, Jochen Könemann, and Ian Post. 2018. A $(1+\epsilon)$-Embedding of Low Highway Dimension Graphs into Bounded Treewidth Graphs. *SIAM J. Comput.* 47, 4 (2018), 1667–1704. https://doi.org/10.1137/16M1067196 preliminary version published in ICALP 2015.

[48] Arnold Filtser. 2019. On Strong Diameter Padded Decompositions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2019, September 20-22, 2019, Massachusetts Institute of Technology, Cambridge, MA, USA*. 6:1–6:21. https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2019.6

[49] A. Filtser and H. Le. 2021. Clan Embeddings into Trees, and Low Treewidth Graphs. *CoRR* abs/2101.01146 (2021). https://arxiv.org/abs/2101.01146.

[50] Arnold Filtser and Hung Le. 2021. Reliable Spanners: Locality-Sensitive Orderings Strike Back. *CoRR* abs/2101.07428 (2021). arXiv:2101.07428 https://arxiv.org/abs/2101.07428

[51] E. Fox-Epstein, P. N. Klein, and A. Schild. 2019. Embedding Planar Graphs into Low-Treewidth Graphs with Applications to Efficient Approximation Schemes for Metric Problems. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '19)*. 1069–1088. https://doi.org/10.1137/1.9781611975482.66

[52] Lee-Ad Gottlieb, Aryeh Kontorovich, and Robert Krauthgamer. 2017. Efficient Regression in Metric Spaces via Approximate Lipschitz Extension. *IEEE Trans. Inf. Theory* 63, 8 (2017), 4838–4849. https://doi.org/10.1109/TIT.2017.2713820 preliminary version published in SIMBAD 2013.

[53] Bernhard Haeupler, D. Ellis Hershkowitz, and Goran Zuzic. 2021. Deterministic Tree Embeddings with Copies for Algorithms Against Adaptive Adversaries. *CoRR* abs/2102.05168 (2021). arXiv:2102.05168 https://arxiv.org/abs/2102.05168

[54] Eran Halperin, Jeremy Buhler, Richard M. Karp, Robert Krauthgamer, and B. Westover. 2003. Detecting protein sequence conservation via metric embeddings. *Bioinformatics* 19, suppl 1 (07 2003), i122–i129. https://doi.org/10.1093/bioinformatics/btg1016 arXiv:https://academic.oup.com/bioinformatics/article-pdf/19/suppl_1/i122/614436/btg1016.pdf

[55] Piotr Indyk. 2001. Algorithmic Applications of Low-Distortion Geometric Embeddings. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*. 10–33. https://doi.org/10.1109/SFCS.2001.959878

[56] Richard M Karp. 1989. A 2k-competitive algorithm for the circle. *Manuscript, August* 5 (1989).

[57] Ioannis Katsikarelis, Michael Lampis, and Vangelis Th. Paschos. 2019. Structural parameters, tight bounds, and approximation for (k, r)-center. *Discret. Appl. Math.* 264 (2019), 90–117. https://doi.org/10.1016/j.dam.2018.11.002 preliminary version published in ISAAC 2017.

[58] Ioannis Katsikarelis, Michael Lampis, and Vangelis Th. Paschos. 2020. Structurally parameterized d-scattered set. *Discrete Applied Mathematics* (2020). https://doi.org/10.1016/j.dam.2020.03.052 preliminary version published in WG 2018.

[59] Maleq Khan, Fabian Kuhn, Dahlia Malkhi, Gopal Pandurangan, and Kunal Talwar. 2012. Efficient distributed approximation algorithms via probabilistic tree embeddings. *Distributed Comput.* 25, 3 (2012), 189–205. https://doi.org/10.1007/s00446-012-0157-9 preliminary version published in PODC 2008.

[60] Robert Krauthgamer, James R. Lee, Manor Mendel, and Assaf Naor. 2005. Measured descent: a new embedding method for finite metrics. *Geometric and Functional Analysis* 15, 4 (2005), 839–858. https://doi.org/10.1007/s00039-005-0527-6 preliminary version published in FOCS 2004.

[61] Felix Lazebnik, Vasiliy A Ustimenko, and Andrew J Woldar. 1995. A new series of dense graphs of high girth. *Bulletin of the American mathematical society* 32, 1 (1995), 73–79. https://doi.org/10.1090/S0273-0979-1995-00569-0

[62] Alex Lemin. 2003. On ultrametrization of general metric spaces. *Proceedings of the American mathematical society* 131, 3 (2003), 979–989. https://doi.org/10.1090/S0002-9939-02-06605-4

[63] Nathan Linial, Eran London, and Yuri Rabinovich. 1995. The Geometry of Graphs and Some of its Algorithmic Applications. *Comb.* 15, 2 (1995), 215–245. https://doi.org/10.1007/BF01200757 preliminary version published in FOCS 1994.

[64] Dániel Marx and Michal Pilipczuk. 2015. Optimal Parameterized Algorithms for Planar Facility Location Problems Using Voronoi Diagrams. In *Algorithms - ESA 2015 - 23rd Annual European Symposium, Patras, Greece, September 14-16, 2015, Proceedings*. 865–877. https://doi.org/10.1007/978-3-662-48350-3_72

[65] Manor Mendel and Assaf Naor. 2007. Ramsey partitions and proximity data structures. *Journal of the European Mathematical Society* 9, 2 (2007), 253–275. https://doi.org/10.4171/JEMS/79 preliminary version published in FOCS 2006.

[66] Assaf Naor and Terence Tao. 2012. Scale-oblivious metric fragmentation and the nonlinear Dvoretzky theorem. *Israel Journal of Mathematics* 192, 1 (2012), 489–504. https://doi.org/10.1007/s11856-012-0039-7

[67] David Peleg and Eli Upfal. 1989. A trade-off between space and efficiency for routing tables. *J. ACM* 36, 3 (1989), 510–530. https://doi.org/10.1145/65950.65953

[68] Yuri Rabinovich and Ran Raz. 1998. Lower Bounds on the Distortion of Embedding Finite Metric Spaces in Graphs. *Discret. Comput. Geom.* 19, 1 (1998), 79–94. https://doi.org/10.1007/PL00009336

[69] Satish Rao. 1999. Small Distortion and Volume Preserving Embeddings for Planar and Euclidean Metrics. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry, Miami Beach, Florida, USA, June 13-16, 1999*. 300–306. https://doi.org/10.1145/304893.304983

[70] N. Robertson and P. D. Seymour. 2003. Graph Minors. XVI. Excluding a Non-planar Graph. *Journal of Combinatoral Theory Series B* 89, 1 (2003), 43–76. https://doi.org/10.1016/S0095-8956(03)00042-X

[71] Kunal Talwar. 2004. Bypassing the embedding: algorithms for low dimensional metrics. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing* (Chicago, IL, USA). ACM Press, 281–290. https://doi.org/10.1145/1007352.1007399

[72] Mikkel Thorup and Uri Zwick. 2001. Compact routing schemes. In *Proceedings of the Thirteenth Annual ACM Symposium on Parallel Algorithms and Architectures, SPAA 2001, Heraklion, Crete Island, Greece, July 4-6, 2001*. 1–10. https://doi.org/10.1145/378580.378581

[73] R Wenger. 1991. Extremal graphs with no C4's, C6's, or C10's. *Journal of Combinatorial Theory, Series B* 52, 1 (1991), 113 – 116. https://doi.org/10.1016/0095-8956(91)90097-4