Hop-Constrained Metric Embeddings and their Applications

Arnold Filtser
Department of Computer Science
Bar-Ilan University
Ramat Gan, Israel
Email: arnold273@gmail.com

***Abstract*—In network design problems, such as compact routing, the goal is to route packets between nodes using the (approximated) shortest paths. A desirable property of these routes is a small number of hops, which makes them more reliable, and reduces the transmission costs.**

Following the overwhelming success of stochastic tree embeddings for algorithmic design, Haeupler, Hershkowitz, and Zuzic (STOC'21) studied hop-constrained Ramsey-type metric embeddings into trees. Specifically, embedding $f : G(V,E) \to T$ has Ramsey hop-distortion (t, M, β, h) (here $t, \beta, h \ge 1$ and $M \subseteq V$) if $\forall u \in M, v \in V$, $d_G^{(\beta\cdot h)}(u,v) \le d_T(u,v) \le t \cdot d_G^{(h)}(u,v)$. t is called the distortion, β is called the hop-stretch, and $d_G^{(h)}(u,v)$ denotes the minimum weight of a $u-v$ path with at most h hops. Haeupler *et al.* constructed embedding where M contains $1-\epsilon$ fraction of the vertices and $\beta = t = O(\frac{\log^2 n}{\epsilon})$. They used their embedding to obtain multiple bicriteria approximation algorithms for hop-constrained network design problems.

In this paper, we first improve the Ramsey-type embedding to obtain parameters $t = \beta = \frac{\tilde{O}(\log n)}{\epsilon}$, and generalize it to arbitrary distortion parameter t (in the cost of reducing the size of M). This embedding immediately implies polynomial improvements for all the approximation algorithms from Haeupler *et al.*. Further, we construct hop-constrained clan embeddings (where each vertex has multiple copies), and use them to construct bicriteria approximation algorithms for the group Steiner tree problem, matching the state of the art of the non constrained version. Finally, we use our embedding results to construct hop constrained distance oracles, distance labeling, and most prominently, the first hop constrained compact routing scheme with provable guarantees. All our metric data structures almost match the state of the art parameters of the non-constrained versions.

The reader is encouraged to read the full version of this paper [47]: arXiv:2106.14969.

***Keywords*-Metric embeddings; Hop constrained tree embeddings; Approximation algorithms; Group Steiner tree; Compact routing scheme; Distance oracle; Distance labelings;**

I. Introduction

Low distortion metric embeddings provide a powerful algorithmic toolkit, with applications ranging from approximation/sublinear/online/distributed algorithms [73], [7], [31], [67] to machine learning [56], biology [64], and vision [11]. The basic idea in order to solve a problem in a "hard" metric space (X, d_X), is to embed the points X into a "simple" metric space (Y, d_Y) that preserves all pairwise distances up to small multiplicative factor. Then one solve the problem in Y, and "pull-back" the solution into X.

A highly desirable target space is trees, as many hard problems become easy once the host space is a tree metric. Fakcharoenphol, Rao, and Talwar [45] (improving over [6], [16], [17], see also [18]) showed that every n point metric space could be embedded into distribution $\mathcal{D}$ over dominating trees with expected distortion $O(\log n)$. Formally, $\forall x, y \in X$, $T \in \text{supp}(\mathcal{D})$, $d_X(x,y) \le d_T(x,y)$, and $\mathbb{E}_{T\sim\mathcal{D}}[d_T(x,y)] \le O(\log n)\cdot d_X(x,y)$. This stochastic embedding enjoyed tremendous success and has numerous applications. However, the distortion guarantee is only in expectation. A different solution is Ramsey type embeddings which have a worst case guarantee, however only w.r.t. a subset of the points. Specifically, Mendel and Naor [76] (see also [30], [20], [22], [77], [25], [2], [19]) showed that for every parameter k, every n-point metric space contains a subset M of at least $n^{1-\frac{1}{k}}$ points, where X could be embedded into a tree such that all the distances in $M \times X$ are preserved up to an $O(k)$ multiplicative factor. Formally, $\forall x, y \in X$, $d_X(x,y) \le d_T(x,y)$, and $\forall x \in M, y \in X$, $d_T(x,y) \le O(k)\cdot d_X(x,y)$. Finally, in order to obtain worst case guarantee w.r.t. all point pairs, the author and Hung [49] introduced clan embeddings into trees. Here each point x is mapped into a subset $f(x) \subseteq Y$ which called copies of x, with a special chief copy $\chi(x) \in f(x)$. For every parameter k, [49] constructed distribution over dominating clan embeddings such that for every pair of vertices x, y, some copy of x is close to the chief of y: $d_X(x,y) \le \min_{x'\in f(x), y'\in f(y)} d_T(x', y') \le \min_{x'\in f(x)} d_T(x', \chi(y)) \le O(k)\cdot d_X(x,y)$, and the expected number of copies of every vertex is bounded: $\mathbb{E}[|f(x)|] = O(n^{\frac{1}{k}})$.

In many applications, the metric space is the shortest path metric d_G of a weighted graph $G = (V, E, w)$. Here, in addition to metric distances, there are often hop-constrains. For instance, one may wish

2575-8454/21/$31.00 ©2021 IEEE
DOI 10.1109/FOCS52979.2021.00056

to route a packet between two nodes, using a path with at most h hops, [1] i.e. the number of edges in the path. Such hop constrains are desirable as each transmission causes delays, which are non-negligible when the number of transmissions is large [5], [27]. Another advantage is that low-hop routes are more reliable: if each transmission is prone to failure with a certain probability, then low-hop routes are much more likely to reach their destination [27], [89], [83]. Electricity and telecommunications distribution network configuration problems include hop constraints that limit the maximum number of edges between a customer and its feeder [27], and there are many other (practical) network design problems with hop constraints [71], [14], [59], [58], [80]. Hop-constrained network approximation is often used in parallel computing [35], [9], as the number of hops governs the number of required parallel rounds (e.g. in Dijkstra). Finally, there is an extensive work on approximation algorithms for connectivity problems like spanning tree and Steiner tree with hop constraints [81], [70], [74], [8], [69], [66], [63], [68], and for some generalizations in [62].

Given a weighted graph $G = (V, E, w)$, $d_G^{(h)}(u, v)$ denotes the minimum weight of a u-v path containing at most h hops. Following the tremendous success of tree embeddings, Haeupler, Hershkowitz and Zuzic [62] suggested to study hop-constrained tree embeddings. That is, embedding the vertex set V into a tree T, such that $d_T(u, v)$ will approximate the h-hop constrained distance $d_G^{(h)}(u, v)$. Unfortunately, [62] showed that $d_G^{(h)}$ is very far from being a metric space, and the distortion of every embedding of $d_G^{(h)}$ is unbounded. To overcome this issue, [62] allowed hop-stretch. Specifically, allowing bi-criteria approximation in the following sense: $d_G^{(\beta \cdot h)}(u, v) \le d_T(u, v) \le t \cdot d_G^{(h)}(u, v)$, for some parameters β, and t. However, even this relaxation is not enough as long as $\beta \cdot h < n - 1$. To see this, consider the unweighted n-path $P_n = \{v_0, \dots, v_{n-1}\}$. Then $d_G^{(n-2)}(v_0, v_{n-1}) = \infty$, implying that for every metric X over V, $\max_i\{d_X(v_i, v_{i+1})\} \ge \frac{1}{n-1}\sum_i d_X(v_i, v_{i+1}) \ge \frac{d_X(v_0, v_{n-1})}{n-1} \ge \frac{d_G^{(n-2)}(v_0, v_{n-1})}{n-1} = \infty$. As a result, some pairs have distortion ∞, and hence no bi-criteria metric embedding is possible. Thus, there is no hop-constrained counterpart for [45].

To overcome this barrier, [62] studied hop-constrained Ramsey-type tree embeddings. [2] Their main result, is that every n-point weighted graph $G = (V, E, w)$ with polynomial aspect ratio,[3] and parameter $\epsilon \in (0, 1)$, there is distribution over pairs (T, M), where $M \subseteq V$, and T is a tree with M as its vertex set such that $\forall u, v \in M$,

$$d_G^{(O(\frac{\log^2 n}{\epsilon}) \cdot h)}(u, v) \le d_T(u, v) \le O(\frac{\log^2 n}{\epsilon}) \cdot d_G^{(h)}(u, v) ,$$

and for every $v \in V$, $\Pr[v \in M] \ge 1 - \epsilon$ (referred to as inclusion probability). Improving the distortion, and hop-stretch, was left as the "main open question" [62].

Question 1 ([62])**.** *Is it possible to construct better hop-constrained Ramsey-type tree embeddings? What are the optimal parameters?*

Adding hop-constrains to a problem often makes it considerably harder. For example, obtaining $o(\log n)$ approximation for the Hop constrained MST problem (where the hop-diameter must be bounded by h) is NP-hard [15], while MST is simple problem without it. Similarly, for every $\epsilon > 0$, obtaining $o(2^{\log^{1-\epsilon} n})$-approximation for the hop-constrained Steiner forest problem is NP-hard [40] (while constant approximation is possible without hop constrains [4]).

[62] used their hop-constrained Ramsey-type tree embeddings to construct many approximation algorithms. Roughly, given that some problem has (non constrained) approximation ratio α for trees, [62] obtain a bicriteria approximation for the hop constrained problem with cost $O(\alpha \cdot \log^3 n)$, and hop-stretch $O(\log^3 n)$. For example, in the group Steiner tree problem we are given k sets $g_1, \dots, g_k \subseteq V$, and a root vertex $r \in V$. The goal is to construct a minimal weight tree spanning the vertex r and at least one vertex from each group (set) g_i. In the hop-constrained version of the problem, we are given in addition a hop-bound h, and the requirement is to find a minimum weight subgraph H, such that for every g_i, $\mathrm{hop}_H(r, g_i) \le h$. Denote by OPT the weight of the optimal solution. In the non-constrained version, where the input metric is a tree, [53] provided a $O(\log n \cdot \log k)$ approximation. Using the scheme of [62], one can find a solution H of weight $O(\log^4 n \cdot \log k) \cdot \mathrm{OPT}$, such that for every g_i, $\mathrm{hop}_H(r, g_i) \le O(\log^3 n) \cdot h$. The best approximation for the non-constrained group Steiner tree problem is $O(\log^2 n \cdot \log k)$ [53]. Clearly, every improvement over the hop-constrained Ramsey-type embedding will imply better bicriteria approximation algorithms. However, even if one would use [62] framework on optimal Ramsey-type embeddings (ignoring hop-constrains),

[1] To facilitate the reading of the paper, all numbers referring to hops are colored in red.

[2] [62] called this embedding "partial metric distribution", rather than Ramsey-type.

[3] The *aspect ratio* (sometimes referred to as spread), is the ratio between the maximal to minimal weight in G $\frac{\max_{e \in E} w(e)}{\min_{e \in E} w(e)}$. Often in the literature the aspect ratio defined as $\frac{\max_{u,v \in V} d_G(u,v)}{\min_{u,v \in V} d_G(u,v)}$: the ratio between the maximal to minimal distance in G, which is equivalent up to a factor n to the definition used here.

the cost factors will be inferior compared to the non-constrained versions (specifically $O(\log^3 n\cdot\log k)$ for GST). It is interesting to understand whether there is a separation between the cost of a bicriteria approximation for hop-constrained problems to their non-constrained counterparts. A more specific question is the following:

Question 2. *Could one match the state of the art cost approximation of the (non hop-constrained) group Steiner tree problem while having* $\mathrm{polylog}(n)$ *hop-stretch?*

Metric data structures such as distance oracles, distance labeling, and compact routing schemes are extensively studied, and widely used throughout algorithmic design. A natural question is whether one can construct similar data structures w.r.t. hop-constrained distances. Most prominently is the question of compact routing scheme [86], where one assigns each node in a network a short label and small local routing table, such that packets could be routed throughout the network with routes not much longer than the shortest paths. Hop-constrained routing is a natural question due to it's clear advantages (bounding transmission delays, and increasing reliability). Indeed, numerous papers developed different heuristics for different models of hop-constrained routing. [4] However, to the best knowledge of the author, no prior provable guarantees were provided for hop-constrained compact routing scheme.

Question 3. *What hop-constrained compact routing schemes are possible?*

A. Our contribution

1) Metric embeddings: Here we present our metric embedding results. Our embeddings will be into ultrametrics, which are structured type of trees having the strong triangle inequality (see Definition 1). Our first contribution is an improved Ramsey-type hop-constrained embedding into trees, partially solving Question 1. Embedding $f:V(G)\to U$ is said to have Ramsey hop-distortion (t,M,β,h) if $\forall u,v\in V$ $d_G^{(\beta\cdot h)}(u,v)\le d_U(u,v)$, and $\forall u\in M,v\in V$, $d_U(u,v)\le t\cdot d_G^{(h)}(u,v)$ (see Definition 4).

Theorem 1 (Hop Constrained Ramsey Embedding)**.** *Consider an n-vertex graph $G=(V,E,w)$ with polynomial aspect ratio, and parameters $k,h\in[n]$. Then there is a distribution $\mathcal{D}$ over dominating ultrametrics, such that:*

1) *Every $U\in\mathrm{supp}(\mathcal{D})$, has Ramsey hop-distortion $(O(k),M,O(k\cdot\log n),h)$, where $M\subseteq V$ is a random variable.*
2) *For every $v\in V$, $\Pr_{U\sim\mathcal{D}}[v\in M]\ge\Omega(n^{-\frac{1}{k}})$.*

In addition, for every $\epsilon\in(0,1)$, there is distribution $\mathcal{D}$ as above such that every $U\in\mathrm{supp}(\mathcal{D})$ has Ramsey hop-distortion $(O(\frac{\log n}{\epsilon}),M,O(\frac{\log^2 n}{\epsilon}),h)$, and for every $v\in V$, $\Pr[v\in M]\ge 1-\epsilon$.

Ignoring the hop-distortion (e.g. setting $h=n$), the tradeoff in Theorem 1 between the distortion to the inclusion probability is asymptotically tight (see [20], [49]). However, it is yet unclear what is the best possible hop-stretch obtainable with asymptotically optimal distortion-inclusion probability tradeoff. Interestingly, we show that by increasing the distortion by a $\log\log n$ factor, one can obtain sub-logarithmic hop-stretch. Further, in this version we can also drop the polynomial aspect ratio assumption.

Theorem 2 (Hop Constrained Ramsey Embedding)**.** *Consider an n-vertex graph $G=(V,E,w)$, and parameters $k,h\in[n]$. Then there is a distribution $\mathcal{D}$ over dominating ultrametrics with V as leafs, such that every $U\in\mathrm{supp}(\mathcal{D})$ has Ramsey hop-distortion $(O(k\cdot\log\log n),M,O(k\cdot\log\log n),h)$, where $M\subseteq V$ is a random variable, and $\forall v\in V$, $\Pr_{U\sim\mathcal{D}}[v\in M_U]\ge\Omega(n^{-\frac{1}{k}})$*
In addition, for every $\epsilon\in(0,1)$, there is distribution $\mathcal{D}$ as above such that every $U\in\mathrm{supp}(\mathcal{D})$ has Ramsey hop-distortion $(O(\frac{\log n\cdot\log\log n}{\epsilon}),M,O(\frac{\log n\cdot\log\log n}{\epsilon}),h)$, and $\forall v\in V$, $\Pr[v\in M]\ge 1-\epsilon$.

Next, we construct hop-constrained clan embedding (f,χ). Embedding $f:V(G)\to 2^U$ is said to have hop-distortion (t,β,h) if $\forall u,v\in V$, $d_G^{(\beta\cdot h)}(x,y)\le\min_{u'\in f(u),v'\in f(v)}d_U(u',v')\le\min_{u'\in f(u)}d_U(u',\chi(v))\le t\cdot d_G^{(h)}(u,v)$. See Definition 3 for formal definition, and Definition 7 of hop-path-distortion.

Theorem 3 (Clan embedding into ultrametric)**.** *Consider an n-vertex graph $G=(V,E,w)$ with polynomial aspect ratio, and parameters $k,h\in[n]$. Then there is a distribution $\mathcal{D}$ over clan embeddings (f,χ) into ultrametrics with hop-distortion $(O(k),O(k\cdot\log n),h)$, hop-path-distortion $(O(k\cdot\log n),h)$, and such that for every vertex $v\in V$, $\mathbb{E}_{f\sim\mathcal{D}}[|f(v)|]\le O(n^{\frac{1}{k}})$.*

In addition, for every $\epsilon\in(0,1)$, there is distribution $\mathcal{D}$ as above such that every $U\in\mathrm{supp}(\mathcal{D})$ has hop-distortion $\left(O(\frac{\log n}{\epsilon}),O(\frac{\log^2 n}{\epsilon}),h\right)$, hop-path-distortion $\left(O(\frac{\log^2 n}{\epsilon}),h\right)$, and such that for every vertex $v\in V$, $\mathbb{E}_{f\sim\mathcal{D}}[|f(v)|]\le 1+\epsilon$.

[4] In fact, at the time of writing, the search "hop-constrained" + routing in google scholar outputs 878 papers.

Similarly to the Ramsey case, the tradeoff in Theorem 3 between the distortion to the expected clan size is asymptotically optimal [49]. However, it is yet unclear what is the best possible hop-stretch obtainable with asymptotically optimal distortion-expected clan size tradeoff. Here as well we show that by increasing the distortion by a $\log\log n$ factor, one can obtain sub-logarithmic hop-stretch. The polynomial aspect ratio assumption is dropped as well.

Theorem 4 (Clan embedding into ultrametric)**.** *Consider an n-vertex graph $G = (V, E, w)$, and parameters $k, h \in [n]$. Then there is a distribution $\mathcal{D}$ over clan embeddings (f, χ) into ultrametrics with hop-distortion $(O(k \cdot \log\log n), O(k \cdot \log\log n), h)$, hop-path-distortion $(O(k \cdot \log n \cdot \log\log n), h)$, and such that for every point $v \in V$, $\mathbb{E}_{f\sim\mathcal{D}}[|f(v)|] \le O(n^{\frac{1}{k}})$.*

In addition, for every $\epsilon \in (0, 1)$, there is distribution $\mathcal{D}$ as above such that every $U \in \operatorname{supp}(\mathcal{D})$ has hop-distortion $\left(O(\frac{\log n \cdot \log\log n}{\epsilon}), O(\frac{\log n \cdot \log\log n}{\epsilon}), h\right)$, hop-path-distortion $\left(O(\frac{\log^2 n \cdot \log\log n}{\epsilon}), h\right)$, and such that for every vertex $v \in V$, $\mathbb{E}_{f\sim\mathcal{D}}[|f(v)|] \le 1 + \epsilon$.

2) Approximation algorithms: [62] used their Ramsey type embeddings to obtain many different approximation algorithms. Our improved Ramsey embeddings imply directly improved approximation factors, and hop-stretch, for all these problems. Most prominently, we improve the cost approximation for the well studied hop constrained k-Steiner tree problem [5] over [68] ([68] approximation is superior to [62]). See [47] for a summary.

We go beyond [62] and apply our hop-constrained clan embeddings to construct better approximation algorithms. A subgraph H of G is h-respecting if for every $u, v \in V(H)$, $d_G^{(h)}(u, v) \le d_H(u, v)$. In the full version [47] we show that there is a one-to-many embedding $f : V(G) \to 2^T$ into a tree, such that for every connected h-respecting subgraph H, there is a connected subgraph H' of T of weight $O(\log n) \cdot w(H)$ and hop diameter $O(\log^2 n) \cdot h$ containing at least one vertex from $f(u)$ for every $u \in V(H)$ (alternatively, we also guarantee a similar subgraph of weight $\tilde{O}(\log n) \cdot w(H)$ and hop diameter $\tilde{O}(\log n) \cdot h$). We use this guarantee to construct bicriteria approximation algorithm for the hop-constrained group Steiner tree problem, the cost of which matches the state of the art for the (non hop-constrained) group Steiner tree problem, thus answering Question 2. Later, for the online hop-constrained group Steiner tree problem, we obtain competitive ratio that matches the competitive ratio non constrained version (see [47]).

Theorem 5. *There is a poly-time algorithm which given an instance of the h-hop-constrained group Steiner tree problem with k groups, returns a subgraph H such that $w(H) \le O(\log^2 n \cdot \log k) \cdot \mathrm{OPT}$, and $\mathrm{hop}_H(r, g_i) \le O(\log^2 n) \cdot h$ for every g_i, where OPT denotes the weight of the optimal solution.*

Alternatively, one can return a subgraph H such that $w(H) \le \tilde{O}(\log^2 n \cdot \log k) \cdot \mathrm{OPT}$, and $\mathrm{hop}_H(r, g_i) \le \tilde{O}(\log n) \cdot h$ for every g_i.

3) Hop-constrained metric data structures: In metric data structures, our goal is to construct data structure that will store (estimated) metric distances compactly, and answer distance (or routing) queries efficiently. As the distances returned by such data structure, do not have to respect the triangle inequality, one might hope to avoid any hop-stretch. This is impossible in general. Consider a complete graph with edge weights sampled randomly from $\{1, \alpha\}$ for some large α. Clearly, from information theoretic considerations, in order to estimate $d_G^{(1)}$ with arbitrarily large distortion (but smaller than α), one must use $\Omega(n^2)$ space. Therefore, in our metric data structures we will allow hop-stretch.

Distance Oracles.: A distance oracle is a succinct data structure DO that (approximately) answers distance queries. Chechik [34] (improving over previous results [87], [82], [76], [90], [33]) showed that any metric (or graph) with n points has a distance oracle of size $O(n^{1+\frac{1}{k}})$,[6] that can report any distance in $O(1)$ time with stretch at most $2k - 1$ (which is asymptotically optimal assuming Erdős girth conjecture). That is on query u, v, the answer $\mathsf{DO}(u, v)$ satisfies $d_G(u, v) \le \mathsf{DO}(u, v) \le (2k - 1) \cdot d_G(u, v)$.

Here we introduce the study of hop-constrained distance oracles. That is, given a parameter h, we would to construct a distance oracle that on query u, v, will return $d^{(h)}(u, v)$, or some approximation of it. Our result is the following:

Theorem 6 (Hop constrained Distance Oracle)**.** *For every weighted graph $G = (V, E, w)$ on n vertices with polynomial aspect ratio, and parameters $k \in \mathbb{N}$, $\epsilon \in (0, 1)$, $h \in \mathbb{N}$, there is an efficiently constructible distance oracle DO of size $O(n^{1+\frac{1}{k}} \cdot \log n)$, that for every distance query (u, v), in $O(1)$ time returns a value $\mathsf{DO}(u, v)$ such that $d_G^{(O(\frac{k\cdot\log\log n}{\epsilon})\cdot h)}(u, v) \le \mathsf{DO}(u, v) \le (2k - 1)(1 + \epsilon) \cdot d_G^{(h)}(u, v)$.*

Distance Labeling.: A *distance labeling* is a distributed version of a distance oracle. Given a graph

[5] In the k-Steiner tree problem we are given a root $r \in V$, and a set $K \subseteq V$ of at least k terminals. The goal is to find a connected subgraph H spanning the root r, and at least k terminals, of minimal weight. In the hop constrained version we are additionally required to guarantee that the hop diameter of H will be at most h.

[6] We measure size in machine words, each word is $\Theta(\log n)$ bits.

$G = (V, E, w)$, each vertex v is assigned a label $\ell(v)$, and there is an algorithm $\mathcal{A}$, that given $\ell(u), \ell(v)$ returns a value $\mathcal{A}(\ell(u), \ell(v))$ approximating $d_G(u, v)$. In their celebrated work, Thorup and Zwick [87] constructed a distance labeling scheme with labels of $O(n^{\frac{1}{k}} \cdot \log n)$ size [6], and such that in $O(k)$ time, $\mathcal{A}(\ell(u), \ell(v))$ approximates $d_G(u, v)$ within a $2k-1$ factor. We refer to [48] for further details on distance labeling (see also [78], [54], [42]).

Matoušek [75] showed that every metric space (X, d) could be embedded into ℓ_∞ of dimension $O(n^{\frac{1}{k}} \cdot k \cdot \log n)$ with distortion $2k-1$ (for the case $k = O(\log n)$, [1] later improved the dimension to $O(\log n)$). As was previously observed, this embedding can serve as a labeling scheme (where the label of each vertex will be the representing vector in the embedding). From this point of view, the main contribution of [87] is the small $O(k)$ query time (as the label size/distortion tradeoff is similar).

Here we introduce the study of hop-constrained labeling schemes, where the goal is to approximate $d_G^{(h)}(u, v)$. Note that $d_G^{(h)}$ is not a metric function, and in particular does not embed into ℓ_∞ with bounded distortion. Hence there is no trivial labeling scheme which is embedding based. Our contribution is the following:

Theorem 7 (Hop constrained Distance Labeling)**.** *For every weighted graph $G = (V, E, w)$ on n vertices with polynomial aspect ratio, and parameters $k \in \mathbb{N}$, $\epsilon \in (0, 1)$, $h \in \mathbb{N}$, there is an efficient construction of a distance labeling that assigns each node v a label $\ell(v)$ of size $O(n^{\frac{1}{k}} \cdot \log^2 n)$, and such that there is an algorithm $\mathcal{A}$ that on input $\ell(u), \ell(v)$, in $O(k)$ time returns a value such that $d_G^{(O(\frac{k \cdot \log\log n}{\epsilon}) \cdot h)}(v, u) \le \mathcal{A}(\ell(v), \ell(u)) \le (2k-1)(1+\epsilon) \cdot d_G^{(h)}(v, u)$.*

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. For fixed k, the state of the art is by Chechik [32] (improving previous works [79], [12], [13], [37], [41], [86]) who obtain stretch $3.68k$ while using $O(k \cdot n^{\frac{1}{k}})$ size [6] tables and labels of size $O(k \cdot \log n)$. For stretch $\Omega(\log n)$ further improvements were obtained by Abraham *et al.* [2], and the Author and Le [49].

Here we introduce the study of hop-constrained routing schemes, where the goal is to route the packet in a short path with small number of hops. Specially, for parameter h, we are interested in a routing scheme that for every u, v will use a route with at most $\beta \cdot h$ hops, and weight $\le t \cdot d_G^{(h)}(u, v)$, for some parameters β, t. If $\mathrm{hop}_G(u, v) \ge h$, the routing scheme is not required to do anything, and any outcome will be accepted. We answer Question 3 in the following:

Theorem 8 (Hop constrained Compact Routing Scheme)**.** *For every weighted graph $G = (V, E, w)$ on n vertices with polynomial aspect ratio, and parameters $k \in \mathbb{N}$, $\epsilon \in (0, 1)$, $h \in \mathbb{N}$, there is an efficient construction of a compact routing scheme that assigns each node a table of size $O(n^{\frac{1}{k}} \cdot k \cdot \log n)$, label of size $O(k \cdot \log^2 n)$, and such that routing a packet from u to v, will be done using a path P such that $\mathrm{hop}(P) = O(\frac{k^2 \cdot \log\log n}{\epsilon}) \cdot h$, and $w(P) \le 3.68k \cdot (1+\epsilon) \cdot d_G^{(h)}(u, v)$.* ‘

B. Related work

Hop-constrained network design problems, and in particular routing, received considerable attention from the operations research community, see [89], [57], [88], [60], [58], [5], [83], [29], [26], [28], [85], [38], [72], [27], [39] for a sample of papers.

Approximation algorithm for hop-constrained problems were previously constructed. They are usually considerably harder than their non hop-constrained counter parts, and often require for bicriteria approximation. Previously studied problems include minimum depth spanning tree [8], degree bounded minimum diameter spanning tree [69], bounded depth Steiner tree [70], [66], hop-constrained MST [81], Steiner tree [74], and k-Steiner tree [63], [68]. We refer to [62] for further details.

Recently, Ghaffari *et al.* [55] obtained a hop-constrained $\mathrm{poly}(\log)$ competitive oblivious routing scheme using techniques based on [62] hop-constrained Ramsey trees. Notably, even though the $O(\frac{\log n}{\epsilon})$ distortion in Theorem 1 is superior to the $O(\frac{\log^2 n}{\epsilon})$ distortion in a similar theorem in [62], [62] showed that the expected distortion in their embedding is only $\tilde{O}(\log \frac{n}{\epsilon})$. This additional property turned out to be important for the oblivious routing application.

Given a graph $G = (V, E, w)$, a *hop-set* is a set of edges G' that when added to G, $d_{G \cup G'}^{(h)}$ well approximates d_G. Formally, an (t, h)-hop-set is a set G' such that $\forall u, v \in V$,

$d_G(u,v) \le d^{(h)}_{G\cup G'}(u,v) \le t \cdot d_G(u,v)$. Most notably, [43] constructed $\left(1+\epsilon, \left(\frac{\log k}{\epsilon}\right)^{\log k - 2}\right)$ hop-sets with $O_{\epsilon,k}(n^{1+\frac{1}{k}})$ edges. Hop-sets were extensively studied, we refer to [44] for a survey. Another related problem is h-hop t-spanners. Here we are given a metric space (X,d), and the goal is to construct a graph G over X such that $\forall x,y \in X$, $d_X(x,y) \le d^{(h)}_G(x,y) \le t \cdot d_X(x,y)$. See [10], [84], [65] for Euclidean metrics, [51] for different metric spaces, [50] for reliable 2-hop spanners, and [9] for low-hop emulators (where $d_G^{(O(\log\log n))}$ respects the triangle inequality).

The idea of one-to-many embedding of graphs was originated by Bartal and Mendel [24], who for $k \ge 1$ constructed embedding into ultrametric with $O(n^{1+\frac{1}{k}})$ nodes and path distortion $O(k \cdot \log n \cdot \log\log n)$. The path distortion was later improved to $O(k \cdot \log n)$ [49], [19]. Recently, Haeupler *et al.* [61] studied approximate copy tree embedding which is essentially equivalent to one-to-many tree embeddings. Their "path-distortion" is inferior: $O(\log^2 n)$, however they were able to bound the number of copies of each vertex by $O(\log n)$ in the worst case (not obtained by previous works). One-to-many embeddings were also studied in the context of minor-free graphs, where Cohen-Addad *et al.* [36] constructed embedding into low treewidth graphs with expected additive distortion. Later, the author and Le [49] also constructed clan, and Ramsey type, embeddings of minor free graph into low treewidth graphs.

Bartal *et al.* [21] showed that even when the metric space is the shortest path metric of a planar graph with constant doubling dimension, the general metric Ramsey type embedding [76] cannot be substantially improved. Finally, there are also versions of Ramsey type embeddings [2], and clan embeddings [49] into spanning trees of the input graph. This embeddings loses a $\log\log n$ factor in the distortion compared with the embeddings into (non-spanning) trees.

C. Paper overview

The paper overview uses terminology presented in the preliminaries Section II.

1) Ramsey type embeddings:

Previous approach.: The construction of Ramsey type embeddings in [62] is based on padded decompositions such as in [16], [46]. Specifically, they show that for every parameter Δ, one can partition the vertices into clusters $\mathcal{C}$ such that for every $C \in \mathcal{C}$, $\mathrm{diam}^{(O(\log^2 n)\cdot h)}(C) = \max_{u,v\in C} d_G^{(O(\log^2 n)\cdot h)}(u,v) \le \Delta$ and every ball $B^{(h)}(v,r) = \{u \mid d^{(h)}_G(u,v) \le r\}$, for $r = \Omega(\frac{\Delta}{\log^2 n})$, belongs to a single cluster (i.e. v is "padded") with probability $\approx 1 - \frac{1}{\log n}$. [62] then recursively partitions the graph to create clusters with geometrically decreasing diameter Δ. The resulting hierarchical partition defines an ultrametric, where the vertices that padded in all the levels belong to the set M. By union bound each vertex belong to M with probability at least $\frac{1}{2}$, and the distortion and hop-stretch equal to the parameters of the decomposition: $O(\log^2 n)$.

Theorem 1: optimal distortion - inclusion probability tradeoff.: We use a more sophisticated approach, similar to previous works on clan embeddings [49], and deterministic construction of Ramsey trees [2]. The main task is to prove a "distributional" version of Theorem 1. Specifically, given a parameter k, and a measure $\mu : X \to \mathbb{R}_{\ge 1}$, we construct a Ramsey type embedding with Ramsey hop-distortion $(16k, M, O(k\cdot\log n), h)$ such that $\mu(M) \ge \mu(X)^{1-\frac{1}{k}}$. Theorem 1 than follows using the minimax theorem. The construction of the distributional lemma is a deterministic recursive ball growing algorithm. Given a cluster X and scale parameter 2^i, we partition X into clusters $X_1, \dots, X_s$ such that each X_i is contained in a hop bounded ball $B_X^{(i\cdot O(k)\cdot h)}(v, 2^{i-2})$. Then we recursively construct a Ramsey type embedding for each cluster X_q (with scale 2^{i-1}), and combine the obtained ultrametrics using a new root with label 2^i. Note that there is a dependence between the radius of the balls, to the number of hops they are allowed. Specifically, in each level we decrease the radius by multiplicative factor of 2, and the number of hops by an additive factor of $h' = O(k)\cdot h$. We maintain a set of active vertices M. Initially all the vertices are active. An active vertex $v \in M$ will cease to be active if the ball $B_G^{(h)}(v, \Theta(\frac{2^i}{k}))$ intersects more than one cluster. The set of vertices that remain active till the end of the algorithm, e.g. vertices that were "padded" in every level, will constitute the set M. The "magic" happens in the creation of the clusters, so that to ensure that large enough portion of the vertices remain active till the end of the algorithm. On an (inaccurate and) intuitive level, inductively, at level $i-1$, the "probability" of a vertex v to be padded in all future levels is $\left(|B^{(i\cdot h')}(v, 2^{i-3})|\right)^{-\frac{1}{k}}$. By a ball growing argument, v is padded in the i'th level with "probability" $\left(\frac{|B^{((i+1)\cdot h')}(v,2^{i-2})|}{|B^{(i\cdot h')}(v,2^{i-3})|}\right)^{-\frac{1}{k}}$. Hence by the induction hypothesis, v is padded at level i and all the future levels with probability $\left(|B^{((i+1)\cdot h')}(v, 2^{i-2})|\right)^{-\frac{1}{k}}$. Note that for this cancellations to work out, we are paying an $h' = O(k\cdot h)$ additive factor in the hop-stretch in each level, as we need that the "minimum possible cluster" at the current level will contain the "maximum possible cluster" in the next one.

Theorem 2: $\log\log n$ hop-stretch.: The construction of the embedding for Theorem 2 is the same as that of

Theorem 1 in all aspects other than the cluster creation. In particular, we first prove a distributional lemma and use the minimax theorem.

The dependence between the different scale levels in the algorithm for Theorem 1 contributes an additive factor of $O(kh)$ to the hop allowance per level. Thus for polynomial aspect ratio it has $O(k\cdot\log n)$ hop-stretch. To decrease the hop-stretch, here we create clusters w.r.t. "hop diameter" at most $O(k\cdot\log\log n)\cdot h$, regardless of the current scale. As a result, it is harder to relate between the different levels (as the "maximum possible cluster" in the next level is not contained in the "minimum possible one" in the current level). To compensate for that, instead, we use a stronger per-level guarantees. In more details, at level i we create a ball cluster inside $B^{(O(k\cdot\log\log n)\cdot h)}(v,2^{i-2})$, however instead of using the ball growing argument on the entire possibilities spectrum, we use it only in a smaller "strip", loosing a $\log\log n$ factor, but obtaining that the probability of v to be padded in the current i level is $\left(\frac{|B^{(O(k\cdot\log\log n)\cdot h)}(v,2^{i-2})|}{|X|}\right)^{\frac{1}{k}}$. Inductively we assume that the "probability" of a vertex $v\in X$ to be padded in all the level is $|X|^{-\frac{1}{k}}$. As the next cluster is surely contained in $B^{(O(k\cdot\log\log n)\cdot h)}(v,2^{i-2})$, the probability that v is padded in all the levels (including i) is indeed $|X|^{-\frac{1}{k}}$.

A similar phenomena occurs in tree embeddings into spanning trees (i.e. subgraphs), where the cluster creation is "independent" for each level as well [3], [2], [49]. An additional advantage of this approach is that it holds for graphs with unbounded aspect ratio (as the hop-stretch is unrelated to the number of scales).

2) Clan embeddings: The construction of the clan embeddings Theorems 3 and 4 are similar to the Ramsey type embeddings. In particular, in both cases the key step is a distributional lemma, where the algorithms for the distributional lemmas are deterministic ball growing algorithms. The main difference, is that while in the Ramsey type embeddings we partition the graph into clusters, in the clan embedding we create a cover. Specifically, given a cluster X and scale 2^i, the created cover is a collection of clusters $X_1,\dots,X_s$ such that each X_i is contained in a hop bounded ball $B_X^{(O(i\cdot k)\cdot h)}(v,2^{i-2})$, and every ball $B_X^{(h)}(u,\Omega(\frac{2^i}{k}))$ is fully contained in some cluster. Each vertex u might belong to several clusters. In the Ramsey type embedding our goal was to minimize the measure of the "close to the boundary" vertices, as they were deleted from the set M. On the other hand, here the goal is to bound the combined multiplicity of all the vertices (w.r.t. the measure μ), which governs the clan sizes.

Path distortion and subgraph preserving embedding: An additional property of the clan embedding is path distortion. Specifically, given an (h-respecting) path $P=(v_0,v_1,\dots,v_m)$ in G, we are guaranteed that there are vertices $v'_0,\dots,v'_m$ where $v'_i\in f(v_i)$ such that $\sum_{i=0}^{m-1}d_U(v'_i,v'_{i+1})\le O(k\cdot\log n)\cdot\sum_{i=1}^{m-1}w(v_i,v_{i+1})$. The proof of the path distortion property is recursive and follow similar lines to [24]. Specifically, for a scale 2^i we iteratively partition the path vertices $v_0,v_1,\dots,v_m$ into the clusters $X_1,\dots,X_s$, while minimizing the number of "split edges": consecutive vertices assigned to different clusters. We inductively construct "copies" for each sub-path internal to a cluster X_j. The "split edges" are used as an evidence that the path P has weight comparable to the scale 2^i times the number of split edges, and thus we could afforded them (as the maximal price of an edge in the ultrametric is 2^i).

Next, we construct the subgraph preserving embedding. Specifically, we construct a clan embedding $f:V\to 2^T$, such that all the vertices in T are copies of G vertices: $V(T)=\cup_{v\in V}f(v)$, and every edge $e=\{u',v'\}\in T$ for $u'\in f(u)$, $v'\in f(v)$ is associated with a path P_e from u to v in G of weight $\le d_T(u',v')$. Concatenating edge paths we obtain an induced path for every pair of vertices u',v' in T. We construct such subgraph preserving embedding, where the number of hops in every induced path is bounded by $O(\log^2 n)\cdot h$, and such that for every connected (h-respecting) subgraph H of G, there is a connected subgraph H' of T of weight $w_T(H')\le O(\log n)\cdot w_G(H)$, where for every $v\in H$, H' contains some vertex from $f(v)$. Then we obtain an alternative tradeoff where the number of hops in every induced path is bounded by $\tilde{O}(\log n)\cdot h$, and the weight of the subgraph H' is bounded by $w_T(H')\le\tilde{O}(\log n)\cdot w_G(H)$.

3) Approximation algorithms: Following the approach of [62], our improved Ramsey type embeddings (Theorems 1 and 2) imply improved approximation algorithms for all the problems studied in [62]. We refer to [62] for details, and to the full version for summary.

For the group Steiner tree problem (and its online version), using the subgraph preserving embedding, we are able to obtain a significant improvement beyond the techniques of [62]. In fact, the cost of our approximation algorithm matches the state of the art for the non hop-restricted case. Our approach is as follows: first we observe that the optimal solution for the hop-constrained group Steiner tree problem is a tree. In particular, the subgraph H of G constituting the optimal solution is $2h$-respecting. It follows that a copy H' of the optimal solution H (of weight $O(\log n)\cdot\mathrm{OPT}$) could be found in the image T of the subgraph preserving embedding f. Next we use known (non hop-constrained) approximation algorithms for the group Steiner tree problem on trees [53] to find a valid solution H'' in T of weight $O(\log^2 n\cdot\log k)\cdot\mathrm{OPT}$. Finally, by combining the associated paths we obtain an induced solution in G of the same cost. The bound on the hops in the induced paths insures an $O(\log^2 n)$ hop-stretch. The competitive algorithm for the online group Steiner tree problem follows similar lines.

4) Hop constrained metric data structures: At first glance, as hop constrained distances are non-metric, constructing metric data structures for them seems to be very challenging, and indeed, the lack of any previous work on this very natural problem serves as an evidence.

Initially, using our Ramsey type embedding we construct distance labelings in a similar fashion to the distance labeling scheme of [76]: construct $O(n^{\frac{1}{k}} \cdot k)$ Ramsey type embedding into ultrametrics $U_1, U_2, \dots$ such that U_i is ultrametric over V with the set M_i of saved vertices. It will hold that every vertex belongs to some set M_i. For trees (and ultrametrics) there are very efficient distance labeling schemes (constant size and query time, and 1.5-distortion [52]). We define the label of each vertex v to be the union of the labels for the ultrametrics, and the index i of the ultrametric where $v \in M_i$. As a result we obtain a distance labeling with label size $O(n^{\frac{1}{k}} \cdot k)$, constant query time, and such that for every u, v, $d_G^{(O(k \cdot \log\log n)\cdot h)}(u,v) \le \mathcal{A}(\ell(v), \ell(u)) \le O(k \cdot \log\log n) \cdot d_G^{(h)}(u,v)$. While the space, query time, and hop-stretch are quite satisfactory, the distortion leaves much to be desired.

To obtain an improvement, our next step is to construct a data structures for a fixed scale. Specifically, for every scale 2^i, we construct a graph G_i such that for every pair u, v where $d_G^{(h)}(u,v) \approx 2^i$, it holds that $d_G^{(O(\frac{k \cdot \log\log n}{\epsilon})\cdot h)}(u,v) \le d_{G_i}(u,v) \le (1+\epsilon) \cdot d_G^{(h)}(u,v)$. Thus, if we were to know that $d_G^{(h)}(u,v) \approx 2^i$, then we could simply use state of the art distance labeling for G_i, ignoring hop constrains. Now, using our first Ramsey based distance labeling scheme we can obtain the required coarse approximation! Our final construction consist of Ramsey based distance labeling, and in addition state of the art distance labeling [87] for the graphs G_i for all possible distance scales. Given a query u, v, we first coarsely approximate $d_G^{(h)}(u,v)$ to obtain some value 2^i. Then we returned the answer from the distance labeling scheme that was prepared in advance for G_i. Our approach for distance oracles, and compact routing scheme is similar.

II. Preliminaries

$\tilde{O}$ notation hides poly-logarithmic factors, that is $\tilde{O}(g) = O(g) \cdot \mathrm{polylog}(g)$. Consider an undirected weighted graph $G = (V, E, w)$, and hop parameter h. A graph is called unweighted if all its edges have unit weight. We denote G's vertex set and edge set by $V(G)$ and $E(G)$, respectively. Often we will abuse notation and wright G instead of $V(G)$. d_G denotes the shortest path metric in G, i.e., $d_G(u,v)$ equals to the minimal weight of a path between u to v. A path $v_0, v_1, \dots, v_h$ is said to have h hops. The h-hop distance between two vertices u, v is

$$d_G^{(h)}(u,v) = \min\{w(P) \mid P \text{ is a path between } u \text{ and } v \text{ with } \mathrm{hop}(P) \le h\} .$$

If there is no path from u to v with at most h hops, then $d_G^{(h)}(u,v) = \infty$. $\mathrm{hop}_G(u,v) = \min\left\{h \mid d_G^{(h)}(u,v) < \infty\right\}$ is the minimal number of hops in a u-v path. The h-hop diameter of the graph is $\mathrm{diam}^{(h)}(G) = \max\{d_G^{(h)}(u,v)\}$ the maximal h-hop distance between a pair of vertices. Note that in many cases, this will be ∞. $B_G(v,r) = \{u \mid d_G(u,v) \le r\}$ is the closed ball around u of radius r, and similarly $B_G^{(h)}(v,r) = \{u \mid d_G^{(h)}(u,v) \le r\}$ is the h-hop constrained ball.

Throughout the paper we will assume that the minimal weight edge in the input graph has weight 1, note that due to scaling this assumption do not lose generality. The *aspect ratio*, [3] is the ratio between the maximal to minimal weight in G $\frac{\max_{e\in E} w(e)}{\min_{e\in E} w(e)}$, or simply $\max_{e\in E} w(e)$ as we assumed the the minimal distance is 1. In many places (similarly to [62]) we will assume that the aspect ratio is polynomial in n (actually in all results other than Theorems 2 and 4). This will always be stated explicitly.

For a subset $A \subseteq V$ of vertices, $G[A] = (A, E_A = E \cap \binom{V}{2}, w_{\restriction E_A})$ denotes the subgraph induced by A. The diameter of A, denoted by $\mathrm{diam}(S)$ is $\max_{u,v\in S} d_{G[S]}(u,v)$. [7] Similarly, for a subset of edges $\tilde{E} \subseteq E$, where $A \subseteq V$ is the subset of vertices in $\cup_{e\in\tilde{E}} e$, the graph induced by $\tilde{E}$ is $G[\tilde{E}] = (A, \tilde{E}, w_{\restriction\tilde{E}})$.

An ultrametric (Z, d) is a metric space satisfying a strong form of the triangle inequality, that is, for all $x, y, z \in Z$, $d(x,z) \le \max\{d(x,y), d(y,z)\}$. The following definition is known to be an equivalent one (see [23]).

Definition 1. *An ultrametric is a metric space (Z,d) whose elements are the leaves of a rooted labeled tree T. Each $z \in T$ is associated with a label $\ell(z) \ge 0$ such that if $q \in T$ is a descendant of z then $\ell(q) \le \ell(z)$ and $\ell(q) = 0$ iff q is a leaf. The distance between leaves $z, q \in Z$ is defined as $d_T(z,q) = \ell(lca(z,q))$ where $lca(z,q)$ is the least common ancestor of z and q in T.*

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

[7] This is often called *strong* diameter. A related notion is the *weak* diameter of a cluster S, defined $\max_{u,v\in S} d_G(u,v)$. Note that for a metric space, weak and strong diameter are equivalent. See [46].

metric embedding f is said to be *dominating* if for every pair of points $x, y \in X$, it holds that $d_X(x,y) \le d_Y(f(x), f(y))$. The distortion of a dominating embedding f is $\max_{x \ne y \in X} \frac{d_Y(f(x),f(y))}{d_X(x,y)}$. Here we will also study a more permitting generalization of metric embedding introduced by Cohen-Addad *et al.* [36], which is called *one-to-many* embedding.

Definition 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 $d_X(x,y) \le \min_{x' \in f(x), y' \in f(y)} d_Y(x', y')$.*

Here we will study the new notion of clan embeddings introduced by the author and Le [49].

Definition 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, where for every $x \in X$, $\chi(x) \in f(x)$. $\chi(x)$ is referred to as the chief of the clan of x (or simply the chief of x). We say that the clan embedding f has distortion t if for every $x, y \in X$, $\min_{y' \in f(y)} d_Y(y', \chi(x)) \le t \cdot d_X(x,y)$.*

Bounded hop distances: We say that an embedding f has distortion t, and *hop-stretch* β, if for every $u, v \in V$ it holds that

$$d_G^{(\beta h)}(u,v) \le d_X(f(u), f(v)) \le t \cdot d_G^{(h)}(u,v) \ .$$

[62] showed an example of a graph G where in every classic embedding of G into a metric space, either the hop-stretch β, or the distortion t must be polynomial in n. In particular, there is no hope for a classic embedding (in particular stochastic) with both hop-stretch and distortion being sub-polynomial.

We will therefore study hop-constrained clan embeddings, and Ramsey type embeddings.

Definition 4 (Hop-distortion of Ramsey type embedding)**.** *An embedding f from a weighted graph $G = (V, E, w)$ to metric space (X, d_X) has Ramsey hop distortion (t, M, β, h) if $M \subseteq V$, for every $u, v \in V$ it holds that $d_G^{(\beta \cdot h)}(u,v) \le d_X(f(v), f(u))$, and for every $u \in V$ and $v \in M$, $d_X(f(v), f(u)) \le t \cdot d_G^{(h)}(u,v)$.*

Definition 5 (Hop-distortion of clan embedding)**.** *A clan embedding (f, χ) from a weighted graph $G = (V, E, w)$ to a metric space (X, d_X) has hop distortion (t, β, h) if for every $u, v \in V$, it holds that $d_G^{(\beta \cdot h)}(u,v) \le \min_{v' \in f(v), u' \in f(u)} d_U(v', u')$, and $\min_{v' \in f(v)} d_U(v', \chi(u)) \le t \cdot d_G^{(h)}(u,v)$.*

The following definitions will be useful to argue that the clan embedding preservers properties of subgraphs, and not only vertex pairs.

Definition 6 (h-respecting)**.** *A subgraph H of G is h-*respecting *if for every $u, v \in H$ it holds that $d_G^{(h)}(u,v) \le d_H(u,v)$. Often we will abuse notation and say that a path P is h-*respecting*, meaning that the subgraph induced by the path is h-respecting.*

Definition 7 (Hop-path-distortion)**.** *We say that the one-to-many embedding f has* hop path distortion (t, h) *if for every h-respecting path $P = (v_0, v_1, \dots, v_m)$ there are vertices $v'_0, \dots, v'_m$ where $v'_i \in f(v_i)$ such that $\sum_{i=0}^{m-1} d_Y(v'_i, v'_{i+1}) \le t \cdot \sum_{i=0}^{m-1} w(v_i, v_{i+1})$.*

References

[1] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. *Advances in Mathematics*, 228(6):3026 – 3126, 2011.

[2] Ittai Abraham, Shiri Chechik, Michael Elkin, Arnold Filtser, and Ofer Neiman. Ramsey spanning trees and their applications. *ACM Trans. Algorithms*, 16(2):19:1–19:21, 2020.

[3] Ittai Abraham and Ofer Neiman. Using petal-decompositions to build a low stretch spanning tree. *SIAM J. Comput.*, 48(2):227–248, 2019.

[4] Ajit Agrawal, Philip N. Klein, and R. Ravi. When trees collide: An approximation algorithm for the generalized steiner problem on networks. *SIAM J. Comput.*, 24(3):440–456, 1995.

[5] Ibrahim Akgün and Barbaros Ç. Tansel. New formulations of the hop-constrained minimum spanning tree problem via miller-tucker-zemlin constraints. *Eur. J. Oper. Res.*, 212(2):263–276, 2011.

[6] Noga Alon, Richard M. Karp, David Peleg, and Douglas B. West. A graph-theoretic game and its application to the k-server problem. *SIAM J. Comput.*, 24(1):78–100, 1995.

[7] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *J. Comput. Syst. Sci.*, 58(1):137–147, 1999.

[8] Ernst Althaus, Stefan Funke, Sariel Har-Peled, Jochen Könemann, Edgar A. Ramos, and Martin Skutella. Approximating k-hop minimum-spanning trees. *Oper. Res. Lett.*, 33(2):115–120, 2005.

[9] Alexandr Andoni, Clifford Stein, and Peilin Zhong. Parallel approximate undirected shortest paths via low hop emulators. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020*, pages 322–335, 2020.

[10] Sunil Arya, David M. Mount, and Michiel H. M. Smid. Randomized and deterministic algorithms for geometric spanners of small diameter. In *35th Annual Symposium on Foundations of Computer Science, Santa Fe, 1994*, pages 703–712, 1994.

[11] Vassilis Athitsos and Stan Sclaroff. Database indexing methods for 3d hand pose estimation. In *Gesture-Based Communication in Human-Computer Interaction, 2003*, pages 288–299, 2003.

[12] Baruch Awerbuch, Amotz Bar-Noy, Nathan Linial, and David Peleg. Improved routing strategies with succinct tables. *J. Algorithms*, 11(3):307–341, 1990.

[13] Baruch Awerbuch and David Peleg. Routing with polynomial communication-space trade-off. *SIAM J. Discret. Math.*, 5(2):151–162, 1992.

[14] Anantaram Balakrishnan and Kemal Altinkemer. Using a hop-constrained model to generate alternative communication network design. *INFORMS J. Comput.*, 4(2):192–205, 1992.

[15] Judit Bar-Ilan, Guy Kortsarz, and David Peleg. Generalized submodular cover problems and applications. *Theor. Comput. Sci.*, 250(1-2):179–200, 2001.

[16] Yair Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science, FOCS 1996*, pages 184–193, 1996.

[17] Yair Bartal. On approximating arbitrary metrices by tree metrics. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, 1998*, pages 161–168, 1998.

[18] Yair Bartal. 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*, pages 89–97, 2004.

[19] Yair Bartal. Advances in metric ramsey theory and its applications. *CoRR*, abs/2104.03484, 2021.

[20] Yair Bartal, Béla Bollobás, and Manor Mendel. Ramsey-type theorems for metric spaces with applications to online problems. *J. Comput. Syst. Sci.*, 72(5):890–921, 2006.

[21] Yair Bartal, Nova Fandina, and Ofer Neiman. Covering metric spaces by few trees. In *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece*, pages 20:1–20:16, 2019.

[22] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric Ramsey-type dichotomies. *Journal of the London Mathematical Society*, 71(2):289–303, 2005.

[23] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. Some low distortion metric ramsey problems. *Discret. Comput. Geom.*, 33(1):27–41, 2005.

[24] Yair Bartal and Manor Mendel. Multiembedding of metric spaces. *SIAM J. Comput.*, 34(1):248–259, 2004.

[25] Guy E. Blelloch, Yan Gu, and Yihan Sun. A new efficient construction on probabilistic tree embeddings. *CoRR*, abs/1605.04651, 2016. https://arxiv.org/abs/1605.04651.

[26] Andreas Bley, S. Mehdi Hashemi, and Mohsen Rezapour. IP modeling of the survivable hop constrained connected facility location problem. *Electron. Notes Discret. Math.*, 41:463–470, 2013.

[27] Jérôme De Boeck and Bernard Fortz. Extended formulation for hop constrained distribution network configuration problems. *Eur. J. Oper. Res.*, 265(2):488–502, 2018.

[28] Quentin Botton, Bernard Fortz, and Luis Eduardo Neves Gouveia. On the hop-constrained survivable network design problem with reliable edges. *Comput. Oper. Res.*, 64:159–167, 2015.

[29] Quentin Botton, Bernard Fortz, Luis Eduardo Neves Gouveia, and Michael Poss. Benders decomposition for the hop-constrained survivable network design problem. *INFORMS J. Comput.*, 25(1):13–26, 2013.

[30] J. Bourgain, T. Figiel, and V. Milman. On Hilbertian subsets of finite metric spaces. *Israel J. Math.*, 55(2):147–152, 1986.

[31] Sébastien Bubeck, Michael B. Cohen, Yin Tat Lee, James R. Lee, and Aleksander Madry. k-server via multiscale entropic regularization. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018*, pages 3–16, 2018.

[32] Shiri Chechik. Compact routing schemes with improved stretch. In *ACM Symposium on Principles of Distributed Computing, PODC '13, Montreal, QC, Canada, July 22-24, 2013*, pages 33–41, 2013.

[33] Shiri Chechik. Approximate distance oracles with constant query time. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 654–663, New York, NY, USA, 2014. ACM.

[34] Shiri Chechik. Approximate distance oracles with improved bounds. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015*, pages 1–10, 2015.

[35] Edith Cohen. Polylog-time and near-linear work approximation scheme for undirected shortest paths. *J. ACM*, 47(1):132–166, 2000.

[36] Vincent Cohen-Addad, Arnold Filtser, Philip N. Klein, and Hung Le. On light spanners, low-treewidth embeddings and efficient traversing in minor-free graphs. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020*, pages 589–600, 2020.

[37] Lenore Cowen. Compact routing with minimum stretch. *J. Algorithms*, 38(1):170–183, 2001.

[38] I. Diarrassouba, Virginie Gabrel, Ali Ridha Mahjoub, Luis Eduardo Neves Gouveia, and Pierre Pesneau. Integer programming formulations for the k-edge-connected 3-hop-constrained network design problem. *Networks*, 67(2):148–169, 2016.

[39] Ibrahima Diarrassouba, Meriem Mahjoub, Ali Ridha Mahjoub, and Hande Yaman. k-node-disjoint hop-constrained survivable networks: polyhedral analysis and branch and cut. *Ann. des Télécommunications*, 73(1-2):5–28, 2018.

[40] Michael Dinitz, Guy Kortsarz, and Ran Raz. Label cover instances with large girth and the hardness of approximating basic k-spanner. *ACM Trans. Algorithms*, 12(2):25:1–25:16, 2016.

[41] Tamar Eilam, Cyril Gavoille, and David Peleg. Compact routing schemes with low stretch factor. *J. Algorithms*, 46(2):97–114, 2003.

[42] Michael Elkin, Arnold Filtser, and Ofer Neiman. Prioritized metric structures and embedding. *SIAM J. Comput.*, 47(3):829–858, 2018.

[43] Michael Elkin and Ofer Neiman. Hopsets with constant hopbound, and applications to approximate shortest paths. *SIAM J. Comput.*, 48(4):1436–1480, 2019.

[44] Michael Elkin and Ofer Neiman. Near-additive spanners and near-exact hopsets, A unified view. *Bull. EATCS*, 130, 2020.

[45] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. Syst. Sci.*, 69(3):485–497, November 2004.

[46] Arnold Filtser. On strong diameter padded decompositions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2019*, pages 6:1–6:21, 2019.

[47] Arnold Filtser. Hop-constrained metric embeddings and their applications. *CoRR*, abs/2106.14969, 2021.

[48] Arnold Filtser, Lee-Ad Gottlieb, and Robert Krauthgamer. Labelings vs. embeddings: On distributed representations of distances. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020*, pages 1063–1075, 2020.

[49] Arnold Filtser and Hung Le. Clan embeddings into trees, and low treewidth graphs. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 342–355, 2021.

[50] Arnold Filtser and Hung Le. Reliable spanners: Locality-sensitive orderings strike back. *CoRR*, abs/2101.07428, 2021.

[51] Arnold Filtser and Ofer Neiman. Light spanners for high dimensional norms via stochastic decompositions. In *26th Annual European Symposium on Algorithms, ESA 2018, August 20-22, 2018, Helsinki, Finland*, pages 29:1–29:15, 2018.

[52] Ofer Freedman, Pawel Gawrychowski, Patrick K. Nicholson, and Oren Weimann. Optimal distance labeling schemes for trees. In *Proceedings of the ACM Symposium on Principles of Distributed Computing, PODC 2017*, pages 185–194, 2017.

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

[54] Cyril Gavoille, David Peleg, Stéphane Pérennes, and Ran Raz. Distance labeling in graphs. *J. Algorithms*, 53(1):85–112, 2004.

[55] Mohsen Ghaffari, Bernhard Haeupler, and Goran Zuzic. Hop-constrained oblivious routing. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1208–1220, 2021.

[56] Lee-Ad Gottlieb, Aryeh Kontorovich, and Robert Krauthgamer. Efficient regression in metric spaces via approximate lipschitz extension. *IEEE Trans. Inf. Theory*, 63(8):4838–4849, 2017.

[57] Luis Gouveia. Multicommodity flow models for spanning trees with hop constraints. *European Journal of Operational Research*, 95(1):178–190, 1996.

[58] Luis Eduardo Neves Gouveia and Thomas L. Magnanti. Network flow models for designing diameter-constrained minimum-spanning and steiner trees. *Networks*, 41(3):159–173, 2003.

[59] Luis Eduardo Neves Gouveia, Pedro Patrício, Amaro de Sousa, and Rui Valadas. MPLS over WDM network design with packet level qos constraints based on ILP models. In *Proceedings IEEE INFOCOM 2003, The 22nd Annual Joint Conference of the IEEE Computer and Communications Societies 2003*, pages 576–586, 2003.

[60] Luis Eduardo Neves Gouveia and Cristina Requejo. A new lagrangean relaxation approach for the hop-constrained minimum spanning tree problem. *Eur. J. Oper. Res.*, 132(3):539–552, 2001.

[61] Bernhard Haeupler, D. Ellis Hershkowitz, and Goran Zuzic. Deterministic tree embeddings with copies for algorithms against adaptive adversaries. *CoRR*, abs/2102.05168, 2021.

[62] Bernhard Haeupler, D. Ellis Hershkowitz, and Goran Zuzic. Tree embeddings for hop-constrained network design. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 356–369, 2021.

[63] Mohammad Taghi Hajiaghayi, Guy Kortsarz, and Mohammad R. Salavatipour. Approximating buy-at-bulk and shallow-light k-steiner trees. *Algorithmica*, 53(1):89–103, 2009.

[64] Eran Halperin, Jeremy Buhler, Richard M. Karp, Robert Krauthgamer, and B. Westover. Detecting protein sequence conservation via metric embeddings. *Bioinformatics*, 19(suppl 1):i122–i129, 07 2003.

[65] Sariel Har-Peled, Piotr Indyk, and Anastasios Sidiropoulos. Euclidean spanners in high dimensions. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013*, pages 804–809, 2013.

[66] Erez Kantor and David Peleg. Approximate hierarchical facility location and applications to the bounded depth steiner tree and range assignment problems. *J. Discrete Algorithms*, 7(3):341–362, 2009.

[67] Maleq Khan, Fabian Kuhn, Dahlia Malkhi, Gopal Pandurangan, and Kunal Talwar. Efficient distributed approximation algorithms via probabilistic tree embeddings. *Distributed Comput.*, 25(3):189–205, 2012.

[68] M. Reza Khani and Mohammad R. Salavatipour. Improved approximations for buy-at-bulk and shallow-light k-steiner trees and (k, 2)-subgraph. *J. Comb. Optim.*, 31(2):669–685, 2016.

[69] Jochen Könemann, Asaf Levin, and Amitabh Sinha. Approximating the degree-bounded minimum diameter spanning tree problem. *Algorithmica*, 41(2):117–129, 2005.

[70] Guy Kortsarz and David Peleg. Approximating shallow-light trees (extended abstract). In *Proceedings of the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, 1997*, pages 103–110, 1997.

[71] Larry J. LeBlanc, Jerome Chifflet, and Philippe Mahey. Packet routing in telecommunication networks with path and flow restrictions. *INFORMS J. Comput.*, 11(2):188–197, 1999.

[72] Markus Leitner. Layered graph models and exact algorithms for the generalized hop-constrained minimum spanning tree problem. *Comput. Oper. Res.*, 65:1–18, 2016.

[73] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Comb.*, 15(2):215–245, 1995.

[74] Madhav V. Marathe, R. Ravi, Ravi Sundaram, S. S. Ravi, Daniel J. Rosenkrantz, and Harry B. Hunt III. Bicriteria network design problems. *J. Algorithms*, 28(1):142–171, 1998.

[75] Jiří Matoušek. On the distortion required for embedding finite metric spaces into normed spaces. *Israel Journal of Mathematics*, 93(1):333–344, 1996.

[76] Manor Mendel and Assaf Naor. Ramsey partitions and proximity data structures. *Journal of the European Mathematical Society*, 9(2):253–275, 2007.

[77] Assaf Naor and Terence Tao. Scale-oblivious metric fragmentation and the nonlinear dvoretzky theorem. *Israel Journal of Mathematics*, 192(1):489–504, 2012.

[78] David Peleg. Proximity-preserving labeling schemes. *J. Graph Theory*, 33(3):167–176, 2000.

[79] David Peleg and Eli Upfal. A trade-off between space and efficiency for routing tables. *J. ACM*, 36(3):510–530, 1989.

[80] Hasan Pirkul and Samit Soni. New formulations and solution procedures for the hop constrained network design problem. *Eur. J. Oper. Res.*, 148(1):126–140, 2003.

[81] R. Ravi. Rapid rumor ramification: Approximating the minimum broadcast time (extended abstract). In *35th Annual Symposium on Foundations of Computer Science, 1994*, pages 202–213, 1994.

[82] Liam Roditty, Mikkel Thorup, and Uri Zwick. Deterministic constructions of approximate distance oracles and spanners. In *Proceedings of the 32Nd International Conference on Automata, Languages and Programming*, ICALP'05, pages 261–272, Berlin, Heidelberg, 2005. Springer-Verlag.

[83] André Rossi, Alexis Aubry, and Mireille Jacomino. Connectivity-and-hop-constrained design of electricity distribution networks. *Eur. J. Oper. Res.*, 218(1):48–57, 2012.

[84] Shay Solomon. Sparse euclidean spanners with tiny diameter. *ACM Trans. Algorithms*, 9(3):28:1–28:33, 2013.

[85] Babacar Thiongane, Jean-François Cordeau, and Bernard Gendron. Formulations for the nonbifurcated hop-constrained multicommodity capacitated fixed-charge network design problem. *Comput. Oper. Res.*, 53:1–8, 2015.

[86] Mikkel Thorup and Uri Zwick. Compact routing schemes. In *Proceedings of the Thirteenth Annual ACM Symposium on Parallel Algorithms and Architectures, SPAA 2001, 2001*, pages 1–10, 2001.

[87] Mikkel Thorup and Uri Zwick. Approximate distance oracles. *J. ACM*, 52(1):1–24, 2005.

[88] Stefan Voß. The steiner tree problem with hop constraints. *Ann. Oper. Res.*, 86:321–345, 1999.

[89] Kathleen A. Woolston and Susan L. Albin. The design of centralized networks with reliability and availability constraints. *Comput. Oper. Res.*, 15(3):207–217, 1988.

[90] Christian Wulff-Nilsen. Approximate distance oracles with improved query time. In *Proceedings of the Twenty-Forth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '13. SIAM, 2013.