On Strong Diameter Padded Decompositions∗

Arnold Filtser †
Bar-Ilan University
Email: arnold.filtser@biu.ac.il

Abstract

Given a weighted graph $G = (V, E, w)$, a partition of $V$ is $\Delta$-bounded if the diameter of each cluster is bounded by $\Delta$. A distribution over $\Delta$-bounded partitions is a $\beta$-padded decomposition if every ball of radius $\gamma \Delta$ is contained in a single cluster with probability at least $e^{-\beta \gamma}$. The weak diameter of a cluster $C$ is measured w.r.t. distances in $G$, while the strong diameter is measured w.r.t. distances in the induced graph $G[C]$. The decomposition is weak/strong according to the diameter guarantee.

Formerly, it was proven that $K_r$ minor free graphs admit weak decompositions with padding parameter $O(r)$, while for strong decompositions only $O(r^2)$ padding parameter was known. Furthermore, for the case of a graph $G$, for which the induced shortest path metric $d_G$ has doubling dimension $\text{ddim}$, a weak $O(\text{ddim})$-padded decomposition was constructed, which is also known to be tight. For the case of strong diameter, nothing was known.

We construct strong $O(r)$-padded decompositions for $K_r$ minor free graphs, matching the state of the art for weak decompositions. Similarly, for graphs with doubling dimension $\text{ddim}$ we construct a strong $O(\text{ddim})$-padded decomposition, which is also tight. We use this decomposition to construct strong $\tilde{O}(\text{ddim})$ sparse cover scheme for such graphs. Our new decompositions and cover have implications to approximating unique games, the construction of light and sparse spanners, and for path reporting distance oracles.

∗This paper is a full version of the proceedings version [Fil19] published in APPROX 2019. In addition to previously published material, this version contains full details of the core clustering in the cops and robbers algorithm (which somewhat simplifies over [AGG⁺ 19]). In addition, there is a new Section 6 on general graphs.

†The reasearch is supported in part by ISF grants No. (1817/17) and (1042/22), by BSF grant No. 2015813, and by the Simons Foundation.
Contents

1 Introduction ........................................... 1
   1.1 Results and Organization .......................... 2
   1.2 Related Work ...................................... 4
   1.3 Follow up work ................................... 4
   1.4 Technical Ideas ................................... 5

2 Preliminaries .......................................... 5

3 Strongly Padded Decomposition ....................... 8
   3.1 Clustering Algorithm Based on Starting Times ........ 8
   3.2 Proof of Theorem 1 ................................ 9

4 Doubling Dimension ................................... 10

5 Minor Free Graphs .................................... 12
   5.1 Strong Padded Partitions for \( K_r \) Minor Free Graphs ........... 13
   5.2 The Core Clustering Algorithm: Proof of Lemma 3 ............ 13

6 General Graphs ....................................... 18
   6.1 Strong Diameter for General Graphs .................. 19
   6.2 Weak diameter for General Metrics .................. 20
   6.3 Lower Bound ..................................... 21

7 Applications .......................................... 22
   7.1 Approximation for Unique Games on Minor Free Graphs ......... 22
   7.2 Spanner for Graphs with Moderate Doubling Dimension ....... 22
   7.3 Path Reporting Distance Oracles .................... 23

8 Conclusion and Open Problems ...................... 24

9 Acknowledgments ...................................... 24
1 Introduction

Divide and conquer is a widely used algorithmic approach. In many distance related graph problems, it is often useful to randomly partition the vertices into clusters, such that small neighborhoods have high probability of being clustered together. Given a weighed graph $G = (V, E, w)$, a partitions is $\Delta$-bounded if the diameter of every cluster is at most $\Delta$. A distribution $\mathcal{D}$ over partitions is called a $(\beta, \delta, \Delta)$-padded decomposition, if every partition is $\Delta$-bounded, and for every vertex $v \in V$ and $\gamma \in [0, \delta]$, the probability that the entire ball $B_G(v, \gamma \Delta)$ of radius $\gamma \Delta$ around $v$ is clustered together, is at least $e^{-\beta \gamma}$ (that is $\Pr[B_G(v, \gamma \Delta) \subseteq P(v)] \geq e^{-\beta \gamma}$, here $P(v)$ is the cluster containing $v$). If $G$ admits a $(\beta, \delta, \Delta)$-padded decomposition for every $\Delta > 0$, we say that $G$ admits $(\beta, \delta)$-padded decomposition scheme. If in addition $\delta = \Omega(1)$ is a universal constant, we say that $G$ is $\beta$-decomposable.

A relaxed notion is that of probabilistic decomposition, where the guarantee is over pairs rather than balls. A distribution $\mathcal{D}$ over partitions is called a $(\beta, \Delta)$-probabilistic decomposition, if every partition is $\Delta$-bounded, and for every pair of vertices vertex $u, v \in V$, $\Pr[P(u) \neq P(v)] \leq \beta \cdot \frac{d_G(u,v)}{\Delta}$. Another relaxation is called threshold probabilistic decomposition (abbreviated ThProbabilistic), where the success probability is fixed and does not goes to 1 as the distance between $u$ and $v$ goes to 0. Formally, a distribution $\mathcal{D}$ over partitions is called a $(\beta, p, \Delta)$-ThProbabilistic decomposition, if every partition is $\Delta$-bounded, and for every pair of vertices vertex $u, v \in V$ at distance at most $\frac{1}{\beta}$, $\Pr[P(u) = P(v)] \geq p$. Similarly to padded decompositions, if $G$ admits a $(\beta, \Delta)$-probabilistic/$(\beta, p, \Delta)$-ThProbabilistic decomposition for every $\Delta > 0$, we say that $G$ admits $\beta$-probabilistic/$(\beta, p)$-ThProbabilistic scheme.

| Stochastic Decomposition type | For | Guarantee |
|------------------------------|-----|----------|
| $(\beta, \delta, \Delta)$-Padded | $\gamma \in [0, \delta], v \in V$ | $\Pr[B_G(v, \gamma \Delta) \subseteq P(v)] \geq e^{-\beta \gamma}$ |
| $(\beta, \Delta)$-Probabilistic | $u, v \in V$ | $\Pr[P(u) \neq P(v)] \leq \beta \cdot \frac{d_G(u,v)}{\Delta}$ |
| $(\beta, p, \Delta)$-ThProbabilistic | $u, v \in V, d_G(u,v) \leq \frac{1}{\beta}$ | $\Pr[P(u) = P(v)] \geq p$ |

We will refer to all the three definitions as stochastic decompositions. Among other applications, stochastic decompositions have been used for multi-commodity flow [KPR93, LR99], metric embeddings [Bar96, Rao99, Rab03, KLMN04, FRT04, LN05, ABN11, ACE$^+20$, FL21, Fil21, BFT23], spanners [HIS13, FN22, HMO21], edge and vertex cut problems [Mat02, FHL08], distance oracles and routing [AGGM06, MN07, ACE$^+20$, FL21, Fil21], near linear SDD solvers [BGK$^+$14], approximation algorithms [CKR04], spectral methods [KLPT09, BLR10], and many more.

The weak diameter of a cluster $C \subseteq V$ is the maximal distance between a pair of vertices in the cluster w.r.t. the shortest path metric in the entire graph $G$, i.e. $\max_{u,v \in C} d_G(u, v)$. The strong diameter is the maximal distance w.r.t. the shortest path metric in the induced graph $G[C]$, i.e. $\max_{u,v \in C} d_G[C](u, v)$. Stochastic decomposition can be weak/strong according to the provided guarantee on the diameter of each cluster. It is considerably harder to construct decompositions with strong diameter. Nevertheless, strong diameter is more convenient to use, and some applications indeed require that (e.g. for routing, spanners e.t.c.).

Previous results on stochastic decompositions are presented in Table 1. In a seminal work, Klein, Plotkin and Rao [KPR93] showed that every $K_r$ minor free graph admits a weak $(O(r^2), \Omega(1))$-padded decomposition scheme. Fakcharoenphol and Talwar [FT03] improved the padding parameter of $K_r$ minor free graphs to $O(r^2)$ (weak diameter). Finally, Abraham, Gavoille, Gupta, Neiman, and Talwar [AGG$^+$19] improved the padding parameter to $O(r)$, still with weak diameter. The first result on strong diameter for $K_r$ minor free graphs is by Abraham, Gavoille, Malkhi, and Wieder.

\[1\]This type of decomposition was also previously called stochastic decomposition [FN22].
[AGMW10], who showed that $K_r$ minor free graph admit a strong $2^{-O(r)}$-probabilistic decomposition scheme. Afterwards, Abraham et al. [AGG+19] (the same paper providing the state of the art for weak diameter), proved that $K_r$ minor free graphs admit strong $(O(r^2), \Omega(\frac{1}{r}))$-padded decomposition scheme. It was conjectured by [AGG+19] that $K_r$ minor free graphs admit padded decomposition scheme with padding parameter $O(\log r)$. However, even improving the padding parameter for the much more basic case of small treewidth graphs remains elusive.

Another family of interest are graphs with bounded doubling dimension. Abraham, Bartal and Neiman [ABN11] showed that a graph with doubling dimension $\ddim$ is weakly $O(\ddim)$-decomposable, generalizing a result from [GKL03]. No prior strong diameter decomposition for this family is known.

General $n$-vertex graphs admit strong $O(\log n)$-probabilistic decomposition scheme as was shown by Bartal [Bar96]. In fact, the same proof can also be used to show that general graphs admit strong $(O(\log n), \Omega(1))$-padded decomposition scheme. Both results are asymptotically tight. Stated alternatively, there is some constant $c$, such that for every $\Delta > 0$, and $k \geq 1$, every $n$-vertex graph admits a distribution over $\Delta$ bounded partitions, such that every pair of vertices at distance at most $\frac{\Delta}{ck}$ will be clustered together with probability at least $n^{-\frac{1}{k}}$. As the padding parameter governs the exponent in the success probability, it is important to optimize the constant $c$. It follows implicitly from the work of Awerbuch and Peleg [AP90] that for $k \in \mathbb{N}$, general $n$-vertex graphs admit a strong $(4k - 2, O(\frac{1}{k} \cdot n^{-\frac{1}{k}}))$-ThProbabilistic decomposition scheme. See the introduction to Section 6 for additional details.

A related notion to padded decompositions is sparse cover. A collection $\mathcal{C}$ of clusters is a weak/strong $(\beta, s, \Delta)$ sparse cover if it is weakly/strongly $\Delta$-bounded, each ball of radius $\frac{\Delta}{\beta}$ is contained in some cluster, and each vertex belongs to at most $s$ different clusters. A graph admits weak/strong $(\beta, s)$ sparse cover scheme if it admits weak/strong $(\beta, s, \Delta)$ sparse cover for every $\Delta > 0$. Awerbuch and Peleg [AP90] showed that for $k \in \mathbb{N}$, general $n$-vertex graphs admit a strong $(4k - 2, 2k \cdot n^\frac{1}{k})$ sparse cover scheme. For $K_r$ minor free graphs, Abraham et al. [AGMW10] constructed strong $(O(r^2), 2^r(r + 1)!)$ sparse cover scheme. Busch, LaFortune and Tirthapura [BLT14] constructed strong $(4, f(r) \cdot \log n)$ sparse cover scheme for $K_r$ minor free graphs.\footnote{A metric space $(X, d)$ has doubling dimension $\ddim$ if every ball of radius $2r$ can be covered by $2^{\ddim}$ balls of radius $r$. The doubling dimension of a graph is the doubling dimension of its induced shortest path metric.}

For the case of graphs with doubling dimension $\ddim$, Abraham et al. [AGGM06] constructed a strong $(4, 4^{\ddim})$ sparse cover scheme. No other tradeoffs are known. In particular, if $\ddim$ is larger than $\log \log n$, the only way to get a sparse cover where each vertex belongs to $O(\log n)$ clusters is through $[AP90]$, with only $O(\log n)$ padding.

### 1.1 Results and Organization

In our first result (Theorem 3 in Section 5), we prove that $K_r$ minor free graphs are strongly $(O(r), \Omega(\frac{1}{r}))$-decomposable. Providing quadratic improvement compared to [AGG+19], and closing the gap between weak and strong padded decompositions on minor free graphs.

Our second result (Corollary 1 in Section 4) is the first strong diameter padded decompositions for doubling graphs, which is also asymptotically tight. Specifically, we prove that graphs with doubling dimension $\ddim$ are strongly $O(\ddim)$-decomposable.

Both of these padded decomposition constructions are based on a technical theorem (Theorem 1 in Section 3). Given a set of centers $N$, such that each vertex has a center at distance $\leq \Delta$, and at most $\tau$
| Partition type | Diameter | Padding/stretch | $\delta$ or $p$ | Ref/Notes |
|----------------|----------|------------------|---------------|-----------|
| Padding        | Weak     | $O(r^3)$         | $\Omega(1)$  | [KPR93]  |
| Padding        | Weak     | $O(r^2)$         | $\Omega(1)$  | [FT03]   |
| Padding        | Weak     | $O(r)$           | $\Omega(1)$  | [AGG*19] |
| Probabilistic  | Strong   | $\exp(r)$        | $\Omega(\frac{1}{r^2})$ | [AGG+19] |
| Padding        | Strong   | $O(r^2)$         | $\Omega(\frac{1}{r^2})$ | Theorem 3 |
| Graphs with doubling dimension $\text{ddim}$ |
| Padding        | Weak     | $O(\text{ddim})$ | $\Omega(1)$  | [GKL03, ABN11] |
| Padding        | Strong   | $O(\text{ddim})$ | $\Omega(1)$  | Corollary 1 |
| General $n$-vertex graphs |
| ThProbabilistic, $k \in \mathbb{N}$ | Strong   | $4k - 2$         | $\Omega(\frac{1}{k} \cdot n^{-\frac{1}{k}})$ | [AP90], implicit |
| Padded         | Strong   | $O(\log n)$     | $\Omega(1)$  | [Bar96]  |
| ThProbabilistic, $k \in \mathbb{R}_{\geq 1}$ | Strong   | $2k$             | $\Omega(n^{-\frac{1}{2-k}})$ | Theorem 4 |
| ThProbabilistic, $k \in \mathbb{N}$ | Weak     | $2k$             | $n^{-\frac{1}{k}}$ | Theorem 5 |

Table 1: Summery of all known and new results on stochastic decomposition schemes for various graph families. The $\delta$ or $p$ column represents the $\delta$ parameter in stochastic decompositions, and the success probability ($p$) in ThProbabilistic decompositions (it is not applicable in probabilistic decompositions).

centers at distance $\leq 3\Delta (\forall v, |B_G(v, 3\Delta) \cap N| \leq \tau)$, we construct a strong $(O(\log \tau), \Omega(1), 4\Delta)$-padded decomposition. All of our decompositions can be efficiently constructed in polynomial time.

Our third contribution is in Section 6, and it is for general graphs. Here we construct strong $(2k, \Omega(n^{-\frac{1}{2-k}}))$-ThProbabilistic decomposition scheme (Theorem 4), and weak $(2k, n^{-\frac{1}{k}})$-ThProbabilistic decomposition scheme (Theorem 5). This bounds are close to being tight, assuming Erdős girth conjecture [Erd64], as for $t < 2k + 1$, if every $n$ point metric space admits a weak $(t, p)$-ThProbabilistic decomposition scheme, then $p = \tilde{O}(n^{-\frac{1}{k}})$ (Theorem 6). Note that this implies that the success probability in our decompositions cannot be improved. However, it might be possible to improve the stretch parameter to $2k - 1$ without changing the success probability. See Table 1 for a summary of results on stochastic decompositions.

Our fourth result (Theorem 2 in Section 4) is a strong sparse cover for doubling graphs. For every parameter $t \geq 1$, we construct a strong $(O(t), O(2^{\text{ddim}/t} \cdot \text{ddim} \cdot \log t))$ sparse cover scheme. Note that for $t = 1$ we (asymptotically) obtain the result of [AGMW10]. However, we also get the entire spectrum of padding parameters. In particular, for $t = \text{ddim}$ we get a strong $(O(\text{ddim}), \tilde{O}(\text{ddim}))$ sparse cover scheme.

Next, we overview some of the previously known applications of strong diameter stochastic decomposition, and analyze the various improvements achieved using our results. Specifically:

1. Given an instance of the unique games problem where the input graph is $K_r$ minor free, Alev and Lau [AL17] showed that if there exist an assignment that satisfies all but an $\epsilon$-fraction of the edges, then there is an efficient algorithm that finds an assignment that satisfies all but an $O(r \cdot \sqrt{\epsilon})$-fraction. Using our padded decompositions for minor-free graphs, we can find an assignment that satisfies all but an $O(\sqrt{r \cdot \epsilon})$-fraction of the edges. See Section 7.1.

2. Using the framework of Filtser and Neiman [FN22], given an $n$ vertex graph, with doubling
dimension $ddim$, for every parameter $t > 1$ we construct a graph-spanner with stretch $O(t)$, lightness $O(2^{ddim} \cdot t \cdot \log^2 n)$ and $O(n \cdot 2^{ddim} \cdot \log n \cdot \log \Lambda)$ edges. The only previous spanner of this type appeared in [FN22], was based on weak diameter decompositions, had the same stretch and lightness, while having no bound whatsoever on the number of edges. See Section 7.2.

3. Elkin, Neiman and Wulff-Nilsen [ENW16] constructed a path reporting distance oracle for $K_r$ minor free graphs with stretch $O(r^2)$, space $O(n \cdot \log \Lambda \cdot \log n)$ and query time $O(\log \log \Lambda)$. That is, on a query $\{u, v\}$ the distance oracle returns a $u - v$ path $P$ of weight at most $O(r^2 \cdot d_G(u, v))$ in $O(|P| + \log \log \Lambda)$ time. Using our strong diameter padded decompositions we improve the stretch to $O(r)$, while keeping the other parameters intact. See Section 7.3.

4. We further use the framework of [ENW16] to create a path reporting distance oracle for graphs having doubling dimension $ddim$ with stretch $O(ddim)$, space $O(n \cdot ddim \log \Lambda)$ and query time $O(\log \log \Lambda)$. This is the first path reporting distance oracle for doubling graphs. The construction uses our strong sparse covers. See Section 7.3.

1.2 Related Work

Miller et al. [MPX13] constructed strong diameter partitions for general graphs, which they later used to construct spanners and hop-sets in parallel and distributed regimes (see also [EN18]). Hierarchical partitions with strong diameter had been studied and used for constructing distributions over spanning trees with small expected distortion [EEST08, AN19], Ramsey spanning trees [ACE+20], spanning clan embeddings [FL21], and for spanning universal Steiner trees [BDR+12]. Another type of diameter guarantee appearing in the literature is when we require only weak diameter, and in addition for each cluster to be connected [EJK+17, FKT19, Fil20].

Related type of partitions are sparse partitions [JLN+05] which is a single partition into bounded diameter clusters such that every small ball intersect small number of clusters (in the worst case), and scattering partitions [Fil20] where the guarantee is that every shortest path intersect small number of clusters (worst case).

Stochastic decompositions were studied for additional graph families. Abraham et al. [AGG+19] showed that pathwidth $r$ graphs admit strong $(O(\log r), \Omega(1))$-padded decomposition scheme, implying that treewidth $r$ graphs admit strong $(O(\log r + \log \log n), \Omega(1))$-padded decomposition scheme (see also [KK17]). In addition, [AGG+19] showed that treewidth $r$ graphs admit strong $(O(r), \Omega(\frac{1}{r}))$-padded decomposition scheme. Finally [AGG+19] proved that genus $g$ graphs admit strong $O(\log g)$-padded decomposition scheme, improving a previous weak diameter version of Lee and Sidiropoulos [LS10].

1.3 Follow up work

Following the techniques developed in this paper, in a companion paper [Fil20] we constructed strong diameter sparse partition. Recently, this techniques were pushed further to construct hierarchical strong sparse partition [BCF+23]. This finally led to a poly-logarithmic stretch solution for the universal Steiner tree problem (UST). Sparse partitions were also recently used to solve the facility locations problem in high dimension in the streaming model [CFJ+23].

---

4Lightness is the ratio between the weight of the spanner to the weight of the MST.

$\Lambda = \max_{u, v \in V} d_G(u, v) / \min_{u, v \in V} d_G(u, v)$ is the aspect ratio.
Our sparse covers for doubling metrics (Theorem 2), and stochastic decompositions for general metrics (Theorem 5) have been recently used to construct ultrametric covers, which in turn were used to construct reliable spanners [FL22, Fil23, FGN23].

1.4 Technical Ideas

The basic approach for creating padded decompositions is by ball carving [Bar96, ABN11]. That is, iteratively create clusters by taking a ball centered around some vertex, with radius drawn according to exponential distribution. The process halts when all the vertices are clustered. Intuitively, if every vertex might join the cluster associated with at most \( \tau \) centers, the padding parameter is \( O(\log \tau) \).

We think of these centers as threateners. This approach worked very well for general graphs as the number of vertices is \( n \). Similarly it also been used for doubling graphs [BFT23], where the number of threateners is bounded by \( 2^{O(\text{dim})} \). However, in doubling graphs ball carving produces only weak diameter clustering.

Our main technical contribution is a proof that the intuition above holds for strong diameter as well. Specifically, we show that if there is a set \( N \) of centers such that each vertex has a center at distance at most \( \Delta \), and at most \( \tau \) centers at distance \( 3\Delta \) (these are the threateners), then the graph admits a strong \( (O(\log \tau), \Omega(1), 4\Delta) \)-padded decomposition scheme. We use the clustering approach of Miller et al. [MPX13] with exponentially distributed starting times. In short, in [MPX13] clustering, each center \( x \) samples a starting time \( \delta_x \). A vertex \( v \) joins the cluster of the center \( x_i \) maximizing \( \delta_x - d_G(x, v) \). This approach guaranteed to creates strong diameter clusters. The key observation is that if for every center \( y \neq x_i \), \( (\delta_{x_i} - d_G(x_i, v)) - (\delta_y - d_G(y, v)) \geq 2\gamma \Delta \), then the ball \( B_G(v, \gamma \Delta) \) is fully contained in the cluster of \( x_i \). Using truncated exponential distribution, we lower bound the probability of this event by \( e^{-\gamma \cdot O(\log \tau)} \). It is the first time [MPX13]-like algorithm is used to create padded decompositions. The [MPX13] algorithm can be efficiently implemented in distributed and parallel settings. Moreover, as each vertex depends only on centers in its local area, we are able to use the Lovász Local Lemma to create a sparse cover from padded decompositions.

Decompositions of \( K_r \) minor free graphs did not use ball carving directly. Rather, they tend to use the topological structure of the graph. We say that a cluster of \( G \) has an \( r \)-core with radius \( \Delta \) if it contains at most \( r \) shortest paths (w.r.t. \( d_G \)) such that each vertex is at distance at most \( \Delta \) from one of these paths. [AGG+19]’s strong decomposition for \( K_r \) minor free graphs is based on a partition into 1-core clusters, such that a ball with radius \( \gamma \Delta \) is cut with probability at most \( 1 - e^{-O(\gamma r^2)} \). This partition is the reason for their \( O(r^2) \) padding parameter. Although not stated explicitly, [AGG+19] also constructed a partition into \( r \)-core clusters, such that a ball with radius \( \gamma \Delta \) is cut with probability at most \( 1 - e^{-O(r^3)} \). Apparently, [AGG+19] lacked an algorithm for partitioning \( r \)-clusters. Taking a union of \( \Delta \)-nets from each shortest path to the center set \( N \), it will follow that each vertex has at most \( O(r) \) centers in its \( O(\Delta) \) neighborhood. In particular, our theorem above implies a clustering of each \( r \)-core cluster into bounded diameter clusters. Our strong decomposition with parameter \( O(r) \) follows.

We include a new and full proof for the \( r \)-core clusters. See Section 5.2 and the discussion therein.

2 Preliminaries

Graphs. We consider connected undirected graphs \( G = (V, E) \) with edge weights \( w : E \to \mathbb{R}_{\geq 0} \). We say that vertices \( v, u \) are neighbors if \( \{v, u\} \in E \). Let \( d_G \) denote the shortest path metric in \( G \).
$B_G(v, r) = \{ u \in V \mid d_G(v, u) \leq r \}$ is the closed ball of radius $r$ around $v$. For a vertex $v \in V$ and a subset $A \subseteq V$, let $d_G(x, A) := \min_{a \in A} d_G(x, a)$, where $d_G(x, \emptyset) = \infty$. For a subset of vertices $A \subseteq V$, $G[A]$ denotes the induced graph on $A$, and $G \setminus A := G[V \setminus A]$.

The diameter of a graph $G$ is $\text{diam}(G) = \max_{v, u \in V} d_G(v, u)$, i.e. the maximal distance between a pair of vertices. Given a subset $A \subseteq V$, the weak-diameter of $A$ is $\text{diam}_W(A) = \max_{v, u \in A} d_G(v, u)$, i.e. the maximal distance between a pair of vertices in $A$, w.r.t. to $d_G$. The strong-diameter of $A$ is $\text{diam}(G[A])$, the diameter of the graph induced by $A$. For illustration, in the figure to the right, consider the lower cluster encircled by a dashed red line. The weak-diameter of the cluster is 4 (as $d_G(v, z) = 4$) while the strong-diameter is 6 (as $d_G[A](v, u) = 6$). A graph $H$ is a minor of a graph $G$ if we can obtain $H$ from $G$ by edge deletions/contractions, and isolated vertex deletions. A graph family $\mathcal{G}$ is $H$-minor-free if no graph $G \in \mathcal{G}$ has $H$ as a minor. Some examples of minor free graphs are planar graphs ($K_5$ and $K_{3,3}$ minor-free), outer-planar graphs ($K_4$ and $K_{3,2}$ minor-free), series-parallel graphs ($K_4$ minor-free) and trees ($K_3$ minor-free).

### Doubling dimension

The doubling dimension of a metric space is a measure of its local “growth rate”. A metric space $(X, d)$ has doubling constant $\lambda$ if for every $x \in X$ and radius $r > 0$, the ball $B(x, 2r)$ can be covered by $\lambda$ balls of radius $r$. The doubling dimension is defined as $\text{ddim} = \log_2 \lambda$. A $d$-dimensional $\ell_p$ space has $\text{ddim} = \Theta(d)$, and every $n$ point metric has $\text{ddim} \leq \log n$. We say that a weighted graph $G = (V, E, w)$ has doubling dimension $\text{ddim}$, if the corresponding shortest path metric $(V, d_G)$ has doubling dimension $\text{ddim}$. The following lemma gives the standard packing property of doubling metrics (see, e.g., [GKL03]).

**Lemma 1** (Packing Property). Let $(X, d)$ be a metric space with doubling dimension $\text{ddim}$. If $S \subseteq X$ is a subset of points with minimum interpoint distance $r$ that is contained in a ball of radius $R$, then $|S| \leq \left( \frac{2R}{r} \right)^{O(\text{ddim})}$.

### Nets

A set $N \subseteq V$ is called a $\Delta$-net, if for every vertex $v \in V$ there is a net point $x \in N$ at distance at most $d_G(v, x) \leq \Delta$, while every pair of net points $x, y \in N$, is farther than $d_G(x, y) > \Delta$. A $\Delta$-net can be constructed efficiently in a greedy manner. In particular, by Lemma 1, given a $\Delta$-net $N$ in a graph of doubling dimension $\text{ddim}$, a ball of radius $R \geq \Delta$, will contain at most $\left( \frac{2R}{\Delta} \right)^{O(\text{ddim})}$ net points.

### Padded Decompositions and Sparse Covers

Consider a partition $\mathcal{P}$ of $V$ into disjoint clusters. For $v \in V$, we denote by $P(v)$ the cluster $P \in \mathcal{P}$ that contains $v$. A partition $\mathcal{P}$ is strongly $\Delta$-bounded (resp. weakly $\Delta$-bounded ) if the strong-diameter (resp. weak-diameter) of every $P \in \mathcal{P}$ is bounded by $\Delta$. If the ball $B_G(v, \gamma \Delta)$ of radius $\gamma \Delta$ around a vertex $v$ is fully contained in $P(v)$, we say that $v$ is $\gamma$-padded by $\mathcal{P}$. Otherwise, if $B_G(v, \gamma \Delta) \not\subseteq P(v)$, we say that the ball is cut by the partition.

**Definition 1** (Padded Decomposition). A distribution $\mathcal{D}$ over partitions of a graph $G = (V, E, w)$ is strong (resp. weak) $(\beta, \delta, \Delta)$-padded decomposition if every $P \in \text{supp}(\mathcal{D})$ is strongly (resp. weakly) $\Delta$-bounded and for any $0 \leq \gamma \leq \delta$, and $z \in V$,

$$\Pr[B_G(z, \gamma \Delta) \subseteq P(z)] \geq e^{-\beta \gamma}.$$
(Probabilistic Decomposition) A distribution $\mathcal{D}$ over partitions of a graph $G = (V, E, w)$ is strong (resp. weak) $(\beta, \Delta)$-probabilistic decomposition if every $P \in \text{supp}(\mathcal{D})$ is strongly (resp. weakly) $\Delta$-bounded and for every pair $u, v \in V$, 

$$\Pr[P(v) \neq P(u)] \leq \beta \cdot \frac{d_G(u, v)}{\Delta}.$$ 

(ThProbabilistic Decomposition) A distribution $\mathcal{D}$ over partitions of a graph $G = (V, E, w)$ is strong (resp. weak) $(\beta, p, \Delta)$-threshold probabilistic (abr. ThProbabilistic) decomposition if every $P \in \text{supp}(\mathcal{D})$ is strongly (resp. weakly) $\Delta$-bounded and for every pair $u, v \in V$, such that $d_G(u, v) \leq \frac{\Delta}{\beta}$,

$$\Pr[P(v) = P(u)] \geq p.$$ 

We say that a graph $G$ admits a strong (resp. weak) $(\beta, \delta, \Delta)$-padded decomposition scheme / $\beta$-probabilistic decomposition scheme / $(\beta, p)$-ThProbabilistic decomposition scheme, if for every parameter $\Delta > 0$ it admits a strong (resp. weak) $(\beta, \delta, \Delta)$-padded decomposition / $(\beta, \Delta)$-probabilistic decomposition / $(\beta, p, \Delta)$-ThProbabilistic decomposition that can be sampled in polynomial time.

We observe that padded decompositions are stronger than probabilistic decompositions, which by themselves stronger than ThProbabilistic decompositions.

**Observation 1.** Suppose that a weighted graph $G = (V, E, w)$ admits a strong/weak $(\beta, \delta, \Delta)$-padded decomposition $\mathcal{D}$ such that $\delta \geq \frac{\beta}{\Delta}$, Then $\mathcal{D}$ is also a strong/weak $(\beta, \Delta)$-probabilistic decomposition.

**Proof.** Let $v, u \in V$ be a pair of vertices. If $d_G(u, v) \geq \frac{\Delta}{\beta}$, then obviously $\Pr[P(v) \neq P(u)] \leq 1 \leq \beta \cdot \frac{d_G(u, v)}{\Delta}$. Thus we can assume $d_G(u, v) \leq \frac{\Delta}{\beta} \leq \delta \Delta$. Set $\gamma = \frac{d_G(u, v)}{\Delta}$. It holds that

$$\Pr[P(v) = P(u)] \geq \Pr[B_G(v, \gamma \Delta) \subseteq P(v)] \geq e^{-\beta \gamma} \geq 1 - \beta \gamma.$$ 

In particular, $\Pr[P(v) \neq P(u)] \leq \beta \gamma = \beta \cdot \frac{d_G(u, v)}{\Delta}$ as required.  

**Observation 2.** Suppose that a weighted graph $G = (V, E, w)$ admits a strong/weak $(\beta, \Delta)$-probabilistic decomposition $\mathcal{D}$. Then for every $\beta \leq \beta'$, $\mathcal{D}$ is also a strong/weak $(\beta', \frac{\beta}{\beta'}, \Delta)$-probabilistic decomposition.

**Proof.** Let $v, u \in V$ be a pair of vertices at distance at most $d_G(u, v) \leq \frac{\Delta}{\beta'}$, then $\Pr[P(v) \neq P(u)] \leq \frac{\beta}{\Delta} \cdot \frac{\Delta}{\beta'} = \frac{\beta}{\beta'}$, and hence $\Pr[P(v) = P(u)] \geq 1 - \frac{\beta}{\beta'}$.  

A related notion to padded decompositions is sparse covers.

**Definition 2 (Sparse Cover).** A collection of clusters $C = \{C_1, ..., C_t\}$ is called a weak/strong $(\beta, s, \Delta)$ sparse cover if the following conditions hold.

1. Bounded diameter: The weak/strong diameter of every $C_i \in C$ is bounded by $\Delta$.
2. Padding: For each $v \in V$, there exists a cluster $C_i \in C$ such that $B_G(v, \frac{s}{\Delta}) \subseteq C_i$.
3. Overlap: For each $v \in V$, there are at most $s$ clusters in $C$ containing $v$.

We say that a graph $G$ admits a weak/strong $(\beta, s, \Delta)$ sparse cover scheme, if for every parameter $\Delta > 0$ it admits a weak/strong $(\beta, s, \Delta)$ sparse cover that can be constructed in expected polynomial time.
Truncated Exponential Distributions. To create padded decompositions, similarly to previous works, we will use truncated exponential distribution. That is, exponential distribution conditioned on the event that the outcome lays in a certain interval. The \([\theta_1, \theta_2]\)-truncated exponential distribution with parameter \(\lambda\) is denoted by \(T_{\text{exp}}[\theta_1, \theta_2](\lambda)\), and the density function is: \(f(y) = \frac{\lambda e^{-\lambda y} \cdot y}{e^{-\lambda \theta_1} - e^{-\lambda \theta_2}}\), for \(y \in [\theta_1, \theta_2]\). For the \([0, 1]\)-truncated exponential distribution we drop the subscripts and denote it by \(T_{\text{exp}}(\lambda)\), the density function is then \(f(y) = \frac{\lambda e^{-\lambda y}}{1 - e^{-\lambda}}\).

3 Strongly Padded Decomposition

In this section we prove the main technical theorem of this paper.

Theorem 1. Let \(G = (V, E, w)\) be a weighted graph and \(\Delta > 0, \tau = \Omega(1)\) parameters. Suppose that we are given a set \(N \subseteq V\) of center vertices such that for every \(v \in V:\)

- **Covering.** There is \(x \in N\) such that \(d_G(v, x) \leq \Delta\).
- **Packing.** There are at most \(\tau\) vertices in \(N\) at distance \(3\Delta\), i.e. \(|B_G(v, 3\Delta) \cap N| \leq \tau\).

Then \(G\) admits a strong \((O(\ln \tau), \frac{1}{10}, 4\Delta)\)-padded decomposition that can be efficiently sampled.

We start with description of the [MPX13] algorithm (with some adaptations), and its properties. Later, in Section 3.2 we will prove Theorem 1.

3.1 Clustering Algorithm Based on Starting Times

As we make some small adaptations, and the role of the clustering algorithm is essential, we provide full details. Let \(\Delta > 0\) be some parameter and let \(N \subseteq V\) be some set of centers such that for every \(v \in V, d_G(v, N) \leq \Delta\). For each center \(x \in N\), let \(\delta_x \in [0, \Delta]\) be some parameter. The choice of \(\{\delta_x\}_{x \in N}\) differs among different implementations of the algorithm. In our case we will sample \(\delta_x\) using truncated exponential distribution. Each vertex \(v\) will join the cluster \(C_x\) of the center \(x \in N\) for which the value \(\delta_x - d_G(x, v)\) is maximized. Ties are broken in a consistent manner, that is we have some order \(x_1, x_2, \ldots\). Among the centers \(x_i\) that minimize \(\delta_{x_i} - d_G(x_i, v)\), \(v\) will join the cluster of the center with minimal index. Note that it is possible that a center \(x \in N\) will join the cluster of a different center \(x' \in N\). An intuitive way to think about the clustering process is as follows: each center \(x\) wakes up at time \(-\delta_x\) and begins to “spread” in a continuous manner. The spread of all centers done in the same unit tempo. A vertex \(v\) joins the cluster of the first center that reaches it.

Claim 1. Every non-empty cluster \(C_x\) created by the algorithm has strong diameter at most \(4\Delta\).

Proof. Consider a vertex \(v \in C_x\). We argue that \(d_G(v, x) \leq 2\Delta\). This will already imply that \(C_x\) has weak diameter \(4\Delta\). Let \(x_v\) be the closest center to \(v\), then \(d_G(v, x_v) \leq \Delta\). As \(v\) joined the cluster of \(x\), it holds that \(\delta_x - d_G(v, x) \geq \delta_{x_v} - d_G(v, x_v)\). In particular \(d_G(v, x) \leq \delta_x + d_G(v, x_v) \leq 2\Delta\).

Let \(I\) be the shortest path in \(G\) from \(v\) to \(x\) (the blue path on the illustration on the right). For every vertex \(u \in I\) and center \(x' \in N\), it holds that

\[\delta(x) - d_G(u, x) = \delta(x) - (d_G(v, x) - d_G(v, u)) \geq \delta(x') - d_G(v, x') + d_G(v, u) \geq \delta(x') - d_G(u, x').\]
If the inequality (*) is strict, then \( \delta(x) - d_G(u, x) > \delta(x') - d_G(u, x') \) and \( u \) will prefer \( x \) over \( x' \). Otherwise, necessarily \( \delta(x) - d_G(v, x) = \delta(x') - d_G(v, x') \) and \( x \) has smaller index then \( x' \) (as it was preferred by \( v \)). In particular \( \delta(x) - d_G(u, x) = \delta(x') - d_G(u, x') \) and hence \( u \) will prefer \( x \) over \( x' \). We conclude that \( u \) will prefer the center \( x \) over any other center. It follows that \( I \subseteq C_x \). In particular, \( d_G[C_x](v, x) \leq 2\Delta \). The claim now follows.

Claim 2. Consider a vertex \( v \), and let \( x_1, x_2, \ldots \) be an ordering of the centers w.r.t. \( \delta(x_1) - d_G(v, x_1) \). That is \( \delta(x_1) - d_G(v, x_1) \geq \delta(x_2) - d_G(v, x_2) \geq \ldots \). Set \( \Upsilon = (\delta(x_1) - d_G(v, x_1)) - (\delta(x_2) - d_G(v, x_2)) \). Then for every vertex \( u \) such that \( d_G(v, u) < \frac{\Upsilon}{2} \) it holds that \( u \in C_{x_1} \).

Proof. For every center \( x_i \neq x_1 \) it holds that,

\[
\delta(x_1) - d_G(u, x_1) > \delta(x_1) - d_G(v, x_1) - \frac{\Upsilon}{2} \geq \delta(x_1) - d_G(v, x_1) + \frac{\Upsilon}{2} > \delta(x_1) - d_G(u, x_1).
\]

In particular, \( u \in C_{x_1} \).

### 3.2 Proof of Theorem 1

For every center \( x \in N \), we sample \( \delta_x^* \in [0, 1] \) according to \( \text{Texp}(\lambda) \) truncated exponential distribution with parameter \( \lambda = 2 + 2\ln \tau \). Set \( \delta_x = \delta_x^* \cdot \Delta \in [0, \Delta] \). We execute the clustering algorithm from Section 3.1 with parameters \( \{\delta_x\}_{x \in N} \) to get a partition \( P \).

According to Claim 1, we created a distribution over strongly \( 4\Delta \)-bounded partitions. Consider some vertex \( v \in V \) and parameter \( \gamma \leq \frac{1}{2} \). We will argue that the ball \( B = B_G(v, \gamma \Delta) \) is fully contained in \( P(v) \) with probability at least \( e^{-O(\gamma \log \tau)} \). Let \( N_v \) be the set of centers \( x \) for which there is non zero probability that \( C_x \) intersects \( B \). Following the calculation in Claim 1, each vertex joins the cluster of a center at distance at most \( 2\Delta \). By triangle inequality, all the centers in \( N_v \) are at distance at most \( (2 + \gamma)\Delta \leq 3\Delta \) from \( v \). In particular \( |N_v| \leq \tau \).

Set \( N_v = \{x_1, x_2, \ldots \} \) ordered arbitrarily. Denote by \( F_i \) the event that \( v \) joins the cluster of \( x_i \), i.e. \( v \in C_{x_i} \). Denote by \( C_i \) the event that \( v \) joins the cluster of \( x_i \), but not all of the vertices in \( B \) joined that cluster, that is \( v \in C_{x_i} \cap B \neq B \). To prove the theorem, it is enough to show that \( \Pr[\bigcup C_i] \leq 1 - e^{-O(\gamma \lambda)} \). Set \( \alpha = e^{-2\gamma \lambda} \).

Claim 3. For every \( i \), \( \Pr[C_i] \leq (1 - \alpha) \left( \Pr[F_i] + \frac{1}{e^{\alpha - 1}} \right) \).

Proof. As the order in \( N_v \) is arbitrary, assume w.l.o.g. that \( i = |N_v| \) and denote \( x = x_{|N_v|} \), \( C = C_i \), \( F = F_i \), \( \delta = \delta_{x_i} \) and \( \delta' = \delta_{x_i}' \). Let \( X \in [0, 1]^{|N_v| - 1} \) be the vector where the \( j \)’th coordinate equals \( \delta_{x_j}^* \).

Set \( \rho_X = \frac{1}{\Delta} \cdot (d_G(x, v) + \max_{j < |N_v|} \{\delta_{x_j}^* - d_G(x_j, v)\}) \).

Note that \( \rho_X \) is the minimal value of \( \delta' \) such that if \( \delta' > \rho_X \), then \( x \) has the maximal value \( \delta_{x} - d_G(x, v) \), and therefore \( v \) will join the cluster of \( x \). Note that it is possible that \( \rho_X > 1 \). Conditioning on the samples having values \( X \), and assuming that \( \rho_X \leq 1 \) it holds that

\[
\Pr[F | X] = \Pr[\delta' > \rho_X] = \int_{\rho_X}^{1} \frac{\lambda \cdot e^{-\lambda y}}{1 - e^{-\lambda}} dy = \frac{e^{-\rho_X \cdot \lambda} - e^{-\lambda}}{1 - e^{-\lambda}}.
\]

If \( \delta' > \rho_X + 2\gamma \) then \( \delta - d_G(x, v) > \max_{j \neq i} \{\delta_{x_j} - d_G(x_j, v)\} + 2\gamma \Delta \). In particular, by Claim 2 the ball
$B$ will be contained in $C_x$. We conclude

$$
\Pr[C | X] \leq \Pr[\rho_X \leq \delta' \leq \rho_X + 2\gamma] \\
= \int_{\rho_X}^{\max(1, \rho_X + 2\gamma)} \lambda \cdot e^{-\lambda y} \frac{1}{1 - e^{-\lambda}} dy \\
\leq e^{-\rho_X \cdot \lambda} - e^{-(\rho_X + 2\gamma) \cdot \lambda} \\
= \left(1 - e^{-2\gamma \cdot \lambda}\right) \cdot \frac{e^{-\rho_X \cdot \lambda}}{1 - e^{-\lambda}} \\
= (1 - \alpha) \cdot \left(\Pr[F | X] + \frac{1}{e^\lambda - 1}\right).
$$

Note that if $\rho_X > 1$ then $\Pr[C | X] = 0 \leq (1 - \alpha) \cdot \left(\Pr[F | X] + \frac{1}{e^\lambda - 1}\right)$ as well. Denote by $f$ the density function of the distribution over all possible values of $X$. Using the law of total probability, we can bound the probability that the cluster of $x$ cuts $B$

$$
\Pr[C] = \int_X \Pr[C | X] \cdot f(X) \, dX \\
\leq (1 - \alpha) \cdot \int_X \left(\Pr[F | X] + \frac{1}{e^\lambda - 1}\right) \cdot f(X) \, dX \\
= (1 - \alpha) \cdot \left(\Pr[F] + \frac{1}{e^\lambda - 1}\right).
$$

We bound the probability that the ball $B$ is cut.

$$
\Pr[\bigcup_i C_i] = \sum_{i=1}^{\left|N_v\right|} \Pr[C_i] \leq (1 - \alpha) \cdot \sum_{i=1}^{\left|N_v\right|} \left(\Pr[F_i] + \frac{1}{e^\lambda - 1}\right) \\
\leq \left(1 - e^{-2\gamma \cdot \lambda}\right) \cdot \left(1 + \frac{\tau}{e^\lambda - 1}\right) \\
\leq \left(1 - e^{-2\gamma \cdot \lambda}\right) \cdot \left(1 + e^{-2\gamma \cdot \lambda}\right) = 1 - e^{-4\gamma \cdot \lambda},
$$

where the last inequality follows as $e^{-2\gamma \lambda} = \frac{e^{-2\gamma \lambda}(e^\lambda - 1)}{e^\lambda - 1} \geq \frac{e^{-2\gamma \lambda}e^\lambda - 1}{e^\lambda - 1} \geq \frac{\tau - 1}{e^\lambda - 1} = \frac{\tau}{e^\lambda - 1}$. To conclude, we obtain a strongly $4\Delta$-bounded partition, such that for every $\gamma \leq \frac{1}{16}$ and $v \in V$, the ball $B_G(v, \gamma \cdot 4\Delta)$ is fully contained in a single cluster with probability at least

$$
\Pr[B(v, \gamma \cdot 4\Delta) \subseteq P(v)] \geq e^{-4(4\gamma \cdot \lambda)} = e^{-\gamma \cdot 32(1 + \ln \tau)}.
$$

### 4 Doubling Dimension

Our strongly padded decompositions for doubling graphs are a simple corollary of Theorem 1.

**Corollary 1.** Let $G = (V, E, w)$ be a weighted graph with doubling dimension $\text{ddim}$. Then $G$ admits a strong $\Theta(O(\text{ddim}), \Omega(1))$-padded decomposition scheme.
Proof. Fix some $\Delta > 0$. Let $N$ be a $\Delta$-net of $X$. According to Lemma 1, for every vertex $v$, the number of net points at distance $3\Delta$ is bounded by $2^{O(\operatorname{ddim})}$. The corollary follows by Theorem 1. \qed

Next, we construct a sparse cover scheme.

**Theorem 2.** Let $G = (V, E, w)$ be a weighted graph with doubling dimension $\operatorname{ddim}$ and parameter $t \geq 1$. Then $G$ admits a strong $(O(t), O(2^{\operatorname{ddim}/t} \cdot \operatorname{ddim} \cdot \log t))$ sparse cover scheme. In particular, there is a strong $(\operatorname{ddim}, O(\operatorname{ddim} \cdot \log \operatorname{ddim}))$ sparse cover scheme.

Proof. Let $\Delta > 0$ be the diameter parameter. Let $\alpha = \theta(1)$ be a constant to be determined later, set $\beta = \alpha \cdot t$. We will construct a strong $(\beta, O(2^{\operatorname{ddim}/t} \cdot \operatorname{ddim} \cdot \log t), 4\Delta)$ sparse cover. As $\Delta$ is arbitrary, this will imply strong $(4\beta, O(2^{\operatorname{ddim}/t} \cdot \operatorname{ddim} \cdot \log t))$ sparse cover scheme.

The sparse cover is constructed by sampling $O(2^{\operatorname{ddim}/t} \cdot \operatorname{ddim} \cdot \log t)$ independent partitions using Corollary 1 with diameter parameter $\Delta$, and taking all the clusters from all the partitions to the cover. The sparsity and strong diameter properties are straightforward. To argue that each vertex is padded in some cluster we will use the constructive version of the Lovász Local Lemma by Moser and Tardos [MT10].

**Lemma 2** (Constructive Lovász Local Lemma [MT10]). Let $\mathcal{P}$ be a finite set of mutually independent random variables in a probability space. Let $\mathcal{A}$ be a set of events determined by these variables. For $A \in \mathcal{A}$ let $\Gamma(A)$ be a subset of $\mathcal{A}$ satisfying that $A$ is independent from the collection of events $\mathcal{A} \setminus (\{A\} \cup \Gamma(A))$. If there exist an assignment of reals $x : \mathcal{A} \rightarrow (0, 1)$ such that

$$\forall A \in \mathcal{A} : \Pr[A] \leq x(A) \cdot \prod_{B \in \Gamma(A)} (1 - x(B)),$$

then there exists an assignment to the variables $\mathcal{P}$ not violating any of the events in $\mathcal{A}$. Moreover, there is an algorithm that finds such an assignment in expected time $\sum_{A \in \mathcal{A}} \frac{x(A)}{1 - x(A)} \cdot \operatorname{poly}(|\mathcal{A}| + |\mathcal{P}|).$

Formally, recall the construction of Theorem 1 used in Corollary 1. Let $N$ be a $\Delta$-net, that we will use as centers. Consider an arbitrary vertex $v \in V$, and fix some sample of the starting times $\{\delta_x\}_{x \in X}$. Let $x(1)$ be the vertex maximizing $\delta_x - d_G(x, v)$ and $x(2)$ the second largest. In other words, $\delta_{x(1)} - d_G(x(1), v) \geq \delta_{x(2)} - d_G(x(2), v) \geq \max_{x \in N \setminus \{x(1), x(2)\}} \{\delta_x - d_G(x, v)\}$. Let $\Psi_v$ be the event that $(\delta_{x(1)} - d_G(x(1), v)) - (\delta_{x(2)} - d_G(x(2), v)) < 4\frac{\Delta}{\beta}$. Recall that the event that the ball of radius $2\frac{\Delta}{\beta}$ around $v$ is cut contained in $\Psi_v$. Following the analysis of Theorem 1, $\Pr[\Psi_v] \leq 1 - e^{-O(\operatorname{ddim} \cdot \frac{1}{\beta})} = 1 - 2^{-\operatorname{ddim}/t}$, where the equality holds for an appropriate choice of $\alpha$.

Let $x_v$ be the closest center to $v$. It holds that $\delta_{x_v} - d_G(x_v, v) \geq -\Delta$, while for every center $x$ at distance larger than $3\Delta$ it holds that $\delta_x - d_G(x, v) \leq -2\Delta$. Therefore $\Psi_v$ depends only on centers at distance at most $3\Delta$. In particular, by triangle inequality, if $v$ and $u$ are farther away than $6\Delta$, $\Psi_v$ and $\Psi_u$ are independent.

We take $m = \alpha_m \cdot 2^{\frac{\operatorname{ddim}}{t}} \cdot \operatorname{ddim} \cdot \log t$ independent partitions of $X$ using Corollary 1, for $\alpha_m = \Theta(1)$ to be determined later. Denote by $\Psi^i_v$ the event representing $\Psi_v$ in the $i$'th partition. Let $\Phi_v = \bigwedge_{i=1}^m \Psi^i_v$ be the event that $v$ “failed” in all the partitions. It holds that

$$\Pr[\Phi_v] \leq \left(1 - 2^{-\operatorname{ddim}/t}\right)^m \leq e^{-2^{-\operatorname{ddim}/t} \cdot m} = e^{-\alpha_m \cdot \operatorname{ddim} \cdot \log t}.$$

Note that if $\Phi_v$ did not occurred, then the ball of radius $2\frac{\Delta}{\beta}$ around $v$ is contained in a single cluster in at least one partition.

Let $Y$ be an $\frac{\Delta}{\beta}$-net of $X$. Set $A = \{\Phi_v\}_{v \in Y}$, to be a set of events determined by $\{\delta^i_x\}_{x \in N, 1 \leq i \leq m} (\delta_x^i$ denotes $\delta_x$ in the $i$'th partition). Each event $\Phi_v$ might depend only on events $\Phi_u$ corresponding to
vertices \( u \) at distance at most \( 6\Delta \) from \( v \). By Lemma 1, \( \Phi_v \) is independent of all, but \( |\Gamma(\Phi_v)| \leq \left( \frac{12\Delta}{\Delta / 3} \right)^{O(\text{ddim})} = 2^{O(\text{ddim} \cdot \log t)} \) events. For every \( \Phi_v \in A \), set \( x(\Phi_v) = p = 2^{-O(\text{ddim} \cdot \log t)} \), such that 
\[
\max_{v \in Y} |\Gamma(v)| \leq \frac{1}{2p}.
\]
Then, for every \( \Phi_v \in A \) it holds that, 
\[
x(\Phi_v) \cdot \Pi_{B \in \Gamma(\Phi_v)} (1 - x(B)) = p \cdot (1 - p)^{|\Gamma(\Phi_v)|} \geq p \cdot (1 - p)^{\frac{1}{2p}} \geq \frac{p}{e} \geq \Pr(\Phi_v),
\]
where the last inequality holds for large enough \( \alpha_m \). By Lemma 2 we can find an assignment to \( \{d_x\}_{x \in N,1 \leq i \leq m} \) such that none of the events \( \{ \Phi_{v_i} \}_{v \in Y} \) occurred in \( \sum_{A \in A} x(A) \cdot \text{poly} (|A| + |P|) \) = \( \frac{1}{1-p} \cdot \text{poly} (n) = \text{poly} (n) \) expected time. Under this assumption, we argue that our sparse cover has the padding property. Consider some vertex \( v \in V \). There is a net point \( u \in Y \) at distance at most \( \frac{\Delta}{\beta} \) from \( v \). As the event \( \Phi_u \) did not occur, there is some cluster \( C \) in the cover in which \( u \) is padded. In particular \( B_G(v, \gamma \Delta) \subseteq B_G(u, 2\gamma \Delta) \subseteq C \) as required.

\[ \square \]

5 Minor Free Graphs

Our clustering algorithm is based on the clustering algorithm of [AGG+19], with a small modification. The clustering of [AGG+19] has two steps. In the first step the graph is partitioned into \( r \)-Core clusters (see Definition 3 below). While \( r \)-core clusters do not have bounded diameter, they do have a simple geometric structure. Moreover, this clustering also has the padding property for small balls. In the second step, each \( r \)-core cluster is partitioned into bounded diameter sub-clusters using Theorem 1.

Definition 3 \((r\text{-Core})\). Given a weighted graph \( G = (V,E,w) \), an \( r \)-core with radius \( \Delta \) is a set of at most \( r \) shortest paths \( T_1, \ldots, T_r \) such that for every \( v \in V \), \( d_G(v, \cup_i T_i) \leq \Delta \).

Given a cluster \( C \subseteq G \), we say that \( C \) is an \( r \)-core cluster with radius \( \Delta \), if \( G[C] \) has an \( r \)-core with radius \( \Delta \). Given a partition \( \mathcal{P} \) of \( G \), we say that it is an \( r \)-core partition with radius \( \Delta \) if each cluster \( C \in \mathcal{P} \), is an \( r \)-core cluster with radius \( \Delta \).

The following theorem was proved implicitly in [AGG+19].

Lemma 3 \((\text{Core Clustering [AGG+19]})\). Given a weighted graph \( G = (V,E,w) \) that excludes \( K_r \) as a minor and a parameter \( \Delta > 0 \), there is a distribution \( \mathcal{D} \) over \( r \)-core partitions with radius \( \Delta \), such that for every vertex \( v \in V \) and \( \gamma \in (0, \Omega(\frac{1}{r})) \) it holds that
\[
\Pr[B_G(v, \gamma \Delta) \subseteq P(v)] \geq e^{-O(r \cdot \gamma)}.\]

We provide a proof of Lemma 3 that differs from [AGG+19], and arguably simplifies it. See Section 5.2, and the discussion therein. We now proceed to proving our main Theorem 3. Our clustering algorithm will be executed in two steps: first we partition the graph into \( r \)-core clusters (Lemma 3) and then we partition each \( r \)-core cluster using Theorem 1.

Historical note. [AGG+19] presented two different algorithms for strong and weak padded decompositions. Each of these algorithms consists of two steps. For weak decompositions, essentially they first partitioning the graph into \( r \)-core clusters. Secondly, instead of partition further each cluster, they pick a net from the \( r \)-cores in all the clusters, and iteratively grow balls around net points, ending with weak diameter guarantee. For strong decompositions, they partition the graph into 1-core clusters (instead of \( r \)-core), ending with a probability of only \( e^{-O(r^2 \cdot \gamma)} \) for a vertex \( x \) to be \( \gamma \)-padded.
5.1 Strong Padded Partitions for $K_r$ Minor Free Graphs

**Lemma 4.** Let $G = (V, E, w)$ be a weighted graph that has an $r$-core with radius $\Delta$. Then $G$ admits a strong $(O(\log r), \frac{1}{16}, 8\Delta)$-padded decomposition.

*Proof.* Let $I_1, I_2, \ldots, I_r$ be the $r$-core of $G$. For each $i$, let $N_i$ be a $\Delta$-net of $I_i$. Set $N = \cup_i N_i$. Every vertex $v \in V$ has some vertex in $N$ at distance at most $2 \cdot \Delta$. Indeed, by definition of $r$-core, there is $x \in I_i$ such that $d_G(v, x) \leq \Delta$. Furthermore, there is a net point $y \in N_i$ at distance at most $\Delta$ from $x$. By triangle inequality $d_G(v, y) \leq 2 \cdot \Delta$. As $I_i$ is a shortest path and $N_i$ is a $\Delta$-net, there are at most $O(1)$ net points at distance $6\Delta$ from $v$ in $N_i$. We conclude that in $N$ there are at most $O(r)$ net points at distance $6\Delta$ from $v$. The lemma now follows by Theorem 1.\hfill $\Box$

**Theorem 3.** Let $G = (V, E, w)$ be a weighted graph that excludes $K_r$ as a minor. Then $G$ admits a strong $(O(r), \Omega(\frac{1}{r}))$-padded decomposition scheme.

*Proof.* Let $\Delta > 0$ be some parameter. We construct the decomposition in two steps. First we sample an $r$-core partition $\mathcal{P}$ with radius parameter $\frac{\Delta}{8}$ using Lemma 3. Next, for every cluster $C \in \mathcal{P}$, we create a partition $\mathcal{P}_C$ using Lemma 4. The final partition is simply $\cup_{C \in \mathcal{P}} \mathcal{P}_C$, the union of all the clusters in all the created partitions. It is straightforward that the created partition has strong diameter $\Delta$. To analyze the padding, consider a vertex $v \in V$ and parameter $0 < \gamma \leq \Omega(\frac{1}{r})$. Denote by $C_v$ the cluster containing $v$ in $\mathcal{P}$, and by $P(v)$ the cluster of $v$ in the final partition. Then,

$$
\Pr [B_G(v, \gamma \Delta) \subseteq P(v)] = \Pr [B_G(v, \gamma \Delta) \subseteq P(v) \mid B_G(v, \gamma \Delta) \subseteq C_v] \cdot \Pr [B_G(v, \gamma \Delta) \subseteq C_v] \\
\geq e^{-O(\gamma r)} \cdot e^{-O(\gamma \log r)} = e^{-O(\gamma r)},
$$

where we used the fact that conditioning on $B_G(v, \gamma \Delta) \subseteq C_v$, it holds that $B_G(v, \gamma \Delta) = B_{G[C_v]}(v, \gamma \Delta)$.\hfill $\Box$

5.2 The Core Clustering Algorithm: Proof of Lemma 3

In this section we prove Lemma 3, which was implicitly proved in [AGG+19]. Our algorithm and proof here follow similar lines to those in [AGG+19]. However, we made some small modifications that made the proof simpler and easier the follow (at least in the author’s subjective opinion). Specifically, we pick the radii in Algorithm 1 using real exponential distribution, as opposed to truncated exponential distribution in [AGG+19]. This change makes the padding property to follow almost immediately, but more importantly the potential function argument is not as vague as in the original proof. The use of unbounded distribution however has the drawback that the resulting clusters might have radius (w.r.t. the $r$-core definition) larger than $\Delta$. We deal with this issue by simply recursively running the algorithm on each such cluster.

Given two disjoint subsets $A, B \subseteq V$, we write $A \sim B$ if there exists an edge from a vertex in $A$ to some vertex in $B$. We denote the partition created by the algorithm by $\mathcal{S}$, and the clusters by $\{S_1, S_2, \ldots\}$. The clusters are constructed iteratively. Initially $G_1 = G$. At step $i$, $G_i = G \setminus \cup_{j=1}^{i-1} S_j$. For a connected component $C \in G_i$, let $\mathcal{K}_{i|C} = \{S_j \mid j < i \land C \sim S_j\}$ be the set of previously created clusters with a neighbor in $C_i$. To create $S_i$, pick arbitrary connected component $C_i$ in $G_i$, and a vertex $x_i \in C_i$. For every neighboring cluster $S_j \in \mathcal{K}_{i|C_i}$, pick an arbitrary vertex $u_j \in C_i$ such that $u_j$ has a neighbor in $S_j$. For each such $u_j$, let $I_j$ be an (arbitrary) shortest path in $G_i$ from $x_i$ to $u_j$. Let $T_i$ be the tree created by the union of $\{I_j\} \cup_{S_j \in \mathcal{K}_{i|C_i}}$.\footnote{Note that there is always a way to pick $\{I_j\} \cup_{S_j \in \mathcal{K}_{i|C_i}}$ such that $T_i$ will be a tree.} Sample a radius parameter $R_i$ using exponential
distribution $\exp(1)$ with parameter 1 (the density function is $f(x) = e^{-x}$). The cluster $S_i$ is defined as $B_{G_i}(T_i, R_i \Delta)$, the set of all vertices at distance at most $R_i \Delta$ from $T_i$ w.r.t. $d_{G_i}$. This finishes the construction of $S_i$. The algorithm halts when all the vertices are clustered. The pseudo-code is presented in Algorithm 1. See also Figure 1 for illustration of the algorithm.

**Algorithm 1 Core-Partition** ($G, \Delta, r$)

1: Let $G_1 \leftarrow G$, $i \leftarrow 1$.
2: Let $\mathcal{S} \leftarrow \emptyset$.
3: while $G_i$ is non-empty do
4: Let $C_i$ be an arbitrary connected component of $G_i$.
5: Pick arbitrary $x_i \in C_i$. For each $S_j \in K_{|C_i|}$, let $u_j \in C_i$ be some vertex with a neighbor in $S_j$.
6: Let $T_i$ be a tree rooted at $x_i$ and consisting of shortest paths towards $\{u_j \mid S_j \in K_{|C_i|}\}$.
7: Sample $R_i \sim \exp(1)$.
8: Let $S_i \leftarrow B_{G_i}(T_i, R_i \Delta)$.
9: Add $S_i$ to $\mathcal{S}$.
10: $G_{i+1} \leftarrow G_i \setminus S_i$.
11: $i \leftarrow i + 1$.
12: end while
13: return $\mathcal{S}$.

Provided that the graph $G$ excludes $K_r$ as a minor, for every $C_i$ it holds that $|K_{|C_i|}| \leq r - 2$. Indeed, by induction for every $S_j, S_j' \in K_{|C_i|}$, there is an edge between $S_j$ to $S_j'$. Assume for contradiction that $|K_{|C_i|}| \geq r - 1$. By contracting all the internal edges in $C_i$ and in the clusters in $K_{|C_i|}$ we will obtain $K_r$ as a minor, a contradiction. It follows that for every $i$, $T_i$ is an $(r - 2)$-core of $S_i$, with radius distributed according to exponential distribution.

Our final algorithm will be as follows: we execute Algorithm 1 to obtain a partition $\mathcal{S}$. Set $\alpha = 7r$. For every cluster $S_i \in \mathcal{S}$ such that $R_i \geq 2\alpha$, we will (recursively) execute Algorithm 1 on the induced graph $G[S_i]$ to obtain an $r$-core partition with radius $2\alpha$. The final clustering $\tilde{\mathcal{S}}$ will consist of the union of all the clusters with radii at most $2\alpha$ in $\mathcal{S}$, and the clusters returned by the all recursive calls. Clearly, $\tilde{\mathcal{S}}$ is an $r$-core partition with radius $2\alpha = O(r) \cdot \Delta$.

For the sake of analysis, instead of discrete graphs and functions, we shall work with their continuous counterparts (see a similar approach in [RR98]). Formally, we will associate our graph with a one-dimensional simplicial complex endowed with a metric. Each edge $e = (v, u)$ of weight $w(e)$ will be an interval of length $w(e)$ equipped with the standard line metric, where one endpoint identified with the vertex $v$ and the other with $u$. Naturally, the edge metrics induce a global metric on the entire structure. Note that the original distance between every pair of vertices is preserved (in particular this holds in every induced subgraph). During the execution of the algorithm we can assume that in all the arbitrary choices of centers $x_i$ in Line 5, the algorithm chooses a real vertex from $V$ (and once no such vertices remain, the algorithm halts). Furthermore, when choosing $u_j \in C_i$ in Line 5, we will simply choose a real vertex in $V$ that has a geodesic path towards a vertex in $S_j$. All in all, the continuous interpretation makes no difference on the execution of the algorithm, or the resulting partition, but will make the next definitions simpler and more natural.

The heart of our analysis will be to show that an arbitrary vertex $z$ belong to a cluster with large radius

---

\footnote{To see this note that there is a path between $u_j$ to $u_{j'}$ in $C_i$. Therefore, when creating $S_{j'}$ (assuming $j < j'$), it was the case that $S_j \in K_{|C_{j'}|}$. In particular, $T_{j'}$ contains a vertex with neighbor in $S_j$.}
Figure 1: The figure illustrates the 6 first steps in Algorithm 1. Here $G$ is the (weighted) grid graph. Note that $G$ excludes $K_5$ as a minor. In step (4), $G_4$ is the graph induced by all the vertices not colored in blue, orange or red. $G_4$ has a single connected component $C_4$. The green vertex defined as $x_4$. $K_{|C_i|$ consist of 3 clusters $S_1, S_2, S_3$ colored respectively by blue, orange and red. $T_4$ is a tree rooted in $x_4$ colored in bold green, that consist of 3 shortest paths. Each of $S_1, S_2, S_3$ has a vertex of $T_4$ as a neighbor. $R_4$ is chosen according to $\exp(1)$. The new cluster $S_4$, colored in green, consist of all vertices in $C_4$ at distance at most $R_4 \Delta$ from $T_4$ w.r.t. $d_{G_4}$.

with probability at most $\frac{1}{3}$. Let $J_z = \{ T_i \mid d_{G_i}(z, S_i) \leq \alpha \Delta \}$ be the cores of the clusters at distance at most $\alpha \Delta$ from $z$. Alternatively, $J_z$ consist of all the cores $T_i$ such that $d_{G_i}(T_i, z) \leq R_i + \alpha \Delta$. Note that $T_i$ can join $J_z$ only if $T_i$ is in the connected component of $z$, and $R_i$ is large enough. In particular $z$ will join a cluster from $J_z$. The following lemma is the technical heart of our proof.

**Lemma 5.** $\mathbb{E} [ |J_z| ] \leq 2r^{\frac{\alpha}{\alpha + r}} \cdot e^{\alpha}$.

**Proof.** We will think of growing the clusters using the exponential distribution as an iterative continues process. We define a potential function to measure the “progress” made towards clustering $z$. Consider step $i$ where $z$ is yet unclustered, and let $\tilde{K}_{C_i} = \{ S_j \in K_{C_i} \mid d_{G_i \cup S_j}(z, S_j) \leq \alpha \Delta \}$ be the subset of $K_{C_i}$ clusters at distance at most $\alpha \Delta$ from $z$.

Suppose that $\tilde{K}_{|C_i|} = \{ S_{i_1}, \ldots, S_{i_l} \}$, and note that necessarily $l \leq r - 2$. Set $\mathbf{x} := (x_1, \ldots, x_l)$ where $x_j = \frac{d_{G_i \cup S_{i_j}}(z, S_{i_j})}{\Delta}$, and

$$
\Phi(\mathbf{x}) = \sum_{j=1}^{l} e^{-x_j}
$$

For every vector $\mathbf{x}$ containing 0 or a negative value set $\Phi(\mathbf{x}) = 2r$ ($\Phi$ is defined on general vectors). While $z$ is unclustered, the value $\Phi(\mathbf{x})$ is upper bounded by $r$. For simplicity of notation, we will assume that $z \in C_i$. That is that until $z$ is clustered, Algorithm 1 in Line 4 always picks the cluster

---

Note that $J_z$ contains (the core of) clusters that been at distance at most $\alpha \Delta$ at the time of their creation, while $\tilde{K}_{C_i}$ contains clusters that are at distance at most $\alpha \Delta$ in the current connected component $C_i$. 

---

15
Note that $h$ is set to be $S_i$ respectively. At the forth step a new cluster is created with core $T_4$.

On the right, just before the creation of $S_4$ the vector $x = (x_1, x_2, x_3)$ where $x_1 = 34$, $x_2 = 6$, and $x_3 = 3$ consist of the distances from $z$ to the clusters $S_1, S_2, S_3$ respectively. The potential at this point is $\Phi(x) = e^{-3} + e^{-6} + e^{-34}$. After the creation of $S_4$ with $y = 3$, there is no longer a path from $S_1$ to $z$, the distance from $S_2$ to $z$ increased from 6 to 8, and the distance from $S_3$ to $z$ remained unchanged. The new vector is $x' = (8, 3, 3)$, and the potential is $\Phi(x') = e^{-3} + e^{-8}$.

Let $h$ be the amount $R_i$ needs to grow so that $T_i$ will join $J_z$, that is $d_{G_i}(z, S_i) \leq \alpha \cdot \Delta$. Formally

$$\tilde{R}_i = \begin{cases} \frac{d_{G_i}(z, T_i)}{\Delta} - \alpha & \text{if } d_{G_i}(z, T_i) > \alpha \cdot \Delta \\ 0 & \text{if } d_{G_i}(z, T_i) \leq \alpha \cdot \Delta \end{cases}.$$  

Let $h = \frac{1}{\Delta} \cdot d_{G_i}(z, T_i) - \tilde{R}_i$, be the amount that $R_i$ need to additionally grow so that $z$ will join $S_i$. Note that $h$ either equals $\alpha$ (if $d_{G_i}(z, T_i) > \alpha \cdot \Delta$) or to $\frac{d_{G_i}(z, T_i)}{\Delta}$. In any case, $h \leq \alpha$. See Figure 2 (left) for an illustration.

If $R_i < \tilde{R}_i$, then no new cluster joins $J_z$, and the potential remains unchanged. Thus nothing happened from our perspective. We will thus assume that $R_i \geq \tilde{R}_i$. Let $\tilde{R}_i = R_i - \tilde{R}_i$. By the memoryless property, $\tilde{R}_i$ is distributed according to exponential distribution. Let $y = h - \tilde{R}_i$. Then if $R_i < d_{G_i}(T_i, z)$ then $y = d_{G_i}(S_i, z)$, and otherwise $y$ has a negative value (and $z$ is clustered). We next analyze how the vector $x$ can change as a result of creating $S_i$. For every $j$, if $x_j \leq y$, then the shortest path from $z$ to $S_j$ is completely disjoint from $S_i$. In particular $x_j$ will remain unaffected. Otherwise, if $x_j > y$ then some vertices in the shortest path from $S_j$ to $z$ might join the new cluster $S_i$. It follows that $x_j$ can increase, disappear or remain unchanged. See Figure 2 (right) for an illustration.

For a sequence of numbers $x$ denote by $x \downarrow y$ the sequence where we delete all the numbers larger than $y$, and add $y$. Let $x(i)$ be the sequence at time $i$. We analyze the change in the potential function. The expectations are bounded from below by $\mathbb{E} [\Phi(x(i + 1)) - \Phi(x(i))] \geq \mathbb{E} [\Phi(x \downarrow (h - \tilde{R}_i)) - \Phi(x)].$  

The inequality holds as in $x \downarrow (h - \tilde{R}_i)$ we nullify all the coordinates smaller than $h - \tilde{R}_i$, while in
\[ \alpha(i+1) \text{ some of them might remain unchanged, or only increased.} \]

**Claim 4.** For every vector \( \mathbf{x} \in (0, \alpha]^l \) for \( l \in [0, r] \), for every \( h \in [0, \alpha] \) and \( \rho \) distributed according to \( \exp(1) \), it holds that that \( \mathbb{E} [\Phi(\mathbf{x} | h - \rho) - \Phi(\mathbf{x})] \geq e^{-\alpha} \cdot (\alpha + r) \).

**Proof.** If \( h = 0 \), then \( \Phi(\mathbf{x} | h - \rho) - \Phi(\mathbf{x}) \geq 2r - r \geq e^{-\alpha} \cdot (\alpha + r) \). Thus we can assume that \( h > 0 \). The gain to \( \Phi \) is a single new addend of value \( e^{-(h-\rho)} \) if \( \rho < h \), or \( 2r \) in case \( \rho \geq h \). Specifically:

\[
\mathbb{E} [\text{gain}] = \int_0^h e^{-\rho} \cdot e^{-(h-\rho)} \, \mathrm{d}\rho + \int_h^\infty e^{-\rho} \cdot 2r \, \mathrm{d}\rho \\
= e^{-h} \int_0^h 1 \, \mathrm{d}\rho + 2r \cdot e^{-h} = e^{-h} (h + 2r) .
\]

From the other hand, for the addend \( x_j \), there is a loss of \( e^{-x_j} \) if \( h - \rho < x_j \), and no change otherwise. If \( h < x_j \), this happens with probability 1 and the loss is \( e^{-x_j} < e^{-h} \). Else, the expected loss is

\[
\mathbb{P} \{ h - \rho < x_j \} \cdot e^{-x_j} = e^{-(h-x_j)} \cdot e^{-x_j} = e^{-h} .
\]

Thus the total expected loss is bounded by \( r \cdot e^{-h} \). In total

\[
\mathbb{E} [\Phi(\mathbf{x} | h - \rho) - \Phi(\mathbf{x})] \geq e^{-h} (h + 2r - r) = e^{-h} \cdot (h + r) .
\]

The function \( e^{-h} \cdot (h + r) \) monotonically decreasing, and has minimum value at \( h = \alpha \). The claim now follows. \( \square \)

Set \( \zeta = e^{-\alpha} \cdot (\alpha + r) \). Then it follows from the claim above, that for every \( i \), and \( \mathbf{x}(i) \), it holds that \( \mathbb{E} [\Phi(\mathbf{x}(i+1)) - \Phi(\mathbf{x}(i)) \mid \Phi(\mathbf{x}(i))] \geq \zeta \). Let \( X_i = \Phi(\mathbf{x}(i)) - t \zeta \). It holds that

\[
\mathbb{E} [X_{t+1} \mid X_1 \ldots X_t] = \mathbb{E} [\Phi(\mathbf{x}(t+1)) - (t+1) \zeta \mid \mathbf{x}(t)] \\
\geq \Phi(\mathbf{x}(t)) + \zeta - (t+1) \zeta = \Phi(\mathbf{x}(t)) - t \zeta = \mathbb{E} [X_t] ,
\]

thus \( X_1, X_2, \ldots, \) is a sub-martingale. Recall that \( \Phi(\mathbf{x}) \) is always bounded by \( 2r \). Denote by \( \tau = |\mathcal{J}_z| \) the time when the process halts. Using Doob’s optional stopping time theorem [BW07] it holds that

\[
\mathbb{E} [\Phi(\mathbf{x}(\tau))] - \mathbb{E} [\tau] \cdot \zeta = \mathbb{E} [X_\tau] \geq \mathbb{E} [X_0] = 0
\]

Implying \( \mathbb{E} [\tau] \leq \frac{1}{\zeta} \cdot \mathbb{E} [\Phi(\mathbf{x}(\tau))] = \frac{2r}{\zeta} \cdot e^\alpha \). \( \square \)

Finally we are ready to bound the probability that \( z \) joins to a cluster of too large radius. Let \( \Psi_z \) be the event that \( z \) joins a cluster \( S_i \), such that at the time that \( z \) joins, \( R_i > \alpha \) (formally, \( z \in S_i \) and \( d_{G_i}(z, T_i) > \alpha \Delta \)). Set \( \mathcal{T}_z = \{ T_i \mid d_{G_i}(z, T_i) > \alpha \Delta \text{ and } d_{G_i}(z, S_i) \leq \alpha \Delta \} \). Denote by \( \mathcal{C}_i \) the event that \( z \in S_i \) and \( T_i \in \mathcal{T}_i \). Recall that \( \mathcal{J}_z = \{ T_i \mid d_{G_i}(z, S_i) \leq \alpha \Delta \} \), thus \( \mathcal{T}_z \subseteq \mathcal{J}_z \). Hence by Lemma 5, \( \mathbb{E} [\mathcal{T}_z] \leq \mathbb{E} [\| \mathcal{J}_z \|] \leq \frac{2r}{\alpha + r} \cdot e^\alpha \). We conclude,

\[
\mathbb{P} [\Psi_z] = \mathbb{P} [\bigcup_i \mathcal{C}_i] = \sum_i \mathbb{P} [\mathcal{C}_i] = \sum_i \mathbb{P} [z \in S_i \wedge T_i \in \mathcal{T}_z] \\
= \sum_i \mathbb{P} [z \in S_i \mid T_i \in \mathcal{T}_z] \cdot \mathbb{P} [T_i \in \mathcal{T}_z] \\
\overset{(\ast)}{=} \sum_i e^{-\alpha} \cdot \mathbb{P} [T_i \in \mathcal{T}_z] \\
= e^{-\alpha} \cdot \mathbb{E} [\| \mathcal{T}_z \|] \leq e^{-\alpha} \cdot \frac{2r}{\alpha + r} \cdot e^\alpha = \frac{2r}{\alpha + r} = \frac{1}{4} ,
\]
where in the equality (*) we used the memoryless property of exponential distribution.

Let \( \Upsilon_z \) be the event that \( z \) joined a cluster \( S_i \), such that after the time that \( z \) joins, \( R_i \) increases by additional \( \alpha \) factor (formally, \( z \in S_i \) and \( R_i \cdot \Delta - d_G(z, T_i) > \alpha \Delta \)). Then by the memoryless property of exponential distribution, \( \Pr[\Upsilon_z] = e^{-\alpha} < \frac{1}{2} \). Denote by \( \Phi_z \) the event that \( z \) belong to a cluster with radius greater than \( 2\alpha \Delta \). Clearly \( \Phi_z \subseteq \Psi_z \cup \Upsilon_z \), as if both \( \Psi_z, \Upsilon_z \) did not occur then \( R_i < 2\alpha \). By union bound it follows that \( \Pr[\Phi_z] \leq \Pr[\Psi_z] + \Pr[\Upsilon_z] \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \).

Next we prove the padding property. Let \( \gamma \in (0, \frac{1}{8}) \) be a padding parameter, and set \( B = B_G(z, \gamma \Delta) \). First we argue that in a single partition executed using Algorithm 1, the ball \( B \) is fully contained in a single cluster with probability at least \( e^{-2\gamma} \). Indeed consider the first index \( i \) such that \( B \cap S_i \neq \emptyset \), and let \( u \) be the closest vertex to \( T_i \) (that is \( u = \arg \min_{v \in V} d_G(v, T_i) \)). As \( B \subseteq G_i \), for every \( v \in B \) it holds that \( d_G(u, v) \leq 2\gamma \cdot \Delta \). Hence by the memoryless property of exponential distribution

\[
\Pr[B \subseteq P(z)] = \Pr[B \subseteq S_i \mid u \in S_i] \\
\geq \Pr[R_i \geq d_G(T_i, u) + 2\gamma \mid R_i \geq d_G(T_i, u)] = \Pr[R_i \geq 2\gamma] = e^{-2\gamma} .
\]

Recall that our algorithm returns a partition \( \tilde{S} \), which constructed by first executing Algorithm 1 to obtain a partition \( S \), and then recursively partitioning each cluster \( S_i \in S \) with radius larger than \( 2\alpha \cdot \Delta \). Let \( \tilde{P}(z) \) be the cluster containing \( z \) in the final partition \( \tilde{S} \). Denote by \( \Gamma_i \) the event that \( z \) participated in at least \( i \) recursive calls, where in the first \( i - 1 \) recursive calls the ball \( B \) was contain in a single cluster, and in the \( i \)'th recursive call, the ball \( B \) was cut. Then \( \Pr[\Gamma_i] \leq \frac{1}{2i} \cdot (1 - e^{-2\gamma}) \), as for \( \Gamma_i \) to occur, \( z \) must be clustered \( i - 1 \) times in a cluster with radius greater than \( 2\alpha \cdot \Delta \) (each occurring with probability at most \( \frac{1}{2} \)), and in the \( i \)'th iteration the ball \( B \) must be cut (which happens with probability at most \( 1 - e^{-2\gamma} \)). We conclude

\[
\Pr \left[ B_G(z, \gamma \Delta) \not\subseteq \tilde{P}(z) \right] \leq \sum_{i \geq 1} \Pr[\Gamma_i] \leq \sum_{i \geq 1} \frac{1}{2i} \cdot (1 - e^{-2\gamma}) \leq 2 \cdot 2\gamma \leq 1 - e^{-8\gamma} .
\]

To conclude we constructed \( r \)-core partitions with radius \( 2\alpha \cdot \Delta \) such that for every vertex \( z \in V \) and \( \gamma \in (0, \frac{1}{8}) \) it holds that \( \Pr \left[ B_G(z, \gamma \Delta) \subseteq \tilde{P}(z) \right] \geq e^{-8\gamma} \). Lemma 3 now follows by scaling.

6 General Graphs

Bartal [Bar96] showed that every \( n \)-point metric space is \( O(\log n) \)-decomposable. In particular, for large enough constant \( c \), and \( \gamma = \frac{1}{c} \), a pair of vertices at distance \( \gamma \Delta \) will be clustered together with probability \( n^{-\frac{1}{8}} \). As the padding parameter governs the exponent in the success probability, it is important to optimize the constant \( c \). Awerbuch and Peleg [AP90] showed that for \( k \in \mathbb{N} \), general \( n \)-vertex graphs admit a strong \( (4k - 2, O(k \cdot n^{\frac{1}{k}})) \) sparse cover scheme. Specifically, [AP90] gave a deterministic construction of \( O(k \cdot n^{\frac{1}{k}}) \) partitions, all strongly \( \Delta \)-bounded and such that every vertex is \( 4k - 2 \)-padded in one of these partitions. It follows that if one samples a single partition from [AP90] uniformly at random, then every vertex is \( 4k - 2 \)-padded with probability at least \( \Omega(\frac{1}{k} \cdot n^{-\frac{1}{k}}) \).

In this section we attempt to optimize the ratio trade-off of ThProbabilistic partitions. We will prove that every \( n \)-vertex weighted graph admits a strong \((2k, \frac{1}{2} \cdot n^{-\frac{1}{k+1}})\)-ThProbabilistic decomposition scheme (Theorem 4), and weak \((2k, n^{-\frac{1}{k}})\)-ThProbabilistic decomposition scheme (Theorem 5). Finally, we show that assuming Erdős girth conjecture [Erd64], for every (integer) \( k \geq 1 \), and (real) \( t < 2k + 1 \), if every \( n \) point metric space admits a weak \((t, p)\)-ThProbabilistic decomposition scheme, then \( p = \)
\(O(n^{-\frac{1}{k}})\). It follows that the success probability in Theorem 5 cannot be (substantially) improved. However, note that it might be possible to improve the stretch parameter to \(2k - 1\) (instead of \(2k\)) while having the same success probability. Closing this gap between \(2k - 1\) to \(2k\), and constructing strong ThProbabilistic decomposition that will match the performance of the weak ThProbabilistic decompositions, are two intriguing open problems.

6.1 Strong Diameter for General Graphs

\textbf{Theorem 4.} For every (real) \(k \geq 1\), every \(n\) point weighted graph \(G = (V, E, w)\) admits a strong \((2k, \frac{1}{k} \cdot n^{-\frac{1}{k-1}})\)-ThProbabilistic decomposition scheme.

\textbf{Proof.} Consider an \(n\) point weighted graph \(G = (V, E, w)\). We will show that \(G\) admits a strong \((2k, 2^{-\frac{1}{k-1}} \cdot n^{-\frac{1}{k-1}})\)-ThProbabilistic decomposition scheme. For \(k \geq 3\), \(2^{-\frac{1}{k-1}} \geq \frac{1}{8}\), so the theorem follows. For \(k < \frac{3}{2}\), \(n^{-\frac{1}{k-1}} < n^{-2}\). Note that every graph admits a strong \((1, n^{-2})\)-ThProbabilistic decomposition scheme.\(^8\) Hence there is nothing to prove.

By scaling, it is enough to construct a ThProbabilistic decomposition for \(\Delta = 2\). We will use a classic ball carving (ala [Bar96]). Fix \(\lambda = \frac{k}{k-1} \cdot \ln(2n)\), and let \(\mathcal{D}\) be exponential distribution with parameter \(\lambda\). The density function is \(\lambda \cdot e^{-\lambda x}\) for \(x \geq 0\). Note that by our choice of \(\lambda\), it holds that

\[ e^{-\frac{\lambda}{k}} = 2n \cdot e^{-\lambda} \]

We create a partition \(\mathcal{P}\) as follows, initially \(Y_1 = V\) is all the unclustered vertices. At step \(i\), after we created the clusters \(C_1, C_2, \ldots, C_{i-1}\), the unclustered vertices are \(Y_i = V \setminus (\cup_{j<i} C_j)\). Pick an arbitrary vertex \(x_i \in Y_i\), and a radius \(r_i \sim \mathcal{D}\). Set the new cluster to be \(C_i = B_{G[Y_i]}(x_i, r_i)\) the ball of radius \(r_i\) around \(x_i\) in the graph induced by the unclustered vertices. The process halts once \(Y_i = \emptyset\).

Consider a pair \(u, v\) such that \(d_G(u, v) = \frac{1}{2k} \cdot \Delta = \frac{1}{k}\). We first argue that \(\Pr[P(x) = P(y)] \geq e^{-\frac{\lambda}{k}}\). Let \(Q\) be some shortest path in \(G\) from \(x\) to \(y\). Let \(i\) be the first index such that \(C_i \cap Q \neq \emptyset\). In particular, \(Q \subseteq Y_i\) and hence \(d_{G[Y_i]}(u, v) = d_G(u, v)\). Let \(z \in Q\) be the vertex closest to \(x_i\) (that is \(z = \arg \min_{s \in Q} d_{G[Y_i]}(x_i, s)\)). Hence \(d_{G[Y_i]}(x_i, z) \leq r_i\). By triangle inequality, for every \(s \in Q\), \(d_{G[Y_i]}(x_i, s) - d_{G[Y_i]}(x_i, z) \leq d_{G[Y_i]}(z, s) \leq \frac{1}{k}\). Using the memoryless property of exponential distribution we conclude

\[
\Pr[P(u) = P(v) \mid Q \subseteq Y_i \text{ and } Q \cap C_i \neq \emptyset] \\
\geq \Pr_{r_i \sim \mathcal{D}} \left[ r_i \geq \max \{ d_{G[Y_i]}(x_i, v), d_{G[Y_i]}(x_i, u) \} \right] \\
= \Pr_{r_i \sim \mathcal{D}} \left[ r_i \geq \max \{ d_{G[Y_i]}(x_i, v) - d_{G[Y_i]}(x_i, z), d_{G[Y_i]}(x_i, u) - d_{G[Y_i]}(x_i, z) \} \right] \\
\geq \Pr_{r_i \sim \mathcal{D}} \left[ r_i \geq \frac{1}{k} \right] = e^{-\frac{\lambda}{k}} .
\]

If all the radii sampled in the process will be at most 1, our partition will be strongly 2-bounded. The probability of a single radius of being larger then 1 is \(e^{-\lambda}\). By union bound the probability that some radius is larger then 1 is at most \(n \cdot e^{-\lambda}\). Hence the probability that the partition is 2 bounded is at

\(^8\)Consider the following random partition: pick u.a.r. a pair of vertices \(u, v\) at distance at most \(\Delta\). \(\mathcal{P}\) will consist of a set containing the shortest path from \(u\) to \(v\), and all the remaining vertices will be in singleton clusters. Clearly this is a strong \((1, n^{-2}, \Delta)\)-ThProbabilistic decomposition.
least $1 - n \cdot e^{-\lambda}$. We will condition the resulting partition to be 2 bounded. Using the law of total probability, the probability that $P(u) = P(v)$ now is:

$$\Pr[P(u) = P(v) \mid \mathcal{P} \text{ is 2-bounded}] = \frac{\Pr[P(u) = P(v)]}{\Pr[\mathcal{P} \text{ is 2-bounded}]} - \frac{\Pr[\mathcal{P} \text{ is not 2-bounded}]}{\Pr[\mathcal{P} \text{ is 2-bounded}]} \cdot \Pr[P(u) = P(v) \mid \mathcal{P} \text{ is not 2-bounded}]$$

$$(*) \Rightarrow \Pr[P(u) = P(v)] - \Pr[\mathcal{P} \text{ is not 2-bounded}] \geq e^{-\frac{k}{2}} - n \cdot e^{-\lambda} = \frac{1}{2} \cdot e^{-\frac{k}{2}} = \frac{1}{2} \cdot (2n)^{-\frac{k}{2}} = 2^{-\frac{k}{2}} \cdot n^{-\frac{1}{k - 1}}.$$ 

where inequality $(*)$ follows as $\Pr[\mathcal{P} \text{ is 2-bounded}] < 1$ and $\Pr[P(u) = P(v) \mid \mathcal{P} \text{ is not 2-bounded}] \leq 1$.

\[\square\]

### 6.2 Weak diameter for General Metrics

Here we present a different clustering scheme with improved parameters, however the diameter guarantee will be only weak. The algorithm here is ball carving as well, however instead of sampling i.i.d. random radii for arbitrarily chosen centers, we pick a single random radius, and the order of centers is chosen randomly. This approach was introduced by Cálinecu, Karloff and Rabani [CKR04] in the context of the 0-extension problem, then Fakcharoenphol, Rao, and Talwar [FRT04] used it to construct stochastic embedding into ultrametrics (alternatively: creating random hierarchical partitions). In particular, one can deduce from [FRT04] analysis an $(O(\log n), O(1))$-ThProbabilistic decomposition scheme. This approach was also the first to construct weak padded decompositions for doubling metrics [GKL03]. Finally, Blelloch, Gu, and Sun [BGS17] used such clustering to construct Ramsey trees. Furthermore, one can deduce from [BGS17] analysis an $(O(k), O(n^{-\frac{1}{k}}))$-ThProbabilistic decomposition scheme. Our contribution here is tighter analysis, improving the stretch constant from $O(k)$ to $2k$.

**Theorem 5.** For every (integer) $k \geq 1$, every $n$ point metric space $(X, d_X)$ admits a weak $(2k, n^{-\frac{1}{k}})$-ThProbabilistic decomposition scheme.

**Proof.** Pick u.a.r. a radius $r \in \{\frac{1}{k}, \frac{2}{k}, \ldots, \frac{k}{k}\}$, and a random permutation $\pi = \{x_1, x_2, \ldots, x_n\}$ over the metric points. Each point $x \in X$ joins the cluster of the first center w.r.t. $\pi$ at distance at most $r \cdot \frac{\Delta}{2}$ from $x$. Formally:

$$C_i = B_X(x_i, r \cdot \frac{\Delta}{2}) \setminus \bigcup_{j<i}B_X(x_j, r \cdot \frac{\Delta}{2}).$$

As a result we obtain a $\Delta$ bounded partition. Fix a pair of points $u, v$ such that $d_X(u, v) \leq \frac{1}{2k} \cdot \Delta$. Let $A_s = \{x \in X \mid d_X(x, \{u, v\}) \leq \frac{s}{k} \cdot \frac{\Delta}{2}\}$ be the set of points at distance at most $\frac{s}{k} \cdot \frac{\Delta}{2}$ from either $u$ or $v$. Then $A_0 = \{u, v\}$. Suppose that $r = \frac{s}{k}$, and let $x_i$ be the vertex with minimal index such that $d_X(\{u, v\}, x_i) \leq \frac{s}{k} \cdot \frac{\Delta}{2}$. Then $u$ and $v$ will not join the clusters $C_1, \ldots, C_{i-1}$, and at least one of them will join $C_i$. Assume w.l.o.g. that $d_X(u, x_i) \leq d_X(v, x_i)$, and suppose farther that $x_i \in A_{s-1}$. By the triangle inequality it follows that $d_X(v, x_i) \leq d_X(v, u) + d_X(u, x_i) \leq \frac{1}{k} \cdot \frac{\Delta}{2} + \frac{s-1}{k} \cdot \frac{\Delta}{2} = \frac{s}{k} \cdot \frac{\Delta}{2}$. Hence both $u, v$ will join the cluster of $x_i$. Using the law of total probability we conclude

$$\Pr[P(u) = P(v)] \geq \frac{1}{k} \sum_{s=1}^{k} \Pr[P(u) = P(v) \mid r = \frac{s}{k}] \geq \frac{1}{k} \sum_{s=1}^{k} \frac{|A_{s-1}|}{|A_s|} \geq \left( \prod_{s=1}^{k} \frac{|A_{s-1}|}{|A_s|} \right)^{\frac{1}{k}} = \left( \frac{|A_0|}{|A_k|} \right)^{\frac{1}{k}} \geq n^{-\frac{1}{k}},$$

20
Proof. Given a metric space \( (X, d_X) \), an \((a, b)\)-gap-distance-oracle, is a data structure that given a pair of points \( x, y \) returns yes if \( d_X(x, y) \leq a \), no if \( d_X(x, y) > b \), and can return either yes or no if \( d_X(x, y) \in (a, b) \). Thorup and Zwick [TZ05] showed that assuming Erdős girth conjecture [Erd64], a \( t \)-distance oracle (a data structure that returns \( t \)-approximation of the distance), must have size \( \Omega(n^{1+\frac{1}{t}}) \). Implicitly, their proof implies also that a gap distance oracle requires \( \Omega(n^{1+\frac{1}{t}}) \) space. Later, we will show how using ThProbabilistic decomposition one can construct a gap distance oracle, and conclude the theorem.

Claim 5. [TZ05], implicit] Assuming Erdős girth conjecture, for every (real) \( t < 2k + 1 \), there is an \( n \)-vertex unweighted graph \( G = (V, E) \) such that every \( (1, t) \)-gap-distance oracle has space \( \Omega(n^{1+\frac{1}{t}}) \).

This claim follows implicitly from [TZ05]. We provide a proof for the sake of completeness.

Proof. The girth of a graph is the length of its shortest cycle. Erdős girth conjecture [Erd64] states that there is a graph \( G = (V, E) \) with girth \( 2k + 2 \) and \( \Omega(n^{1+\frac{1}{t}}) \) edges.\(^9\) Let \( G \) be all the subgraph of \( G \). Consider two different subgraphs \( G_1, G_2 \). Then there is an edge \( \{u, v\} \) that w.l.o.g. belongs to \( G_1 \) but not \( G_2 \). It follows that \( d_{G_1}(u, v) \geq 2k + 1 \), as otherwise \( G \) will contain a cycle with at most \( 2k + 1 \) edges. Let \( D_1, D_2 \) be \( (1, t) \)-gap distance oracles for \( G_1, G_2 \) respectively. Given the query \( (u, v) \), \( D_1 \) will return yes, while \( D_2 \) will return no. In particular \( D_1 \neq D_2 \). As there are \( 2|E| \) different subgraphs, there are at least \( 2^{|E|} \) different gap distance oracles. It follows that there is some graph \( G' \in G \), such that the space of every gap distance oracle \( D \) of \( G' \), is at least \( \log 2^{|E|} = |E| = \Omega(n^{1+\frac{1}{t}}) \).

Claim 6. If an \( n \)-point metric space admits \( (t, p, \Delta) \)-ThProbabilistic decomposition for \( t > 1 \), then it has a \((\frac{\Delta}{t}, \Delta)\)-gap distance oracle with \( \tilde{O}(n \cdot p) \) space.

Proof. Set \( s = p^{-1} \cdot 2 \log n \), and let \( P_1, P_2, \ldots, P_s \) be random partition drawn from the \((t, p, \Delta)\)-ThProbabilistic decomposition. Note that all the clusters in all the partitions have diameter at most \( \Delta \). Consider a pair \( u, v \) such that \( d_X(u, v) \leq \frac{\Delta}{t} \). In every partition \( P_i \), \( u \) and \( v \) belong to the same cluster with probability at least \( p \). The probability that in no partition \( u \) and \( v \) are clustered together is thus at most \((1 - p)^s < e^{-ps} = n^{-2} \). By union bound, as there are at most \( \binom{n}{2} \) pairs at distance at most \( \frac{\Delta}{t} \), with probability at least \( \frac{1}{2} \) every such pair is clustered together in some partition. We thus can assume that we drawn partitions with this property.

For our gap distance oracle we will simply store all the \( s \) partitions. Given a query \( u, v \), we will return yes if there is a cluster in some partition containing both \( u, v \). Clearly, given a pair \( u, v \) such that \( d_X(u, v) \leq \frac{\Delta}{t} \), we will return yes as we insured that there is a cluster in one of the partitions containing both \( u, v \). Otherwise, if \( d_X(u, v) > \Delta \) then as all the partitions are \( \Delta \)-bounded, no cluster will contain both \( u, v \) and we will return no.

\(^9\)Often in the literature the conjecture is referred to as stating that there is a graph \( G \) with girth \( 2k + 1 \) and \( \Omega(n^{1+\frac{1}{t}}) \) edges. However, as every graph contains a bipartite graph with at least half the edges, the conjecture implies a graph with girth at least \( 2k + 2 \).
Storing a single partition takes $\tilde{O}(n)$ space, and hence the total size of our gap distance oracle is $\tilde{O}(n \cdot s) = \tilde{O}(n \cdot p)$.

Using Claim 5 and Claim 6, Theorem 6 follows.

7 Applications

Applying Observation 1 on Corollary 1 and Theorem 3 we conclude,

**Corollary 2.** Let $G$ be a weighted graph and $\Delta > 0$ some parameter.

- If $G$ excludes $K_r$ as a minor, it admits a strong $(O(r), \Delta)$-probabilistic decomposition scheme.
- If $G$ has doubling dimension $\text{ddim}$, it admits a strong $(O(\text{ddim}), \Delta)$-probabilistic decomposition scheme.

7.1 Approximation for Unique Games on Minor Free Graphs

In the *Unique Games* problem we are give a graph $G = (V, E)$, an integer $k \geq 1$ and a set of permutations $\Pi = \{\pi_{uv}\}_{uv \in E}$ on $[k]$ satisfying $\pi_{uv} = \pi_{vu}^{-1}$ (each permutation $\pi_{uv}$ is a bijection from $[k]$ to $[k]$). Given an assignment $x : V \rightarrow [k]$, the edge $uv \in E$ is satisfied if $\pi_{uv}(x(u)) = x(v)$. The problem is to find an assignment that maximizes the number of satisfied edges. The Unique Games Conjecture of Khot [Kho02] postulates that it is NP-hard to distinguish whether a given instance of unique games is almost satisfiable or almost unsatisfiable. The unique games conjecture was thoroughly studied. The conjecture has numerous implications.

Alev and Lau [AL17] studied a special case of the unique games problem, where the graph $G$ is $K_r$ minor free. Given an instance $(G, \Pi)$ where the optimal assignment violates $\epsilon$-fraction of the edge constraints, Alev and Lau used an LP-based approach to efficiently find an assignment that violates at most $O(\sqrt{\epsilon \cdot r})$-fraction. Specifically, in the rounding step of their LP, they used strong diameter probabilistic decompositions with parameter $O(r^2)$. Using instead our decompositions from Corollary 2 with parameter $O(r)$ we obtain a quadratic improvement in the dependence on $r$.

**Theorem 7.** Consider an instance $(G, \Pi)$ of the unique games problem, where the graph $G$ is $K_r$ minor free. Suppose that the optimal assignment violates at most an $\epsilon$-fraction of the edge constraints. There is an efficient algorithm that find an assignment that violates at most an $O(\sqrt{\epsilon \cdot r})$-fraction.

7.2 Spanner for Graphs with Moderate Doubling Dimension

Given a weighted graph $G = (V, E, w)$, a weighted graph $H = (V, E_H, w_H)$ is a $t$-spanner of $G$, if for every pair of vertices $v, u \in V$, $d_G(v, u) \leq d_H(v, u) \leq t \cdot d_X(v, u)$. If in addition $H$ is a subgraph of $G$ (that is $E_H \subseteq E$ and $w_H$ agrees with $w$ on $E_H$) then $H$ is a graph spanner. The factor $t$ is called the *stretch* of the spanner. The number of edges $|E_H|$ is the *sparsity* of the spanner. The weight of $H$ is $w_H(H) = \sum_{e \in E_H} w_H(e)$ the sum of its edge weights. The *lightness* of $H$ is $\frac{w_H(H)}{w(MST(G))}$ the ratio between the weight of the spanner to the weight of the MST of $G$. The tradeoff between stretch and sparsity/lightness of spanners had been the focus of an intensive research effort, and low stretch graph spanners are used in a plethora of applications.
There is an extensive study of spanners for doubling metrics. Recently, for an $n$-vertex graph with doubling dimension $\ddim$, Borradaile, Le and Wulff-Nilsen [BLW19] contrasted a graph spanner with $1+\epsilon$ stretch, $\epsilon^{-O(\ddim)}$ lightness and $n\cdot \epsilon^{-O(\ddim)}$ sparsity (improving [Smi09, Got15, FS20]). This result is also asymptotically tight. Note that the dependency on $\ddim$ is exponential, which is unavoidable for small, $1+\epsilon$ stretch. In cases where $\ddim$ is moderately large (say $\sqrt{\log n}$), it might be preferable to accept larger stretch in order to obtain small lightness.

In a recent work, Filtser and Neiman [FN22], for every stretch parameter $t \geq 1$, constructed a spanner with stretch $O(t)$, lightness $O(2^{\ddim} \cdot t \cdot \log^2 n)$ and $O(n \cdot 2^{\ddim} \cdot \log n \cdot \log t)$ edges. However, this spanner was not a subgraph. Most applications require a graphic spanner. It is possible to transform [FN22] into a graphic spanner, but the number of edges becomes unbounded. The spanner construction of [FN22] is based on a variant of probabilistic decompositions, where they used a weak-diameter version. If we replaced this with our strongly padded decompositions Corollary 1, and plug this into Theorem 3 from [FN22], we obtain a spanner with the same stretch to lightness ratio, but also with an additional sparsity guarantee.

**Corollary 3.** Let $G = (V, E, w)$ be an $n$ vertex graph, with doubling dimension $\ddim$ and aspect ratio $\Lambda = \frac{\max_{e \in E} w(e)}{\min_{e \in E} w(e)}$. Then for every parameter $t > 1$ there is a graph-spanner of $G$ with stretch $O(t)$, lightness $O(2^{\ddim} \cdot t \cdot \log^2 n)$ and $O(n \cdot 2^{\ddim} \cdot \log n \cdot \log \Lambda)$ edges.

### 7.3 Path Reporting Distance Oracles

Given a weighted graph $G = (V, E, w)$, a **distance oracle** is a data structure that supports distance queries between pairs $u, v \in V$. The distance oracle has stretch $t$, if for every query $\{u, v\}$, the estimated distance $\text{est}(u, v)$ is within $d_G(u, v)$ and $t \cdot d_G(u, v)$. The studied objects are stretch, size the query time. An additional requirement that been recently studied [EP16] is path reporting: in addition to distance estimation, the distance oracle should also return a path of the promised length. In this case, we say that distance oracle has query time $q$, if answering a query when the reported path has $m$ edges, takes $q + O(m)$ time.

Path reporting distance oracles were studied for general graphs [EP16, ENW16]. For the special case of graphs excluding $K_r$ as a minor, Elkin, Neiman and Wulff-Nilsen [ENW16] constructed a path reporting distance oracles with stretch $O(r^2)$, space $O(n \cdot \log \Lambda \cdot \log n)$ and query time $O(\log \log \Lambda)$, where $\Lambda = \frac{\max_{u, v} d_G(u, v)}{\min_{u, v} d_G(u, v)}$ is the aspect ratio. For this construction they used the strongly padded decomposition of [AGG+19] (in fact strong sparse covers). Implicitly, given a graph $G$ that admits a strong $(\beta, s)$ sparse cover scheme, [ENW16] constructs a path reporting distance oracle with stretch $\beta$, size $O(n \cdot s \cdot \log \beta \Lambda)$ and query time $O(\log \log \Lambda)$. Following similar arguments to [ENW16] (taking $O(\log n)$ independent copies and using union bound), our padded decompositions from Theorem 3 implies that every $K_r$ minor free graph admits a strong $(O(r), O(\log n))$ sparse cover scheme. We conclude:

**Corollary 4.** Given an $n$-vertex weighted graph $G = (V, E, w)$ which excludes $K_r$ as a minor, with aspect ratio $\Lambda$, there is a path reporting distance oracle with stretch $O(r)$, space $O(n \cdot \log_r \Lambda \cdot \log n)$ and query time $O(\log \log \Lambda)$.

For the case of graphs with doubling dimension $\ddim$, we constructed the first strong sparse covers. Plugging our Theorem 2 into the framework of [ENW16], we obtain the first path reporting distance oracle for doubling graphs. The only relevant previous distance oracle for doubling metrics is by Bartal et al. [BGK+11]. However, they focused on the $1+\epsilon$-stretch regime, where inherently the oracle size has exponential dependency on $\ddim$. 


Corollary 5. Given an $n$-vertex weighted graph $G = (V, E, w)$ with doubling dimension $ddim$ and aspect ratio $\Lambda$, for every parameter $t \geq \Omega(1)$, there is a path reporting distance oracle with stretch $O(t)$, space $O(n \cdot 2^{ddim/t} \cdot ddim \cdot \log \Lambda)$ \footnote{This is assuming $\Lambda > \log t$, otherwise simply using an arbitrary shortest path tree will provide a distance oracle with stretch $O(\log t)$.} and query time $O(\log \log \Lambda)$.

In particular, there is a path reporting distance oracle with stretch $O(ddim)$, space $O(n \cdot ddim \cdot \log \Lambda)$ and query time $O(\log \log \Lambda)$.

8 Conclusion and Open Problems

In this paper we closed the gap left in [AGG+19] between the padding parameters of strong and weak padded decompositions for minor free graphs. Our second contribution is tight strong padded decomposition scheme for graphs with doubling dimension $ddim$, which we also use to create sparse cover schemes. Finally, we addressed the most basic question of strong and weak ThProbabilistic decompositions for general graphs, optimizing the stretch parameter. Some open questions remain:

1. Prove/disprove that $K_r$ minor free graphs admit strong/weak decompositions with padding parameter $O(\log r)$, as conjectured by [AGG+19].

2. The question above is already open for the more restricted family of treewidth $r$ graphs.

3. The $\delta$ parameter: [AGG+19] constructed weak $(O(r), \Omega(1))$-padded decomposition scheme, while we constructed strong $(O(r), \Omega(\frac{1}{r}))$-padded decomposition scheme. It will be nice to construct strong $(O(r), \Omega(1))$-padded decomposition scheme. Such a decomposition will imply a richer spectrum of sparse covers (with $o(r)$ stretch).

4. Sparse covers for $K_r$ minor free graphs: [AGMW10] constructed strong $(O(r^2), 2^r (r+1)!)$ sparse cover scheme, while [BLT14] constructed strong $(4, f(r) \cdot \log n)$ sparse cover scheme. An interesting open question is to create additional sparse cover schemes. Specifically, our padded decompositions suggest that an $(O(r), g(r))$-strong sparse cover scheme for some function $g$ independent of $n$, should be possible. Currently it is unclear how to construct such a cover. Optimally, we would like to construct $(O(1), g(r))$-sparse cover scheme.

5. Close the gap between the lower and upper bounds of weak ThProbabilistic decompositions for general metrics. Specifically show that general $n$-point metrics admit $(2k−1, n^{−\frac{1}{k}})$-ThProbabilistic decomposition scheme, or the impossibility of such a scheme.

6. Close the gap between weak to strong stochastic decompositions for general graphs. Specifically, construct strong $(2k, n^{−\frac{1}{k}})$-stochastic decompositions scheme, or show the impossibility of such a scheme.

9 Acknowledgments

The author would like to thank Ofer Neiman for helpful discussions.
References

[ABN11] I. Abraham, Y. Bartal, and O. Neiman. Advances in metric embedding theory. *Advances in Mathematics*, 228(6):3026 – 3126, 2011, doi:https://doi.org/10.1016/j.aim.2011.08.003, 1, 2, 3, 5

[ACE+20] I. Abraham, S. Chechik, M. Elkin, A. Filtser, and O. Neiman. Ramsey spanning trees and their applications. *ACM Trans. Algorithms*, 16(2):19:1–19:21, 2020, doi:10.1145/3371039, 1, 4

[AGG+19] I. Abraham, C. Gavoille, A. Gupta, O. Neiman, and K. Talwar. Cops, robbers, and threatening skeletons: Padded decomposition for minor-free graphs. *SIAM J. Comput.*, 48(3):1120–1145, 2019, doi:10.1137/17M112406, 1, 2, 3, 4, 5, 12, 13, 23, 24

[AGGM06] I. Abraham, C. Gavoille, A. V. Goldberg, and D. Malkhi. Routing in networks with low doubling dimension. In *26th IEEE International Conference on Distributed Computing Systems (ICDCS 2006), 4-7 July 2006, Lisboa, Portugal*, page 75, 2006, doi:10.1109/ICDCS.2006.72, 1, 2

[AGMW10] I. Abraham, C. Gavoille, D. Malkhi, and U. Wieder. Strong-diameter decompositions of minor free graphs. *Theory Comput. Syst.*, 47(4):837–855, 2010, doi:10.1007/s00224-010-9283-6, 2, 3, 4, 5, 12, 13, 23, 24

[AL17] V. L. Alev and L. C. Lau. Approximating unique games using low diameter graph decomposition. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2017, August 16-18, 2017, Berkeley, CA, USA*, pages 18:1–18:15, 2017, doi:10.4230/LIPIcs.APPROX-RANDOM.2017.18, 3, 22

[AN19] I. Abraham and O. Neiman. Using petal-decompositions to build a low stretch spanning tree. *SIAM Journal on Computing*, 48(2):227–248, 2019, arXiv:https://doi.org/10.1137/17M1115575, doi:10.1137/17M1115575, 4

[AP90] B. Awerbuch and D. Peleg. Sparse partitions (extended abstract). In *31st Annual Symposium on Foundations of Computer Science, St. Louis, Missouri, USA, October 22-24, 1990, Volume II*, pages 503–513, 1990, doi:10.1109/SFCS.1990.89571, 2, 3, 18

[Bar96] Y. Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science, FOCS ’96, Burlington, Vermont, USA, 14-16 October, 1996*, pages 184–193, 1996, doi:10.1109/SFCS.1996.548477, 1, 2, 3, 5, 18, 19

[BCF+23] C. Busch, D. Q. Chen, A. Filtser, D. Hathcock, D. E. Hershkowitz, and R. Rajaraman. One tree to rule them all: Poly-logarithmic universal steiner tree. *CoRR*, abs/2308.01199, 2023. To appear in FOCS 2023, arXiv:2308.01199, doi:10.48550/ARXIV.2308.01199, 4

[BDR+12] C. Busch, C. Dutta, J. Radhakrishnan, R. Rajaraman, and S. Srivathsan. Split and join: Strong partitions and universal steiner trees for graphs. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 81–90, 2012, doi:10.1109/FOCS.2012.45, 4

[BFT23] S. Bhore, A. Filtser, and C. D. Tóth. Online duet between metric embeddings and minimum-weight perfect matchings. *CoRR*, abs/2310.14078, 2023. To appear in SODA 2024, arXiv:2310.14078, doi:10.48550/ARXIV.2310.14078, 1, 5

[BGK+11] Y. Bartal, L. Gottlieb, T. Kopelowitz, M. Lewenstein, and L. Roditty. Fast, precise and dynamic distance queries. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011*, pages 840–853, 2011, doi:10.1137/1.9781611973082.66, 23

[BGK+14] G. E. Blelloch, A. Gupta, I. Koutis, G. L. Miller, R. Peng, and K. Tangwongsan. Nearly-linear work parallel SDD solvers, low-diameter decomposition, and low-stretch subgraphs. *Theory Comput. Syst.*, 55(3):521–554, 2014, doi:10.1007/s00224-013-9444-5, 1
[BGS17] G. E. Blelloch, Y. Gu, and Y. Sun. Efficient construction of probabilistic tree embeddings. In I. Chatzigiannakis, P. Indyk, F. Kuhn, and A. Muscholl, editors, 44th International Colloquium on Automata, Languages, and Programming, ICALP 2017, July 10-14, 2017, Warsaw, Poland, volume 80 of LIPIcs, pages 26:1–26:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017, doi:10.4230/LIPIcs.ICALP.2017.26. 20

[BLR10] P. Biswal, J. R. Lee, and S. Rao. Eigenvalue bounds, spectral partitioning, and metrical deformations via flows. J. ACM, 57(3):13:1–13:23, 2010, doi:10.1145/1706590.1706593. 1

[BLT14] C. Busch, R. LaFortune, and S. Tirthapura. Sparse covers for planar graphs and graphs that exclude a fixed minor. Algorithmica, 69(3):658–684, 2014, doi:10.1007/s00453-013-9757-4. 2, 24

[BLW19] G. Borradaile, H. Le, and C. Wulff-Nilsen. Greedy spanners are optimal in doubling metrics. In Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019, pages 2371–2379, 2019, doi:10.1137/1.9781611975482.145. 23

[BW07] R. Bhatnacharya and E. C. Waymire. A basic course in probability theory. Analysis, 2007, doi:10.1007/978-0-387-71939-9. 17

[CFJ+23] A. Czumaj, A. Filtser, S. H. C. Jiang, R. Krauthgamer, P. Veselý, and M. Yang. Streaming facility location in high dimension via geometric hashing, 2023, arXiv:2204.02095. 4

[CKR04] G. Călinescu, H. J. Karloff, and Y. Rabani. Approximation algorithms for the 0-extension problem. SIAM J. Comput., 34(2):358–372, 2004, doi:10.1137/S0097539701395978. 1, 20

[Erd64] P. Erdős. Extremal problems in graph theory. Theory of Graphs and Its Applications (Proc. Sympos. Smolenice), pages 29–36, 1964. see here. 3, 18, 21

[EN18] M. Elkin and O. Neiman. Efficient algorithms for constructing very sparse spanners and emulators. ACM Trans. Algorithms, 15(1):4:1–4:29, November 2018, doi:10.1145/3274651. 4

[ENW16] M. Elkin, O. Neiman, and C. Wulff-Nilsen. Space-efficient path-reporting approximate distance oracles. Theor. Comput. Sci., 651:1–10, 2016, doi:10.1016/j.tcs.2016.07.038. 4, 23

[EP16] M. Elkin and S. Pettie. A linear-size logarithmic stretch path-reporting distance oracle for general graphs. ACM Trans. Algorithms, 12(4):50:1–50:31, 2016, doi:10.1145/2888397. 23

[Fil19] A. Filtser. On strong diameter padded decompositions. In D. Achlioptas and L. A. Végh, editors, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2019, September 20-22, 2019, Massachusetts Institute of Technology, Cambridge, MA, USA, volume 145 of LIPIcs, pages 6:1–6:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019, doi:10.4230/LIPICS.APPROX-RANDOM.2019.6. 1

[Fil20] A. Filtser. Scattering and sparse partitions, and their applications. In A. Czumaj, A. Dawar, and E. Merelli, editors, 47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference), volume 168 of LIPIcs, pages 47:1–47:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, doi:10.4230/LIPIcs.ICALP.2020.47. 4
[JLN+05] L. Jia, G. Lin, G. Noubir, R. Rajaraman, and R. Sundaram. Universal approximations for TSP, steiner tree, and set cover. In Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005, pages 386–395, 2005, doi:10.1145/1060590.1060649. 4

[Kho02] S. Khot. On the power of unique 2-prover 1-round games. In Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada, pages 767–775, 2002, doi:10.1145/509907.510017. 22

[KK17] L. Kamma and R. Krauthgamer. Metric decompositions of path-separable graphs. Algorithmica, 79(3):645–653, 2017, doi:10.1007/s00453-016-0213-0. 4

[KKLMN04] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: A new embedding method for finite metrics. In 45th Symposium on Foundations of Computer Science (FOCS 2004), 17-19 October 2004, Rome, Italy, Proceedings, pages 434–443, 2004, doi:10.1109/FOCS.2004.41. 1

[KLPT09] J. A. Kelner, J. R. Lee, G. N. Price, and S. Teng. Higher eigenvalues of graphs. In 50th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2009, October 25-27, 2009, Atlanta, Georgia, USA, pages 735–744. IEEE Computer Society, 2009, doi:10.1109/FOCS.2009.69. 1

[KPR93] P. N. Klein, S. A. Plotkin, and S. Rao. Excluded minors, network decomposition, and multicommodity flow. In Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing, May 16-18, 1993, San Diego, CA, USA, pages 682–690, 1993, doi:10.1145/167088.167261. 1, 3

[KL05] J. R. Lee and A. Naor. Extending lipschitz functions via random metric partitions. Inventiones Mathematicae, 160(1):59–95, 2005, doi:10.1007/s00222-004-0400-5. 1

[LR99] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. J. ACM, 46:787–832, November 1999, doi:http://doi.acm.org/10.1145/331524.331526. 1

[LS10] J. R. Lee and A. Sidiropoulos. Genus and the geometry of the cut graph. In Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010, pages 193–201, 2010, doi:10.1137/1.9781611973075.18. 4

[Mat02] J. Matoušek. Lectures on discrete geometry. Springer-Verlag, New York, 2002, doi:10.1007/978-1-4613-0039-7. 1

[MN07] M. Mendel and A. Naor. Ramsey partitions and proximity data structures. Journal of the European Mathematical Society, 9(2):253–275, 2007, doi:10.1109/FOCS.2006.65. 1

[MPX13] G. L. Miller, R. Peng, and S. C. Xu. Parallel graph decompositions using random shifts. In G. E. Blelloch and B. Vöcking, editors, 25th ACM Symposium on Parallelism in Algorithms and Architectures, SPAA ’13, Montreal, QC, Canada - July 23 - 25, 2013, pages 196–203. ACM, 2013, doi:10.1145/2486159.2486180. 4, 5, 8

[MT10] R. A. Moser and G. Tardos. A constructive proof of the general lovász local lemma. J. ACM, 57(2):11:1–11:15, 2010, doi:10.1145/1667053.1667060. 11

[Rab03] Y. Rabinovich. On average distortion of embedding metrics into the line and into L1. In Proceedings of the 35th Annual ACM Symposium on Theory of Computing, June 9-11, 2003, San Diego, CA, USA, pages 456–462, 2003, doi:10.1145/780542.780609. 1

[Rao99] S. Rao. Small distortion and volume preserving embeddings for planar and euclidean metrics. In Proceedings of the fifteenth annual symposium on Computational geometry, SCG ’99, pages 300–306, New York, NY, USA, 1999. ACM, doi:10.1145/304893.304983. 1

[RR98] Y. Rabinovich and R. Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. Discret. Comput. Geom., 19(1):79–94, 1998, doi:10.1007/PL00009336. 14

[RS03] N. Robertson and P. D. Seymour. Graph minors. XVI. excluding a non-planar graph. J. Comb. Theory, Ser. B, 89(1):43–76, 2003, doi:10.1016/S0095-8956(03)00042-X. 2
[Smi09] M. H. M. Smid. The weak gap property in metric spaces of bounded doubling dimension. In Efficient Algorithms, Essays Dedicated to Kurt Mehlhorn on the Occasion of His 60th Birthday, pages 275–289, 2009, doi:10.1007/978-3-642-03456-5_19.

[TZ05] M. Thorup and U. Zwick. Approximate distance oracles. J. ACM, 52(1):1–24, 2005, doi:10.1145/1044731.1044732.