

CUTTING CORNERS CHEAPLY, OR HOW TO REMOVE STEINER POINTS*

LIOR KAMMA[†], ROBERT KRAUTHGAMER[†], AND HUY L. NGUYỄN[‡]

Abstract. Our main result is that the Steiner point removal (SPR) problem can always be solved with polylogarithmic distortion, which answers in the affirmative a question posed by Chan, Xia, Konjevod, and Richa in 2006. Specifically, we prove that for every edge-weighted graph $G = (V, E, w)$ and a subset of terminals $T \subseteq V$, there is a graph $G' = (T, E', w')$ that is isomorphic to a minor of G such that for every two terminals $u, v \in T$, the shortest-path distances between them in G and in G' satisfy $d_{G,w}(u, v) \leq d_{G',w'}(u, v) \leq O(\log^5 |T|) \cdot d_{G,w}(u, v)$. Our existence proof actually gives a randomized polynomial-time algorithm. Our proof features a new variant of metric decomposition. It is well known that every finite metric space (X, d) admits a β -separating decomposition for $\beta = O(\log |X|)$, which means that for every $\Delta > 0$ there is a randomized partitioning of X into clusters of diameter at most Δ , satisfying the following separation property: for every $x, y \in X$, the probability that they lie in different clusters of the partition is at most $\beta d(x, y)/\Delta$. We introduce an additional requirement in the form of a tail bound: for every shortest-path P of length $d(P) \leq \Delta/\beta$, the number of clusters of the partition that meet the path P , denoted by Z_P , satisfies $\Pr[Z_P > t] \leq 2e^{-\Omega(t)}$ for all $t > 0$.

Key words. Steiner point removal, approximate distances, metric spaces

AMS subject classifications. 05C12, 05C83, 05C85, 52C99, 05C85, 68W20

DOI. 10.1137/140951382

1. Introduction. *Graph compression* describes the transformation of a given graph G into a small graph G' that preserves certain features (quantities) of G , such as distances or cut values. Notable examples for this genre include graph spanners, distance oracles, cut sparsifiers, and spectral sparsifiers; see, e.g., [PS89, TZ05, BK96, BSS09] and references therein. The algorithmic utility of such graph transformations is clear—once the “compressed” graph G' is computed as a preprocessing step, further processing can be performed on G' instead of on G , using fewer resources, like runtime and memory, or achieving better accuracy (when the solution is approximate). See section 1.3 for more details.

Within this context, we study *vertex sparsification*, where G has a designated subset of vertices T , and the goal is to reduce the number of vertices in the graph while maintaining certain properties of T . A prime example for this genre is vertex sparsifiers that preserve terminal versions of (multicommodity) cut and flow problems, a successful direction that was initiated by Moitra [Moi09] and extended in several followups [LM10, CLLM10, MM10, EGK⁺10, Chu12]. Our focus here is different, on preserving distances, a direction that was seeded by Gupta [Gup01] more than a decade ago.

*Received by the editors January 3, 2014; accepted for publication (in revised form) June 18, 2015; published electronically August 4, 2015. A preliminary version of this paper appeared in *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2014.

<http://www.siam.org/journals/sicomp/44-4/95138.html>

[†]The Weizmann Institute, Rehovot 76100, Israel (lior.kamma@weizmann.ac.il, robert.krauthgamer@weizmann.ac.il). The work of these authors was supported in part by the Israel Science Foundation (grant 97/13), the US-Israel BSF (grant 2010418), and the Citi Foundation. Part of this work was done while the second author was visiting Microsoft Research New England.

[‡]Princeton University, Princeton, NJ 08540 (hnguyen@princeton.edu). This author’s work was supported in part by NSF CCF 0832797, and a Gordon Wu Fellowship. Part of this work was done while this author was interning at Microsoft Research New England.

Throughout the paper, all graphs are undirected and all edge weights are positive.

Steiner point removal (SPR). Let $G = (V, E, w)$ be an edge-weighted graph, and let $T = \{t_1, \dots, t_k\} \subseteq V$ be a designated set of k terminals. Here and throughout, $d_{G,w}(\cdot, \cdot)$ denotes the shortest-path metric between vertices of G according to the weights w . The *SPR* problem asks one to construct on the terminals a new graph $G' = (T, E', w')$ such that (i) distances between the terminals are *distorted* at most by factor $\alpha \geq 1$, formally,

$$\forall u, v \in T, \quad d_{G,w}(u, v) \leq d_{G',w'}(u, v) \leq \alpha \cdot d_{G,w}(u, v);$$

and (ii) the graph G' is (isomorphic to) a minor of G . This formulation of the SPR problem was proposed by Chan et al. [CXKR06, section 5], who posed the problem of bounding the distortion α (existentially and/or using an efficient algorithm). Our main result is to answer their open question.

Requirement (ii) above expresses structural similarity between G and G' ; for instance, if G is planar, then so is G' . The SPR formulation above actually came about as a generalization to a result of Gupta [Gup01], which asserts that if G is a tree, then there exists a tree G' , which preserves terminal distances with distortion $\alpha = 8$. Later Chan et al. [CXKR06] observed that this same G' is actually a minor of the original tree G and proved the factor of 8 to be tight. The upper bound for trees was later extended by Basu and Gupta [BG08], who achieved distortion $\alpha = O(1)$ for the larger class of outerplanar graphs.

How to construct minors. We now describe a general methodology that is natural for the SPR problem. The first step constructs a minor G' with vertex set T , but without any edge weights, and is prescribed by Definition 1.2. The second step determines edge weights w' such that $d_{G',w'}$ dominates $d_{G,w}$ on the terminals T and is given in Definition 1.3. These steps are illustrated in Figure 1. Our definitions are actually more general (anticipating the technical sections) and consider G' , whose vertex set is sandwiched between T and V .

DEFINITION 1.1. A partial partition of a set V is a collection V_1, \dots, V_k of pairwise disjoint subsets of V , referred to as clusters.

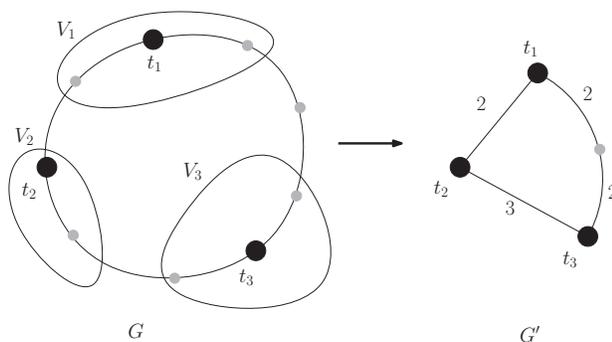


FIG. 1. The graph G , a 9-cycle with unit edge weights, is depicted on the left with three terminals and disjoint subsets V_1, V_2, V_3 . Its terminal-centered minor G' and the standard-restriction edge weights are shown on the right.

DEFINITION 1.2 (terminal-centered minor). Let $G = (V, E)$ be a graph with k terminals $T = \{t_1, \dots, t_k\}$, and let V_1, \dots, V_k be a partial partition of V , such that each induced subgraph $G[V_j]$ is connected and contains t_j . The graph $G' = (V', E')$

obtained by contracting each $G[V_j]$ into a single vertex that is identified with t_j is called the terminal-centered minor of G induced by V_1, \dots, V_k .

By identifying the “contracted supernode” V_j with t_j , we may think of the vertex set V' as containing T and (possibly) some vertices from $V \setminus T$, which implies $V' \subset V$. A terminal-centered minor G' of G can also be described by a mapping $f : V \rightarrow T \cup \{\perp\}$ such that $f|_T \equiv \text{id}$ and $f^{-1}(\{t_j\})$ is connected in G for all $j \in [k]$. Indeed, simply let $V_j = f^{-1}(\{t_j\})$ for all $j \in [k]$, and thus $V \setminus (\cup_j V_j) = f^{-1}(\{\perp\})$.

DEFINITION 1.3 (standard restriction). *Let $G = (V, E, w)$ be an edge-weighted graph with terminal set T , and let $G' = (V', E')$ be a terminal-centered minor of G . (Recall that we view $V' \subset V$.) The standard restriction of w to G' is the edge weight w' given by the respective distances in G , formally,*

$$\forall (x, y) \in E', \quad w'_{xy} := d_{G,w}(x, y).$$

This edge weight w' is optimal in the sense that $d_{G',w'}$ dominates $d_{G,w}$ (where it is defined, i.e., on V'), and the weight of each edge $(x, y) \in E'$ is minimal under this domination condition.

1.1. Main result. Our main result below gives an efficient algorithm achieving $\text{polylog}(k)$ distortion for the SPR problem. Its proof spans sections 3 and 4, though the former contains the heart of the matter.

THEOREM 1.4. *Let $G = (V, E, w)$ be an edge-weighted graph with k terminals $T \subseteq V$. Then there exists a terminal-centered minor $G' = (T, E', w')$ of G that attains distance distortion $O(\log^5 k)$, i.e.,*

$$\forall u, v \in T, \quad 1 \leq \frac{d_{G',w'}(u, v)}{d_{G,w}(u, v)} \leq O(\log^5 k).$$

Moreover, w' is the standard restriction of w , and G' is computable in randomized polynomial time.

This theorem answers a question of Chan et al. [CXKR06]. The only distortion lower bound known for general graphs is a factor of 8 (which actually holds for trees) [CXKR06], and thus it remains a challenging open question whether $O(1)$ distortion can be achieved in general graphs.

Our proof of Theorem 1.4 begins similarly to the proof of Englert et al. [EGK⁺10], by iterating over the “distance scales” 2^i , going from the smallest distance $d_{G,w}(u, v)$ among all terminals $u, v \in T$ towards the largest such distance. Each iteration i first employs a “stochastic decomposition,” which is basically a randomized procedure that finds clusters of V whose diameter is at most 2^i . Then, some clusters are contracted to a nearby terminal, which must be “adjacent” to the cluster; this way, the current graph is a minor of the previous iteration’s graph and thus also of the initial G . After iteration i is executed, we roughly expect “neighborhoods” of radius proportional to 2^i around the terminals to be contracted. As i increases, these neighborhoods get larger until eventually all the vertices are contracted into terminals, at which point the edge weights are set according to the standard restriction. To eventually get a minor, it is imperative that every contracted region be connected. To guarantee this, we perform the iteration i decomposition in the graph resulting from previous iterations’ contractions (rather than the initial G), which introduces further dependencies between the iterations.

The main challenge is to control the distortion, and this is where we crucially deviate from [EGK⁺10] (and differ from all previous work). In their randomized construction of a minor G' , for every two terminals $u, v \in T$ it is shown that G' contains

a uv -path of *expected length* at most $O(\log k)d_G(u, v)$. Consequently, they design a *distribution* D over minors G' such that the stretch $d_{G'}(u, v)/d_G(u, v)$ between any $u, v \in T$ has expectation at most $O(\log k)$. Note, however, that it is possible that no $G' \in \text{supp}(D)$ achieves a low stretch simultaneously for all $u, v \in T$. In contrast, in our randomized construction of G' , the stretch between $u, v \in T$ is polylogarithmic *with high probability*, say at least $1 - 1/k^3$. Applying a simple union bound over the $\binom{k}{2}$ terminal pairs, we can then obtain a single graph G' achieving a polylogarithmic distortion. Technically, these bounds follow by fixing in G a shortest-path P between two terminals $u, v \in T$ and then tracking the execution of the randomized algorithm to analyze how the path P evolves into a uv -path P' in G' . In [EGK⁺10], the length of P' is analyzed in expectation, which by linearity of expectation follows from analyzing the case where P consists of a single edge; in contrast, we provide for P a high-probability bound, which inevitably must consider (anti)correlations along the path.

The next section features a new tool that we developed in our quest for high-probability bounds and which may be of independent interest. For the sake of clarity, we provide below a vanilla version that excludes technical complications such as terminals, strong diameter, and consistency between scales. The proof of Theorem 1.4 actually does require these complications and thus cannot use the generic form described below.

1.2. A key technique: Metric decomposition with concentration.

Metric decomposition. Let (X, d) be a metric space, and let Π be a partition of X . Every $S \in \Pi$ is called a *cluster*, and for every $x \in X$ we use $\Pi(x)$ to denote the unique cluster $S \in \Pi$ such that $x \in S$. In general, a stochastic decomposition of the metric (X, d) is a distribution μ over partitions of X , although we usually impose additional requirements. The following definition is perhaps the most basic version, often called a separating decomposition or a Lipschitz decomposition.

DEFINITION 1.5. *A metric space (X, d) is called β -decomposable if for every $\Delta > 0$ there is a probability distribution μ over partitions of X satisfying the following requirements:*

- (a) *Diameter bound: for all $\Pi \in \text{supp}(\mu)$ and all $S \in \Pi$, $\text{diam}(S) \leq \Delta$.*
- (b) *Separation probability: for all $x, y \in X$, $\Pr_{\Pi \sim \mu}[\Pi(x) \neq \Pi(y)] \leq \frac{\beta d(x, y)}{\Delta}$.*

We note that the diameter bound holds with respect to distances in X ; in the case of a shortest-path metric in a graph, this is known as a *weak-diameter* bound.

Bartal [Bar96] proved that every n -point metric is $O(\log n)$ -decomposable and that this bound is tight. We remark that by now there is a rich literature on metric decompositions, and different variants of this notion may involve terminals or (in a graphical context) connectivity requirements inside each cluster; see, e.g., [LS93, Bar96, CKR01, FRT04, Bar04, LN05, GNR10, EGK⁺10, MN07, AGMW10, KR11].

Degree of separation. Let $P = (x_0, x_1, \dots, x_\ell)$ be a *shortest path*, i.e., a sequence of points in X such that $\sum_{i \in [\ell]} d(x_{i-1}, x_i) = d(x_0, x_\ell)$. We denote its length by $d(P) := d(x_0, x_\ell)$ and say that P *meets* a cluster $S \subseteq X$ if $S \cap P \neq \emptyset$. Given a partition Π of X , define the *degree of separation* $Z_P(\Pi)$ as the number of different clusters in the partition Π that meet P . Formally,

$$(1.1) \quad Z_P(\Pi) := \sum_{S \in \Pi} \mathbb{1}_{\{P \text{ meets } S\}}.$$

Throughout, we omit the partition Π when it is clear from the context. When we consider a random partition $\Pi \sim \mu$, the corresponding $Z_P = Z_P(\Pi)$ is actually

a random variable. If this distribution μ satisfies requirement (b) of Definition 1.5, then

$$(1.2) \quad \mathbb{E}_{\Pi \sim \mu} [Z_P] \leq 1 + \sum_{i \in [\ell]} \Pr_{\Pi \sim \mu} [\Pi(x_{i-1}) \neq \Pi(x_i)] \leq 1 + \sum_{i \in [\ell]} \frac{\beta d(x_{i-1}, x_i)}{\Delta} = 1 + \frac{\beta d(P)}{\Delta}.$$

But what about the concentration of Z_P ? More precisely, can every finite metric be decomposed such that every shortest path P admits a tail bound on its degree of separation Z_P ?

A tail bound. We answer this last question in the affirmative using the following theorem. We prove it, or actually a stronger version that does involve terminals, in section 2.

THEOREM 1.6. *For every n -point metric space (X, d) and every $\Delta > 0$ there is a probability distribution μ over partitions of X that satisfies, for $\beta = O(\log n)$, requirements (a)–(b) of Definition 1.5 and furthermore meets the following requirement:*

(c) *Degree of separation: For every shortest path P of length $d(P) \leq \frac{\Delta}{\beta}$,*

$$(1.3) \quad \forall t \geq 1, \quad \Pr_{\Pi \sim \mu} [Z_P > t] \leq 2e^{-\Omega(t)}.$$

The tail bound (1.3) can be compared to a naive estimate that holds for every β -decomposition μ : using (1.2) we have $\mathbb{E}[Z_P] \leq 2$, and then by Markov’s inequality $\Pr[Z_P \geq t] \leq 2/t$.

We remark that for general metric spaces, it is known that $\beta = O(\log n)$ is tight [Bar96]. However, for requirements (a)–(b) of Definition 1.5 several decompositions are known to have better values of β for special families of metric spaces (e.g., metrics induced by planar graphs [KPR93]). We leave it open whether for these families the bounds of Theorem 1.6 can be improved, say to $\beta = O(1)$.

1.3. Related work.

Applications. Vertex sparsification, and the “graph compression” approach in general, is obviously beneficial when G' can be computed from G very efficiently, say in linear time, and then G' may be computed on the fly rather than in advance. But compression may also be valuable in scenarios that require the storage of many graphs, like archiving and backups, or rely on low-throughput communication, like distributed or remote processing. For instance, the succinct nature of G' may be indispensable for computations performed frequently, say on a smartphone, with preprocessing done in advance on a powerful machine.

We do not have new theoretical applications that leverage our SPR result, although we anticipate these will be found later. Either way, we believe this line of work will prove technically productive and may influence, e.g., work on metric embeddings and on approximate min-cut/max-flow theorems.

Probabilistic SPR. Here, the objective is not to find a single graph $G' = (T, E', w')$, but rather a distribution D over graphs $G' = (T, E', w')$, such that every graph $G' \in \text{supp}(D)$ is isomorphic to a minor of G and its distances $d_{G', w'}$ dominate $d_{G, w}$ (on $T \times T$), and such that the distortion inequalities hold in expectation, that is,

$$\forall u, v \in T, \quad \mathbb{E}_{G' \sim D} [d_{G', w'}(u, v)] \leq \alpha \cdot d_{G, w}(u, v).$$

This problem, first posed by Chan et al. in [CXKR06], was answered by Englert et al. in [EGK⁺10] with $\alpha = O(\log |T|)$.

Distance preserving minors. This problem is a relaxation of SPR in which the minor G' may contain a few nonterminals while preserving terminal distances exactly. Formally, the objective is to find a small graph $G' = (V', E', w')$ such that (i) G' is isomorphic to a minor of G ; (ii) $T \subseteq V' \subseteq V$; and (iii) for every $u, v \in T$, $d_{G', w'}(u, v) = d_{G, w}(u, v)$. This problem was originally defined by Krauthgamer, Nguyễn, and Zondiner [KNZ14], who showed an upper bound $|V'| \leq O(|T|^4)$ for general graphs and a lower bound of $\Omega(|T|^2)$ that holds even for planar graphs.

2. Metric decomposition with concentration. In this section we prove a slightly stronger result than that of Theorem 1.6, stated as Theorem 2.2 below. Let (X, d) be a metric space, and let $\{t_1, \dots, t_k\} \subseteq X$ be a designated set of terminals. Recall that a partial partition Π of X is a collection of pairwise disjoint subsets of X . For a shortest path P in X , define $Z_P = Z_P(\Pi)$ using (1.1), which is similar to before, except that now Π is a partial partition. We first extend Definition 1.5.

DEFINITION 2.1. *We say that X is β -terminal-decomposable with concentration if for every $\Delta > 0$ there is a probability distribution μ over partial partitions of X satisfying the following properties:*

- Diameter bound: *For all $\Pi \in \text{supp}(\mu)$ and all $S \in \Pi$, $\text{diam}(S) \leq \Delta$.*
- Separation probability: *For every $x, y \in X$,*

$$\Pr_{\Pi \sim \mu} [\exists S \in \Pi \text{ such that } |S \cap \{x, y\}| = 1] \leq \frac{\beta d(x, y)}{\Delta}.$$

- Terminal cover: *For all $\Pi \in \text{supp}(\mu)$, we have $T \subseteq \bigcup_{S \in \Pi} S$.*
- Degree of separation: *For every shortest path P and every $t \geq 1$,*

$$\Pr_{\Pi \sim \mu} \left[Z_P > t \max \left\{ \frac{\beta d(P)}{\Delta}, 1 \right\} \right] \leq O \left(\min \left\{ k\beta, \left\lceil \frac{\beta d(P)}{\Delta} \right\rceil \right\} \right) e^{-\Omega(t)}.$$

THEOREM 2.2. *Every finite metric space with k terminals is $(4 \log k)$ -terminal-decomposable with concentration.*

Define the *truncated exponential* with parameters $\lambda, \Delta > 0$, denoted by $\text{Texp}(\lambda, \Delta)$, to be distribution given by the probability density function $g_{\lambda, \Delta}(x) = \frac{1}{\lambda(1-e^{-\Delta/\lambda})} e^{-x/\lambda}$ for all $x \in [0, \Delta)$.

We are now ready to prove Theorem 2.2. For simplicity of notation, we prove the result with cluster diameter at most 2Δ instead of Δ . Fix a desired $\Delta > 0$, and set for the rest of the proof $\lambda := \frac{\Delta}{\log k}$ and $g := g_{\lambda, \Delta}$. For $x \in X$ and $r > 0$, we use the standard notation of a closed ball $B(x, r) := \{y \in X : d(x, y) \leq r\}$. We define the distribution μ via the following procedure that samples a partial partition Π of X .

```

1: for  $j = 1, 2, \dots, k$  do
2:   choose  $R_j \sim \text{Texp}(\lambda, \Delta)$  independently at random, and let  $B_j = B(t_j, R_j)$ .
3:   set  $S_j = B_j \setminus \bigcup_{m=1}^{j-1} B_m$ .
4: return  $\Pi = \{S_1, \dots, S_k\} \setminus \{\emptyset\}$ .

```

The diameter bound and terminal partition properties hold by construction. The proof of the separation event property is identical to the one in [Bar96, section 3]. The following two lemmas prove the degree of separation property, which will conclude the proof of Theorem 2.2. Fix a shortest path P in X , and let us assume that $t/2$

is a positive integer; a general $t \geq 1$ can be reduced to this case up to a loss in the unspecified constant.

LEMMA 2.3. *If $d(P) < \lambda$, then $\Pr[Z_P > t] \leq 2e^{-\Omega(t)}$.*

Proof. Split the k terminals into $J_{\text{far}} := \{j \in [k] : d(t_j, P) > \Delta - 2\lambda\}$ and $J_{\text{near}} := [k] \setminus J_{\text{far}}$. Define random variables $Z_{\text{far}} := \#\{j \in J_{\text{far}} : B_j \cap P \neq \emptyset\}$ and $Z_{\text{near}} := \#\{j \in J_{\text{near}} : S_j \cap P \neq \emptyset\}$. Then $Z_P \leq Z_{\text{far}} + Z_{\text{near}}$ and

$$\Pr[Z_P > t] \leq \Pr[Z_{\text{far}} + Z_{\text{near}} > t] \leq \Pr[Z_{\text{far}} > t/2] + \Pr[Z_{\text{near}} > t/2].$$

For every $j \in J_{\text{far}}$,

$$\Pr[B_j \cap P \neq \emptyset] \leq \Pr[R_j \geq \Delta - 2\lambda] = \int_{\Delta - 2\lambda}^{\Delta} g(x) dx = \frac{k}{k-1} (e^{-\frac{\Delta-2\lambda}{\lambda}} - e^{-\frac{\Delta}{\lambda}}) \leq \frac{8}{k},$$

and therefore $\mathbb{E}[Z_{\text{far}}] \leq 8$. Since Z_{far} is the sum of independent indicators, by the Chernoff bound, $\Pr[Z_{\text{far}} > t/2] \leq 2^{-t/2}$ for all $t \geq 32e$. For smaller t , observe that $\Pr[Z_{\text{far}} = 0] \geq (1 - 8/k)^k \geq \Omega(1)$, and thus for every $t \geq 1$ we have $\Pr[Z_{\text{far}} > t/2] \leq e^{-\Omega(t)}$.

Next, consider the balls among $\{B_j : j \in J_{\text{near}}\}$ that have nonempty intersection with P . Let m denote the number of such balls, and let $j_1 < \dots < j_m$ denote their indices. In other words, we condition henceforth on an event $\mathcal{E} \in \{0, 1\}^{J_{\text{near}}}$ that determines whether $R_j \geq d(t_j, P)$ occurs or not for each $j \in J_{\text{near}}$. The indices of coordinates of \mathcal{E} that are equal to 1 are exactly j_1, \dots, j_m . For $a \in [m]$, let Y_a be the indicator variable for the event that the ball B_{j_a} does *not* contain P . Note that since $\{R_j\}_{j \in [k]}$ are independent, then so are $\{Y_a\}_{a \in [m]}$. Then

$$\begin{aligned} \Pr[Y_a = 1 \mid \mathcal{E}] &= \Pr[P \not\subseteq B_{j_a} \mid P \cap B_{j_a} \neq \emptyset] \\ &\leq \Pr[R_j < d(t_{j_a}, P) + \lambda \mid R_j \geq d(t_{j_a}, P)] \leq \frac{1 - e^{-1}}{1 - e^{-2}} \leq \frac{3}{4}. \end{aligned}$$

Having conditioned on \mathcal{E} , the event $\{Z_{\text{near}} > t/2\}$ implies that $m > t/2$ and, moreover, $Y_a = 1$ for all $a \in [t/2]$, and since $\{Y_a\}_{a \in [m]}$ are independent, $\Pr[Z_{\text{near}} > t/2 \mid \mathcal{E}] \leq (3/4)^{t/2} \leq e^{-Ct}$ for an appropriate constant $C > 0$. The last inequality holds for all such events \mathcal{E} (with the same constant $C > 0$) and thus also without any such conditioning.

Altogether, we conclude that $\Pr[Z_P > t] \leq 2e^{-\Omega(t)}$. \square

LEMMA 2.4. *If $d(P) \geq \lambda$, then*

$$\Pr[Z_P > td(P)/\lambda] \leq O\left(\min\left\{k \log k, \left\lceil \frac{d(P)}{\lambda} \right\rceil\right\}\right) e^{-\Omega(t)}.$$

Proof. Treating P as a continuous path, subdivide it into $r := \lceil d(P)/\lambda \rceil$ segments, say segments of equal length that are (except for the last one) half open and half closed. The induced subpaths P_1, \dots, P_r of P are disjoint (as subsets of X) and have length at most λ each, though some of subpaths may contain only one or even zero points of X . Writing $Z_P = \sum_{i \in [r]} Z_{P_i}$, we can apply a union bound and then Lemma 2.3 on each P_i to obtain

$$\Pr[Z_P > td(P)/\lambda] \leq \Pr[\exists i \in [r] \text{ such that } Z_{P_i} > t/2] \leq O\left(\left\lceil \frac{d(P)}{\lambda} \right\rceil\right) \cdot e^{-\Omega(t)}.$$

Furthermore, for every $j \in [k]$, let $\mathcal{A}_j := \{i \in [r] : P_i \cap B(t_j, \Delta) \neq \emptyset\}$, and since P is a shortest path, let $|\mathcal{A}_j| \leq 4\Delta/\lambda = 4 \log k$. Observe that $Z_{P_i} = 0$ (with certainty) for all $i \notin \cup_j \mathcal{A}_j$, and hence

$$\Pr[Z_P > td(P)/\lambda] \leq \Pr \left[\exists i \in \cup_{j \in [k]} \mathcal{A}_j \text{ such that } Z_{P_i} > t/2 \right] \leq 4k \log k e^{-\Omega(t)}. \quad \square$$

By substituting $\beta = 4 \log k$ and $\lambda = \Delta/\log k$, it is easy to verify that Lemmas 2.3 and 2.4 complete the proof of Theorem 2.2.

3. Terminal-centered minors: Main construction. This section proves Theorem 1.4 when $\mathcal{D} := \frac{\max_{u,v \in T} d_G(u,v)}{\min_{u,v \in T} d_G(u,v)}$ satisfies the following assumption (the extension to the general case is proved in section 4).

ASSUMPTION 3.1. $\mathcal{D} \leq 2^{k^3}$.

By scaling all edge weights, we may assume that $\min_{u,v \in T} d_G(u,v) = 1$.

NOTATION 3.2. Let $V_1, \dots, V_k \subseteq V$. For $S \subseteq [k]$, denote $V_S := \cup_{j \in S} V_j$. In addition, denote $V_\perp := V \setminus V_{[k]}$ and $V_{\perp+j} := V_\perp \cup V_j$ for any $j \in [k]$.

We now present a randomized algorithm that, given a graph $G = (V, E, w)$ and terminals $T \subset V$, constructs a terminal-centered minor G' as stated in Theorem 1.4. The algorithm maintains a partial partition $\{V_1, V_2, \dots, V_k\}$ of V , starting with $V_j = \{t_j\}$ for all $j \in [k]$. The sets grow monotonically during the execution of the algorithm. We may also think of the algorithm as if it maintains a mapping $f : V \rightarrow T \cup \{\perp\}$, starting with $f(t_j) = t_j$ for all $j \in [k]$ and gradually assigning a value in T to additional vertices, which correspond to the set $V_{[k]}$. Thus, we will also refer to the vertices in $V_{[k]}$ as *assigned* and to vertices in V_\perp as *unassigned*. The heart of the algorithm is two nested loops (lines 4–9). During every iteration of the outer loop, the inner loop performs k iterations, one for every terminal t_j . Every inner-loop iteration picks a random radius (from an exponential distribution) and “grows” V_j to that radius (but without overlapping any other set), thus removing nodes from V_\perp and assigning them to V_j . Every outer-loop iteration increases the expectation of the radius distribution. Eventually, all nodes are assigned; i.e., $\{V_1, V_2, \dots, V_k\}$ is a partition of V . Note that the algorithm does not actually contract the clusters at the end of each iteration of the outer loop. However, subsequent iterations grow each V_j only in the subgraph of G induced by the respective $V_{\perp+j}$, which is effectively the same as contracting clusters to their respective terminals at the end of each outer-loop iteration.

Every cluster in the partial partition maintained by the algorithm needs to induce a connected subgraph of G , and thus we cannot directly use the result of section 2 but rather apply more subtle arguments which use the same idea. In particular, the algorithm has to grow the clusters so that they do not overlap. We therefore require the following definition.

DEFINITION 3.3. For $U \subseteq V$, let $G[U]$ denote the subgraph of G induced by U , with induced edge lengths (i.e., $w|_{E(G[U])}$). For a subgraph H of G with induced edge lengths, a vertex $v \in V(H)$, and $r > 0$, denote $B_H(v, r) := \{u \in V(H) : d_H(u, v) \leq r\}$, where d_H is the shortest-path metric in H induced by w .

Input: $G = (V, E, w)$, $T = \{t_1, \dots, t_k\} \subseteq V$.
Output: A partition $\{V_1, V_2, \dots, V_k\}$ of V .

- 1: set $b \leftarrow 1 + 1/(45 \log k)$.
- 2: for every $j \in [k]$ set $V_j \leftarrow \{t_j\}$, $r_j = 0$.
- 3: set $i \leftarrow 0$. // i is the iteration number of the outer loop.
- 4: **while** $V_{[k]} \neq V$ **do**
- 5: $i \leftarrow i + 1$.
- 6: **for all** $j \in [k]$ **do**
- 7: choose independently at random $R_j^i \sim \exp(b^i)$.
- 8: $r_j \leftarrow r_j + R_j^i$.
- 9: $V_j \leftarrow V_j \cup B_{G[V_{\perp+j}]}(t_j, r_j)$. // This is the same as $V_j \leftarrow B_{G[V_{\perp+j}]}(t_j, r_j)$.
- 10: **return** $\{V_1, V_2, \dots, V_k\}$.

Algorithm 1: Partitioning V .

CLAIM 3.4. *The following properties hold throughout the execution of the algorithm:*

1. For all $j \in [k]$, V_j is connected in G , and $t_j \in V_j$.
2. For every $j_1, j_2 \in [k]$, if $j_1 \neq j_2$, then $V_{j_1} \cap V_{j_2} = \emptyset$.
3. For every outer loop iteration i and every $j \in [k]$, if V_j^i denotes the set V_j at the beginning of the i th iteration (of the outer loop) and V_j^{i+1} denotes the set V_j at the end of that iteration, then $V_j^i \subseteq V_j^{i+1}$.

In what follows, we analyze the stretch in distance between a fixed pair of terminals. We show that with probability at least $1 - O(k^{-5})$ the distance between these terminals in G' is at most $O(\log^5 k)$ times their distance in G . By a union bound over all $\binom{k}{2}$ pairs of terminals, we deduce Theorem 1.4. Let $s, t \in T$, and let P^* be a shortest st -path in G . Due to the triangle inequality, we may focus on pairs which satisfy $V(P^*) \cap T = \{s, t\}$, where $V(P^*)$ is the node set of P^* . We denote $\ell := w(P^*) = d_{G,w}(s, t)$.

3.1. High-level analysis. Following an execution of the algorithm, we maintain a (dynamic) path P between s and t . In a sense, in every step of the algorithm, P simulates an st -path in the terminal-centered minor induced by V_1, V_2, \dots, V_k . At the beginning of the execution, set P to be simply P^* . During the course of the execution, update P to satisfy two invariants. At every step of the algorithm, the weight of P is an upper bound on the distance between s and t in the terminal-centered minor induced by V_1, \dots, V_k (in that step). In addition, if I is a subpath of P whose inner vertices are all unassigned, then I is a subpath of P^* . Throughout the analysis, we think of P as directed from s to t , thus inducing a linear ordering of the vertices in P .

DEFINITION 3.5. *A subpath of P will be called active if it is a maximal subpath whose inner vertices are unassigned.*

Note that a single edge whose endpoints are both assigned will not be considered active.

We now describe how P is updated during the execution of the algorithm. Consider line 9 of the algorithm for the i th iteration of the outer loop and some $j \in [k]$. We say that the ball $B = B_{G[V_{\perp+j}]}(t_j, r_j)$ punctures an active subpath A of P if there is an inner node of A that belongs to the ball. If B does not puncture any active subpath of P , we do not change P . Otherwise, denote by u, v the first and last unassigned nodes (possibly not in the same active subpath) in $V(P) \cap B$, respectively. Then we do the following.

We replace the entire subpath of P between u and v with a concatenation of a shortest ut_j -path and a shortest t_jv -path that lie in B ; this is possible since $G[B]$ is connected, and $u, t_j, v \in B$. This addition to P will be called a *detour* from u to v through t_j . The process is illustrated in Figures 2(a)–2(b). Beginning with P^* , the figure describes the update after the first four balls. Note that the detour might not be a simple path. It is also worth noting that here u and v may belong to different active subpaths of P . For example, in Figure 2(c), the new ball punctures two active subpaths, and therefore in Figure 2(d) the detour goes from a node in one active subpath to a node in another active subpath. Note that in this case, we remove from P portions which are not active.

It is worth noting that this update process implies that at any given time, there is at most one detour that goes through t_j . If, for some iteration $i' < i$ of the outer loop and for some $u', v' \in V(P)$, we add a detour from u' to v' through t_j in iteration i' , we keep only one detour through t_j , from the first node between u, u' and to the last between v, v' . For example, in Figure 2(e), the ball centered in t_3 punctures an active subpath. Only one detour is kept in Figure 2(f).

The total weight of all detours during the execution will be called the *additional weight* to P (ignoring portions of P that are deleted from P). Denote the set of active subpaths of P at the beginning of the i th iteration of the outer loop by \mathcal{A}_i .

Let $V_1^{\text{fin}}, \dots, V_k^{\text{fin}}$ be the partition returned by the algorithm, let G' be the terminal-centered minor induced by that partition, and let w' be the standard restriction of w to G' . Denote by P^{fin} the path obtained at the end of the execution.

CLAIM 3.6. *At every step of the algorithm the following hold:*

1. *The weight of P is an upper bound on the distance between s and t in the terminal-centered minor induced by V_1, \dots, V_k . Moreover, once $\mathcal{A}_i = \emptyset$ (namely, P has no active subpaths), the weight of P is an upper bound on the distance between s and t in the terminal-centered minor induced by $V_1^{\text{fin}}, \dots, V_k^{\text{fin}}$ (actually, from this point on, $P = P^{\text{fin}}$).*
2. *If A is a subpath of P whose inner points are all in V_\perp , then A is a subpath of P^* .*
3. *If A_1, A_2 are two different active subpaths of P , they are internally disjoint.*
4. *$|\mathcal{A}_i| \leq k$ for all i .*

Proof. The proof follows easily by induction on i, j . □

COROLLARY 3.7. $d_{G', w'}(s, t) \leq w(P^{\text{fin}})$.

Let $A \in \mathcal{A}_i$. During the execution of the inner loop, either A is removed from P entirely, or some subpaths of A remain active (perhaps A remains active entirely). Therefore, for every $A' \in \mathcal{A}_{i+1}$, either A' is a nontrivial subpath of A (by nontrivial we mean $|V(A')| \geq 3$), or A' and A are internally disjoint. Therefore there is a laminar structure on $\bigcup_i \mathcal{A}_i$.

We describe this structure using a tree \mathcal{T} , whose node set is $\{\langle i, A \rangle \mid A \in \mathcal{A}_i\}$. The root of \mathcal{T} is $\langle 1, P^* \rangle$, and for every i and every $A \in \mathcal{A}_i$ the children of $\langle i, A \rangle$, if any, are all pairs $\langle i + 1, A' \rangle$, where $A' \in \mathcal{A}_{i+1}$ is a subpath of A . Whenever we update P , we log the weight of the detour by charging it to one of the nodes of \mathcal{T} as follows. Consider a detour from u to v in the i th iteration of the outer loop for some i . Before adding this detour, u and v are unassigned nodes in P . Because u is unassigned, u is an inner vertex of some active subpath. In either case, there is exactly one active subpath containing u . The weight of the detour is charged to the unique active subpath $A \in \mathcal{A}_i$ such that $u \in A$. For every i and $A \in \mathcal{A}_i$, let $w_{i,A}$

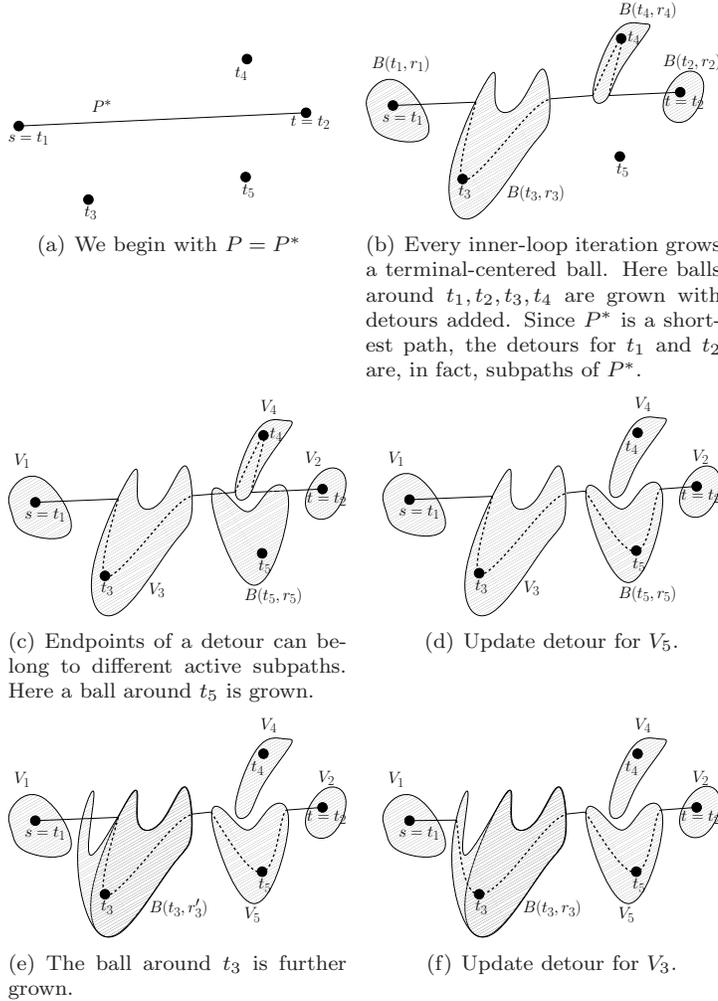


FIG. 2. Updating P .

be the total weight charged to $\langle i, A \rangle$. If the node is never charged, the weight of the node is set to 0. Therefore,

$$(3.1) \quad w(P^{\text{fin}}) \leq w(P^*) + \sum_{\langle i, A \rangle \in \mathcal{T}} w_{i, A}.$$

Equation (3.1) together with Corollary 3.7 implies that if we show that with high probability the total weight charged to the tree is at most $O(\log^5 k) \cdot \ell$, we can deduce Theorem 1.4. For the rest of this section, we therefore prove the following lemma.

LEMMA 3.8. *With probability at least $1 - O(k^{-5})$, the total weight charged to the tree is at most $O(\log^5 k)\ell$.*

Consider an iteration $i \geq 1$ and an active subpath $A \in \mathcal{A}_i$. Informally, since the distortion is measured relatively to $\ell = w(P^*)$, if the expected radius b^i is small compared to $w(A)$, then with high probability a detour will not add “much” to the

distortion, and thus we are more concerned with the opposite case, where $w(A)$ is small relative to the current expected radius.

Formally, let $p = 1/100$. An active subpath $A \in \mathcal{A}_i$ will be called *short* if $w(A) \leq pb^i$. Otherwise, A will be called *long*. Notice that $P^* \in \mathcal{A}_1$ is long, and for $i \geq \log_b(\ell/p)$, every $A \in \mathcal{A}_i$ is short.

DEFINITION 3.9. *Let $i > 1$, and let $A \in \mathcal{A}_i$ be a short subpath. Denote by $\mathcal{T}_{i,A}$ the subtree of \mathcal{T} rooted in $\langle i, A \rangle$. Denote the parent of $\langle i, A \rangle$ in \mathcal{T} by $\langle i - 1, A' \rangle$ for $A' \in \mathcal{A}_{i-1}$. If A' is long, $\mathcal{T}_{i,A}$ will be called a short subtree of \mathcal{T} .*

Once an active subpath becomes short (during the course of the iterations), we want all its vertices to be assigned quickly and by a few detours. For this reason, the height and weight of short subtrees will play an important role in the analysis of the height and weight of \mathcal{T} .

To bound the total weight of the tree \mathcal{T} , we analyze separately the weights charged to long active subpaths at each level and the weights of short subtrees rooted at each level. More formally, for every $i \leq \log_b(\ell/p)$, denote by l_i the total weight charged to nodes of the form $\langle i, A \rangle$, where $A \in \mathcal{A}_i$ is a long active subpath. Denote by s_i the total weight charged to short subtrees rooted at the level i of \mathcal{T} . For $i \geq \log_b(\ell/p)$, every $A \in \mathcal{A}_i$ is short, and thus $\langle i, A \rangle$ belongs to some short subtree rooted at level at most $\log_b(\ell/p)$. Therefore,

$$(3.2) \quad \sum_{\langle i,A \rangle \in \mathcal{T}} w_{i,A} = \sum_{i=1}^{\log_b(\ell/p)} (l_i + s_i).$$

Similarly to the proof of Theorem 2.2, we will first analyze the behavior of short subpaths of active paths and then use it to bound s_i . To bound the weight of long active paths, we will divide them into short segments, similarly to the proof of Lemma 2.4, and then sum everything up to bound l_i .

3.1.1. The effect of a single ball on a short segment. Let $i_0 \geq 1$, and let I be a subpath of P such that all the inner nodes of I are unassigned in the beginning of the i_0 th iteration of the outer loop, and $w(I) \leq pb^{i_0}$. Note that I is not necessarily maximal with that property and therefore is not necessarily an active subpath. However, I is a subpath of some (unique) active subpath $A \in \mathcal{A}_{i_0}$. We first consider the effect of a single ball over I in some iteration $i \geq i_0$.

Fix some $i \geq i_0$ and some $j \in [k]$. Let X denote the number of active subpaths A' such that $V(A') \cap V(I) \neq \emptyset$ at the beginning of the j th iteration of the inner loop (during the i th iteration of the outer loop). Note that every such subpath A' is necessarily a subpath of A , due to the laminar structure of active subpaths. Since every such active subpath will add at least one detour before it is completely assigned, we want to show that X is rapidly decreasing. Let X' denote the number of active subpaths A' such that $V(A') \cap V(I) \neq \emptyset$ at the end of the j th iteration. Denote by B the ball considered in this iteration, namely $B := B_{G[V_{1+j}]}(t_j, r_j)$.

PROPOSITION 3.10. *With certainty, $X' \leq X + 1$.*

Proof. Let A_1, A_2, \dots, A_X be all active subpaths of A which intersect I and are active in the beginning of the j th iteration ordered by their location on P . For $\alpha \in [X]$, denote by u_α, v_α the first and last unassigned nodes in A_α , respectively. If B does not puncture any of these subpaths, then $X' \leq X < X + 1$ (note that subpaths of A can still be removed if B punctures active subpaths of P not contained in A). So assume B punctures A_α . Assume first that A_α is the only subpath of P which is active and is punctured by B . Then there are three options: If both $u_\alpha, v_\alpha \in B$, then A_α is replaced

and removed entirely from P when adding the detour, and $X' \leq X - 1 < X + 1$. If $u_\alpha \in B$ and $v_\alpha \notin B$, let v' be the last node in $V(A_\alpha) \cap B$; then the $u_\alpha v'$ segment of A_α is replaced, and the segment $v'v_\alpha$ remains active. Therefore, $X' \leq X < X + 1$. The argument is similar if $u_\alpha \notin B$ and $v_\alpha \in B$. Otherwise, some of the inner portion of A_α is replaced by a nonactive path, and both end segments of A_α remain active, and therefore $X' = X + 1$. Next, assume the ball punctures several active subpaths of A and maybe more subpaths of P . Denote by I_α, I_β the first and last subpaths of A punctured by B . Denote by u the first node in $V(I_\alpha) \cap B$ and v the last node in $V(I_\beta) \cap B$. When updating P , the entire subpath of P between u and v is removed. Thus $X' \leq X - (\beta - \alpha + 1) \leq X < X + 1$. \square

We now want to show that if some unassigned vertex $v \in V(I)$ gets assigned due to a ball B , i.e., B punctures some active subpath intersecting I , then X is likely to decrease. Recalling that unassigned nodes in P must be in P^* , this goal is stated formally as

$$\Pr[X' \geq X \mid B \cap V(I) \cap V(P^*) \neq \emptyset] \leq p.$$

However, this statement is not sufficient for our needs, as it does not imply that with high probability a short active subpath is assigned quickly. Indeed, let I be a short active subpath, and suppose no ball punctures any subpath of I for many iterations following i_0 ; then the detour that will eventually be added to replace a subpath of I might be too long relative to I (as expected radii increase exponentially). Therefore, when arguing that with reasonable probability X decreases, we shall condition on a more refined event, which generalizes the notion of a ball puncturing an active subpath. Loosely speaking, we consider events in which the ball B includes a vertex $v \notin P^*$ (i.e., v is already assigned), and assume there is an unassigned $u \in V(I)$, such that uv is an edge in G (since P^* is a shortest path, this edge uv must be part of P^*), which means there is an active subpath intersecting I adjacent to v (in particular, v is one of its endpoints). By the memoryless property of the exponential distribution, conditioned on $v \in B$, with reasonable probability B covers that subpath. The formal definition follows.

DEFINITION 3.11. *Let I be a subpath of P . We say that a ball B reaches I if there is $v \in V(I) \cap B$ such that either $v \in V(P^*)$ is unassigned, or v has an unassigned neighbor which is in $V(I)$.*

Consider again the case where I is active. Then both its endpoints are assigned. Note that the endpoints of I cannot both be assigned to the same terminal (otherwise I would have been removed entirely). By the definition of the balls in the algorithm, $B \subseteq G[V_{\perp+j}]$, and therefore B may reach I and not puncture it if and only if I has exactly one endpoint in V_j . Note that all active subpaths are reached at least twice in every iteration of the outer loop (by the clusters which contain their endpoints). Therefore, in every iteration of the outer loop at least two balls reach I with certainty, even though it could be the case that no ball punctures I .

PROPOSITION 3.12. $\Pr[X' \geq X \mid B \text{ reaches } I] \leq p.$

Proof. Assume that B reaches I . Then there exists a node $v \in V(I) \cap B$ such that either v is unassigned, or v has an unassigned neighbor $u \in V(I)$. Let $d = d_{G[V_{\perp+j}]}(t_j, v)$. Assume first that $v \in V(P^*)$ is unassigned. Let A' be the active subpath such that $v \in V(A')$. Following the analysis of the previous proof, if $X' \geq X$, then B punctures exactly one active subpath (namely A') that intersects I and does not cover the part of A' contained in I , or B punctures exactly two such active subpaths and covers neither of them. In either case, B punctures A' and does not

cover the part of A' contained in I . Since $w(I) \leq pb^{i_0} \leq pb^i$, the length of A' is at most pb^i . We conclude that $r_j + R_j^i \geq d$, and $r_j + R_j^i < d + pb^i$. If v has an unassigned neighbor $u \in V(I)$, we get the same conclusion since this again means $r_j + R_j^i \geq r_j \geq d$. By the memoryless property,

$$\Pr[X' \geq X \mid B \text{ reaches } I] \leq \Pr[R_j^i < d - r_j + pb^i \mid R_j^i \geq d - r_j] \leq 1 - e^{-p} \leq p. \quad \square$$

3.1.2. The effect of a sequence of balls on a short segment. Consider now the first N balls that reach I , starting from the beginning of iteration i_0 of the outer loop and perhaps during several iterations of that loop. For every $a \in [N]$, let Y_a be the indicator random variable for the event that the a th ball reaching I decreased the number of active subpaths intersecting I . In these notations, Proposition 3.12 stated that

$$\forall a \in [N], \quad \Pr[Y_{a+1} = 1 \mid Y_1, \dots, Y_a] \geq 1 - p.$$

Let $Y = \sum_{a \in [N]} Y_a$, and let $Z \sim \text{Bin}(N, 1 - p)$. Simple induction on N implies the following claim.

CLAIM 3.13. *For all k , $\Pr[Y > k] \geq \Pr[Z > k]$.*

LEMMA 3.14. *With probability at least $1 - 1/k^{10}$, after $90 \log k$ balls have reached I , there are no active subpaths intersecting I .*

Proof. Assume $N = 70 \log k$. Since whenever $Y_a = 0$ the number of active subpaths increases by at most 1 and whenever $Y_a = 1$ the number of active subpaths decreases by at least 1, if $Y > N/2$, then there are no active subpaths intersecting I . Therefore, by the Chernoff bound,

$$\begin{aligned} \Pr[\text{there are no active subpaths intersecting } I \text{ after } N \text{ balls reach } I] \\ \geq \Pr[Y > N/2] \geq \Pr[Z > N/2] \geq 1 - 1/k^{10}. \quad \square \end{aligned}$$

3.2. The behavior of short subtrees. As stated before, the most crucial part of the proof is to bound the weight and height of short subtrees of \mathcal{T} . Let $i_0 > 1$, and let $A \in \mathcal{A}_{i_0}$ be a short subpath such that $\mathcal{T}_{i_0,A}$ is a short subtree of \mathcal{T} . Clearly, A' is short for every node $\langle i', A' \rangle$ of $\mathcal{T}_{i_0,A}$. In order to bound the height of $\mathcal{T}_{i_0,A}$ we combine the fact that not too many balls may reach A with the fact that at least two balls reach A during each iteration of the outer loop.

CLAIM 3.15. *With probability at least $1 - 1/k^{10}$, the height of $\mathcal{T}_{i_0,A}$ is at most $45 \log k$.*

Proof. In the notations of Lemma 3.14, consider some $i \geq i_0$. If A has an active subpath at the end of the i th iteration of the outer loop, then at least two times during the i th iteration an active subpath of A is reached by a ball. After $45 \log k$ iterations of the outer loop, if A has an active subpath, then $N \geq 90 \log k$. By arguments similar to Lemma 3.14,

$$\begin{aligned} \Pr[\text{The height of } \mathcal{T}_{i_0,A} \text{ is at most } N/2] &\geq \Pr[Y > N/2] \geq \Pr[Z > N/2] \\ &\geq 1 - 1/k^{10}. \quad \square \end{aligned}$$

We denote by \mathbf{E}_1 the event that for every i and every $A \in \mathcal{A}_i$, if $\mathcal{T}_{i,A}$ is a short subtree, then after at most $90 \log k$ balls reach an active subpath of A , A has no more active subpaths and, in addition, the height of $\mathcal{T}_{i,A}$ is at most $45 \log k$.

LEMMA 3.16. $\Pr[\mathbf{E}_1] \geq 1 - 1/k^5$.

Proof. Fix some i and $A \in \mathcal{A}_i$. Assume that $\mathcal{T}_{i,A}$ is a short subtree. By definition, $\langle i, A \rangle$ has no short ancestor. For $i' = \log_b(\ell/p) \leq \log_b(\mathcal{D}/p)$, P^* itself is short, since $b^{i'} = \ell/p$, and thus all tree nodes in level i' (and lower) are short. Therefore, $i \leq i'$. Since there are at most k nodes in every level of the tree, the number of short subtrees of \mathcal{T} is at most $k \cdot \log_b(\mathcal{D}/p) = k \cdot (\log_b \mathcal{D} + \log_b 100) \leq k^4 \log k + O(k \log k) \leq O(k^4 \log k)$. By the previous lemma and a union bound over all short subtrees, the result follows. \square

Since every node in level $\log_b(\ell/p)$ of the tree belongs to some short subtree, we get the following corollary.

COROLLARY 3.17. *With probability at least $1 - 2/k^5$, the height of \mathcal{T} is at most $\log_b(\ell/p) + O(\log k) \leq 10 \log_b \mathcal{D}$.*

We denote by \mathbf{E}_2 the event that for all $i \leq 10 \log_b \mathcal{D}$ and $j \in [k]$ the radius of the j th ball of the i th iteration of the outer loop is at most $O(b^i \log k)$. We wish to prove that \mathbf{E}_2 holds with high probability. We will need the following lemma, which gives a concentration bound on the sum of independent exponential random variables.

LEMMA 3.18. *Let X_1, \dots, X_n be independent random variables such that each $X_j \sim \exp(\lambda_j)$ for $\lambda_j > 0$, and denote $\lambda = \max_j \lambda_j$. Then $X = \sum_j X_j$ has expectation $\mu = \mathbb{E}[X] = \sum_j \lambda_j$ and satisfies*

$$\forall \delta > 1, \quad \Pr[X > (1 + \delta)\mu] \leq e^{(1-\delta)\frac{\mu}{2\lambda}}.$$

Proof. We proceed by applying Markov's inequality to the moment generating function (similarly to proving Chernoff bounds). Let $j \in [n]$, and consider $0 \leq t \leq \frac{1}{2\lambda_j}$. Then the moment generating function of X_j is known and can be written as $\mathbb{E}[e^{tX_j}] = \frac{1}{1-t\lambda_j} \leq 1 + 2t\lambda_j \leq e^{2t\lambda_j}$. Now set $t = \frac{1}{2\lambda}$ and use Markov's inequality to get

$$\begin{aligned} \Pr[X > (1 + \delta)\mu] &= \Pr[e^{tX} > e^{t(1+\delta)\mu}] \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu}} = \frac{\prod_{j \in [n]} \mathbb{E}[e^{tX_j}]}{e^{t(1+\delta)\mu}} \\ &\leq \frac{e^{2t\mu}}{e^{t(1+\delta)\mu}} = e^{(1-\delta)\frac{\mu}{2\lambda}}. \quad \square \end{aligned}$$

LEMMA 3.19. $\Pr[\mathbf{E}_2] \geq 1 - 1/k^5$.

Proof. Fix $i \leq 10 \log_b \mathcal{D}$ and $j \in [k]$. In the notations of Algorithm 1, let r_j be the radius of the j th ball in the i th iteration of the outer loop. Then $r_j = \sum_{i' \leq i} R_j^{i'}$ is the sum of independent exponential random variables. $\mathbb{E}[r_j] = \sum_{i' \leq i} b^{i'} = b \cdot \frac{b^i - 1}{b - 1} \geq 20b^i \log k$. Applying Lemma 3.18 we get that

$$\Pr[r_j > 40b^i \log k] = \Pr[r_j > 2\mathbb{E}[r_j]] \leq e^{-\frac{\mathbb{E}[r_j]}{b^i}} \leq k^{-10}.$$

By Assumption 3.1, $10 \log_b \mathcal{D} = O(\log \mathcal{D} \log k) = O(k^3 \log k)$. Thus by a union bound over all values of i and j in question,

$$\begin{aligned} \Pr[\forall i, j, r_j \leq 40b^i \log k \text{ in the } i\text{th iteration of the outer loop}] &\geq 1 - \frac{k \cdot 10 \log_b \mathcal{D}}{k^{10}} \\ &\geq 1 - \frac{1}{k^5}. \quad \square \end{aligned}$$

Summing everything up, we can now bound with high probability the weights of all short subtrees of \mathcal{T} .

CLAIM 3.20. *Conditioned on the events \mathbf{E}_1 and \mathbf{E}_2 , for every $i_0 > 1$ and $A \in \mathcal{A}_{i_0}$, if $\mathcal{T}_{i_0,A}$ is a short subtree of \mathcal{T} , then the total weight charged to nodes of $\mathcal{T}_{i_0,A}$ is at most $O(b^{i_0} \log^2 k)$ with certainty.*

Proof. Conditioned on \mathbf{E}_1 , at most $90 \log k$ detours are charged to nodes of every short subtree and for $i = i_0 + 45 \log k$ there are no more active subpaths of A . Conditioned on \mathbf{E}_2 , the most expensive detour being of weight at most $O(b^{i_0+45 \log k} \log k)$, we get that the total weight charged to nodes of the subtree is $90 \log k \cdot O(b^{i_0} \cdot b^{45 \log k} \cdot \log k) \leq O(b^{i_0} \log^2 k)$ since $b^{45 \log k} = O(1)$. \square

3.3. Bounding the weight of \mathcal{T} . We are now ready to bound the total weight charged to the tree. Recall that for every $i \leq \log_b(\ell/p)$ we denoted by l_i the total weight charged to nodes of the form $\langle i, A \rangle$, where $A \in \mathcal{A}_i$ is a long active subpath, and by s_i the total weight charged to short subtrees rooted in the i th level of \mathcal{T} . Since $\ell \geq 1$, $P^* \in \mathcal{A}_1$ is long. Therefore, $s_1 = 0$. We can therefore rearrange (3.2) to get

$$(3.3) \quad \sum_{\langle i,A \rangle \in \mathcal{T}} w_{i,A} = \sum_{i=1}^{\log_b(\ell/p)} (l_i + s_{i+1}).$$

Let $i \leq \log_b(\ell/p)$. Let $A \in \mathcal{A}_i$ be a long active subpath. That is, $w(A) \geq pb^i$. Thinking of A as a continuous path, divide A into $w(A)/(pb^i)$ segments of length pb^i . Some segments may contain no nodes. Let I be a segment of A , and assume I contains nodes (otherwise, no cost is charged to A on account of detours from I). Following Lemma 3.14, we get the following.

LEMMA 3.21. *With probability at least $1 - 1/k^{10}$, no more than $90 \log k$ balls reach I .*

Denote by \mathbf{E}_3 the event that for every $i \geq \log_b(\ell/k^2)$ and for every long active subpath $A \in \mathcal{A}_i$, in the division of A to segments of length pb^i , every such subsegment is reached by at most $90 \log k$ balls.

LEMMA 3.22. $\Pr[\mathbf{E}_3] \geq 1 - 1/k^5$.

Proof. Since for every $i \geq \log_b(\ell/p)$ every $A \in \mathcal{A}_i$ is short, the number of relevant iterations (of the outer loop) is at most $\log_b(\ell/p) - \log_b(\ell/k^2) = \log_b(k^2/p) \leq O(\log^2 k)$. For every $i \geq \log_b(\ell/k^2)$ and every long path $A \in \mathcal{A}_i$, the number of segments of A is at most $w(A)/(pb^i) \leq w(A)/(p\ell/k^2) \leq k^2/p$. Therefore the number of relevant segments for all $i \geq \log_b(\ell/k^2)$ and for all long $A \in \mathcal{A}_i$ is at most $O(k^2 \log^2 k)$. Applying a union bound over all relevant segments, the result follows. \square

Since $\Pr[\mathbf{E}_1] \geq 1 - 1/k^5$ and $\Pr[\mathbf{E}_2] \geq 1 - 1/k^5$, we get the following corollary.

COROLLARY 3.23. $\Pr[\mathbf{E}_1 \wedge \mathbf{E}_2 \wedge \mathbf{E}_3] \geq 1 - O(k^{-5})$.

It follows that it is enough for us to prove that conditioned on \mathbf{E}_1 , \mathbf{E}_2 , and \mathbf{E}_3 , with probability 1 the total weight charged to the tree is at most $O(\log^5 k)\ell$.

LEMMA 3.24. *Conditioned on \mathbf{E}_2 and \mathbf{E}_3 , $l_i \leq O(b^i k \log k)$. In addition, if $i \geq \log_b(\ell/k^2)$, then $l_i \leq O(\log^2 k) \cdot \ell$ with probability 1.*

Proof. To see the first bound, observe that by the update process of P at most k detours are added to P during the i th iteration. Conditioned on \mathbf{E}_2 , each one of them is of weight at most $O(b^i \log k)$. To see the second bound, let $A \in \mathcal{A}_i$ be a long active subpath. The additional weight resulting from detours from vertices of A is at most the number of segments of A of length pb^i times the additional weight to each segment. Therefore, the additional weight is at most

$$w_{i,A} \leq w(A)/pb^i \cdot O(\log k) \cdot O(b^i \log k) = O(\log^2 k) \cdot w(A).$$

Since all paths in \mathcal{A}_i are internally disjoint subpaths of P^* , we get

$$l_i = \sum_{\text{long } A \in \mathcal{A}_i} w_{i,A} \leq \sum_{\text{long } A \in \mathcal{A}_i} O(\log^2 k) \cdot w(A) \leq O(\log^2 k) \cdot \ell. \quad \square$$

LEMMA 3.25. *Conditioned on events \mathbf{E}_1 , \mathbf{E}_2 , and \mathbf{E}_3 , $s_{i+1} \leq O(b^{i+1}k \log^2 k)$. In addition, if $i \geq \log_b(\ell/k^2)$, then $s_{i+1} \leq O(\log^3 k) \cdot \ell$.*

Proof. Conditioned on \mathbf{E}_1 and \mathbf{E}_2 , we proved in Claim 3.20 that the total weight charged to a short subtree rooted in level $i + 1$ is at most $O(b^{i+1} \log^2 k)$ with certainty. Since there are at most k such subtrees, the first bound follows. To get the second bound, note that by the definition of a short subtree, for every short subtree \mathcal{T}' rooted at level $i + 1$, the parent of the root of \mathcal{T}' consists of a long active subpath A of level i . Conditioned on \mathbf{E}_3 , every segment of A is intersected by at most $90 \log k$ balls. Therefore, $\langle i, A \rangle$ can have at most $(w(A)/pb^i) \cdot 90 \log k$ children, in particular children consisting of short active subpaths. The cost of a short subtree rooted in the $i + 1$ level of \mathcal{T} is at most $O(b^i \log^2 k)$. Thus the total cost of all short subtrees rooted in children of A is bounded by

$$(w(A)/pb^i) \cdot 90 \log k \cdot O(b^i \log^2 k) \leq O(\log^3 k) \cdot w(A).$$

Summing over all (internally disjoint) long subpaths of level i , the result follows. \square

We now turn to prove Lemma 3.8.

Proof of Lemma 3.8. Since $\Pr[\mathbf{E}_1 \wedge \mathbf{E}_2 \wedge \mathbf{E}_3] \geq 1 - O(k^{-5})$, it is enough to show that conditioned on \mathbf{E}_1 , \mathbf{E}_2 , and \mathbf{E}_3 , the total weight charged to the tree is at most $O(\log^5 k)\ell$ with certainty. Recall that $\sum_{\langle i,A \rangle \in \mathcal{T}} w_{i,A} = \sum_{i=1}^{\log_b(\ell/p)} (l_i + s_{i+1})$. Following Lemmas 3.24 and 3.25 we get that

$$\begin{aligned} \sum_{i=1}^{\log_b(\ell/k^2)} (l_i + s_{i+1}) &\leq \sum_{i=1}^{\log_b(\ell/k^2)} O(b^i k \log k + b^{i+1} k \log^2 k) \\ &\leq O(k \log^2 k) \sum_{i=1}^{\log_b(\ell/k^2)} b^i = O(k \log^2 k) \frac{b}{b-1} \cdot \frac{\ell}{k^2} = o(1) \cdot \ell. \end{aligned}$$

In addition,

$$\begin{aligned} \sum_{i=\log_b(\ell/k^2)+1}^{\log_b(\ell/p)} (l_i + s_{i+1}) &\leq \sum_{i=\log_b(\ell/k^2)+1}^{\log_b(\ell/p)} O(\log^2 k)\ell + O(\log^3 k)\ell \\ &\leq O(\log^3 k)\ell \cdot (\log_b(\ell/p) - \log_b(\ell/k^2)) \\ &\leq O(\log^3 k)\ell \cdot O(\log^2 k) = O(\log^5 k) \cdot \ell. \quad \square \end{aligned}$$

4. Terminal-centered minors: Extension to general case. In this section we complete the proof of Theorem 1.4 by reducing it to the special case where Assumption 3.1 holds (which we proved in section 3). We first outline the reduction, which is implemented using a recursive algorithm, as follows. The algorithm initially rescales edge weights of the graph so that minimal terminal distance is 1. If $\mathcal{D} < 2^{k^3}$, then we apply Algorithm 1 and we are done. Otherwise, we construct a set of at most $k - 1$ low-diameter balls which are mutually far apart and whose union contains all terminals. Then, for each of the balls, we apply Algorithm 1 on the graph induced by that ball. Each ball is then contracted into a “superterminal.” We apply

the algorithm recursively on the resulting graph \tilde{G} with the set of superterminals as the terminal set. Going back from the recursion, we “stitch” together the output of Algorithm 1 on the balls in the original graph with the output of the recursive call on \tilde{G} to construct a partition of V as required. The detailed algorithm and proof of correctness are described in section 4.1. Before that, we need a few definitions.

Assume that the edge weights are already so that the minimum interterminal distance is 1. Denote by D the set of all distances between terminals, rounded down to the nearest powers of 2. Note that $|D| < k^2$. Consider the case $\mathcal{D} > 2^{k^3}$. There must exist $0 \leq m_0 \leq k^3 - k$ such that $D \cap \{2^{m_0}, 2^{m_0+1}, \dots, 2^{m_0+k}\} = \emptyset$. Define $R := \{(x, y) \in T^2 : d_G(x, y) < 2^{m_0}\}$.

CLAIM 4.1. *R is an equivalence relation.*

Proof. Reflexivity and symmetry of R follow directly from the definition of a metric. To see that R is transitive, let $x, y, z \in T$, and assume $(x, y), (y, z) \in R$. Therefore, $d_G(x, y) < 2^{m_0}$ and $d_G(y, z) < 2^{m_0}$. By the triangle inequality, $d_G(x, z) < 2^{m_0+1}$. Since $D \cap \{2^{m_0}, \dots, 2^{m_0+k}\} = \emptyset$, $d_G(x, z) < 2^{m_0}$, and therefore $(x, z) \in R$. \square

For every equivalence class $U \in T/\mathbb{R}$, we pick an arbitrary $u \in U$ and define $\hat{U} = B_G(u, 2^{m_0})$.

CLAIM 4.2. *$\mathcal{U} := \{\hat{U}\}_{U \in T/\mathbb{R}}$ is a partial partition of V . Moreover, for every $U \in T/\mathbb{R}$, $U \subseteq \hat{U}$, $G[\hat{U}]$ is connected and of diameter at most $2^{m_0+1} < 2^{k^3}$.*

Proof. Let $U \in T/\mathbb{R}$. Let $u \in U$ be such that $\hat{U} = B_G(u, 2^{m_0})$. For every $x \in U$, by the definition of R , $d(x, u) < 2^{m_0}$, and thus $x \in \hat{U}$. Therefore, $U \subseteq \hat{U}$. By the definition of a ball, $G[\hat{U}]$ is connected and of diameter at most $2^{m_0+1} < 2^{k^3}$. To see that \mathcal{U} is a partial partition of V , take $U' \in T/\mathbb{R}$ such that $U \neq U'$, and let $u' \in U'$ be such that $\hat{U}' = B_G(u', 2^{m_0})$. Since $(u, u') \notin R$, $d_G(u, u') \geq 2^{m_0}$, and since $D \cap \{2^{m_0}, \dots, 2^{m_0+k}\} = \emptyset$, $d_G(u, u') \geq 2^{m_0+k+1}$, thus $\hat{U} \cap \hat{U}' = \emptyset$. \square

4.1. Detailed algorithm. Our algorithm for the general case of Theorem 1.4 is given as Algorithm 2 (which makes calls to Algorithm 1). It is clear that this algorithm returns a partition of V . In addition, since every level of recursion decreases the number of terminals in the graph, the depth of the recursion is at most k . During each level of the recursion, Algorithm 1 is invoked at most k times. Therefore, Algorithm 1 is invoked at most k^2 times, each time on a set of at most k terminals. Note that the result of Lemma 3.8 still applies if k is only an upper bound on the number of terminals, and not the exact number of terminals. Therefore, we get that there exists $C_0 > 0$ such that all $O(k^2)$ times that the algorithm is invoked, it achieves a weight stretch factor of at most $C_0 \log^5 k$, with probability at least $1 - O(k^{-3})$. Applying a union bound, we get that with high probability the stretch bound is obtained in all invocations of the algorithm. It remains to show that this suffices to achieve the desired stretch factor in G .

LEMMA 4.3. *With probability at least $1 - 1/k$, on a graph with $\tau \leq k$ terminals, Algorithm 2 obtains a stretch factor of at most $C_0 \log^5 k + \log^5 k \cdot 2^{-k} \sum_{k' \leq \tau} 2(k')^2 \leq 2C_0 \log^5 k$.*

Proof. It is enough to show that conditioned on the event that every invocation of Algorithm 1 achieves a stretch factor of at most $C_0 \log^5 k$, the generalized algorithm achieves the desired stretch factor. We prove this by induction on k . For the case $k = 2$, rescaling the weights ensures that $\mathcal{D} = 1 \leq 2^{k^3}$, and therefore Algorithm 1 is applied on G . The result follows from the proof of section 3. Assuming correctness for every $\tilde{k} < k$, we prove correctness for k . Let $s, t \in T$. If $(s, t) \in R$, then s

Input: $G = (V, E, w)$, $T = \{t_1, \dots, t_k\} \subseteq V$.
Output: A partition $\{V_1, V_2, \dots, V_k\}$ of V .

- 1: rescale the edge weights so that the minimal terminal distance is 1.
- 2: **if** $\mathcal{D} := \max_{u,v \in T} d_G(u, v) \leq 2^{k^3}$ **then**
- 3: run Algorithm 1, and return its output.
- 4: **else**
- 5: define R and \mathcal{U} as above.
- 6: **for all** $\hat{U} \in \mathcal{U}$ **do**
- 7: run Algorithm 1 independently on $G[\hat{U}]$.
- 8: contract \hat{U} to a single “superterminal,” maintaining edge weights of all remaining edges.
- 9: denote the resulting graph \tilde{G} .
- 10: run Algorithm 2 recursively on \tilde{G} with the set of superterminals.
- 11: **for all** superterminals $u \in \tilde{G}$ **do**
- 12: let u_1, \dots, u_r be the terminals contracted to u in line 8 in an arbitrary order.
- 13: **for all** vertices v assigned to u in the recursive call **do**
- 14: assign v to its nearest terminal among u_1, \dots, u_r .
 Break ties by the ordering of u_1, \dots, u_r . // *Making sure we construct a minor.*
- 15: **return** the resulting partition of V .

Algorithm 2: Partitioning V —The General Case.

and t are in the same set in \mathcal{U} , and by the conditioning, the stretch factor of the distance between s and t is at most $C_0 \log^5 k$. This does not change in steps 4–7 of the algorithm. Otherwise, let \tilde{s}, \tilde{t} be the terminals (or superterminals) associated with s and t in \tilde{G} , respectively. Denote $d = d_{G,w}(s, t)$ and $\tilde{d} = d_{\tilde{G},\tilde{w}}(\tilde{s}, \tilde{t})$. Denote by $G' = (V', E', w')$ the terminal-centered minor induced by the partition returned by the recursive call and by $G'' = (V'', E'', w'')$ the terminal-centered minor induced by the partition returned in the final step of the algorithm. Denote $d' = d_{G',w'}(\tilde{s}, \tilde{t})$ and $d'' = d_{G'',w''}(s, t)$. Let u be a superterminal on a shortest path P' between \tilde{s} and \tilde{t} in G' . Let P'' be the path obtained from P' in G'' in the following manner. In the place of every superterminal u in P' , originating in some node set \hat{U} , we add a path between the corresponding terminals in \hat{U} (based on the terminal-centered minor constructed for $G[\hat{U}]$ in step 3). The edges of P' are also replaced with corresponding edges in G'' .

Recall that in G' the weight of every edge is the distance between its endpoints in \tilde{G} (by the definition of a terminal-centered minor). In G'' the weight of every edge is the distance between its endpoints in G . Therefore, the weight P' contained at most $k - 1$ edges. In G'' the weight of each such edge increases by at most $k2^{m_0}$. In addition, every expansion of a superterminal adds at most $k2^{m_0}$ to the path. Therefore, $w''(P'') \leq w'(P') + 2k^2 2^{m_0} \leq d' + d \cdot 2k^2 2^{-k}$. By the induction hypothesis,

$$d' \leq \left(C_0 \log^5 k + \log^5 k \cdot 2^{-k} \sum_{k' \leq k-1} (k')^2 \right) \tilde{d}.$$

Since $\tilde{d} \leq d$, we get that

$$\begin{aligned} d'' &\leq \left(C_0 \log^5 k + \log^5 k \cdot 2^{-k} \sum_{k' \leq k-1} 2(k')^2 \right) d + d \cdot 2k^2 2^{-k} \\ &\leq \left(C_0 \log^5 k + \log^5 k \cdot 2^{-k} \sum_{k' \leq k} 2(k')^2 \right) d. \end{aligned}$$

This completes the proof of Lemma 4.3 (and in fact also of Theorem 1.4). \square

Acknowledgments. We thank the anonymous reviewers for useful comments, and particularly for suggesting to improve the proof of Lemma 3.19 using a tail bound for sum of exponential random variables, which saves a factor of $\log k$ in our main result, Theorem 1.4.

REFERENCES

- [AGMW10] I. ABRAHAM, C. GAVOILLE, D. MALKHI, AND U. WIEDER, *Strong-diameter decompositions of minor free graphs*, Theory Comput. Syst., 47 (2010), pp. 837–855.
- [Bar96] Y. BARTAL, *Probabilistic approximation of metric spaces and its algorithmic applications*, in Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science, 1996, pp. 184–193.
- [Bar04] Y. BARTAL, *Graph decomposition lemmas and their role in metric embedding methods*, in Proceedings of the 12th Annual European Symposium on Algorithms, Lecture Notes in Comput. Sci. 3221, Springer, Berlin, 2004, pp. 89–97.
- [BG08] A. BASU AND A. GUPTA. *Steiner Point Removal in Graph Metrics*, unpublished manuscript; available online from <http://www.math.ucdavis.edu/~abasu/papers/SPR.pdf>, 2008.
- [BK96] A. A. BENCÚR AND D. R. KARGER, *Approximating s-t minimum cuts in $\tilde{O}(n^2)$ time*, in Proceedings of the 28th Annual ACM Symposium on Theory of Computing, 1996, pp. 47–55.
- [BSS09] J. D. BATSON, D. A. SPIELMAN, AND N. SRIVASTAVA, *Twice-Ramanujan sparsifiers*, in Proceedings of the 41st Annual ACM Symposium on Theory of Computing, 2009, pp. 255–262.
- [Chu12] J. CHUZHUY, *On vertex sparsifiers with Steiner nodes*, in Proceedings of the 44th Annual ACM Symposium on Theory of Computing, 2012, pp. 673–688.
- [CKR01] G. CALINESCU, H. KARLOFF, AND Y. RABANI, *Approximation algorithms for the 0-extension problem*, in Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms, 2001, pp. 8–16.
- [CLLM10] M. CHARIKAR, T. LEIGHTON, S. LI, AND A. MOITRA, *Vertex sparsifiers and abstract rounding algorithms*, in Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science, 2010, pp. 265–274.
- [CXKR06] T. CHAN, D. XIA, G. KONJEVOD, AND A. RICHA, *A tight lower bound for the Steiner point removal problem on trees*, in Proceedings of the 9th International Workshop on Approximation, Randomization, and Combinatorial Optimization, Lecture Notes in Comput. Sci. 4110, Springer, Berlin, 2006, pp. 70–81.
- [EGK⁺10] M. ENGLERT, A. GUPTA, R. KRAUTHGAMER, H. RÄCKE, I. TALGAM-COHEN, AND K. TALWAR, *Vertex sparsifiers: New results from old techniques*, in Proceedings of the 13th International Workshop on Approximation, Randomization, and Combinatorial Optimization, Lecture Notes in Comput. Sci. 6302, Springer, Berlin, 2010, pp. 152–165.
- [FRT04] J. FAKCHAROENPHOL, S. RAO, AND K. TALWAR, *A tight bound on approximating arbitrary metrics by tree metrics*, J. Comput. System Sci., 69 (2004), pp. 485–497.
- [GNR10] A. GUPTA, V. NAGARAJAN, AND R. RAVI, *Improved approximation algorithms for requirement cut*, Oper. Res. Lett., 38 (2010), pp. 322–325.
- [Gup01] A. GUPTA, *Steiner points in tree metrics don't (really) help*, in Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms, 2001, pp. 220–227.

- [KNZ14] R. KRAUTHGAMER, H. L. NGUYEN, AND T. ZONDINER, *Preserving terminal distances using minors*, SIAM J. Discrete Math., 28 (2014), pp. 127–141.
- [KPR93] P. KLEIN, S. A. PLOTKIN, AND S. RAO, *Excluded minors, network decomposition, and multicommodity flow*, in Proceedings of the 25th Annual ACM Symposium on Theory of Computing, 1993, pp. 682–690.
- [KR11] R. KRAUTHGAMER AND T. ROUGHGARDEN, *Metric clustering via consistent labeling*, Theory Comput., 7 (2011), pp. 49–74.
- [LM10] F. T. LEIGHTON AND A. MOITRA, *Extensions and limits to vertex sparsification*, in Proceedings of the 42nd Annual ACM Symposium on Theory of Computing, 2010, pp. 47–56.
- [LN05] J. R. LEE AND A. NAOR, *Extending Lipschitz functions via random metric partitions*, Invent. Math., 1 (2005), pp. 59–95.
- [LS93] N. LINIAL AND M. SAKS, *Low diameter graph decompositions*, Combinatorica, 13 (1993), pp. 441–454.
- [MM10] K. MAKARYCHEV AND Y. MAKARYCHEV, *Metric extension operators, vertex sparsifiers and Lipschitz extendability*, in Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science, 2010, pp. 255–264.
- [MN07] M. MENDEL AND A. NAOR, *Ramsey partitions and proximity data structures*, J. Eur. Math. Soc. (JEMS), 9 (2007), pp. 253–275.
- [Moi09] A. MOITRA, *Approximation algorithms for multicommodity-type problems with guarantees independent of the graph size*, in Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science, 2009, pp. 3–12.
- [PS89] D. PELEG AND A. A. SCHÄFFER, *Graph spanners*, J. Graph Theory, 13 (1989), pp. 99–116.
- [TZ05] M. THORUP AND U. ZWICK, *Approximate distance oracles*, J. ACM, 52 (2005), pp. 1–24.