Electronic Colloquium on Computational Complexity, Report No. 8 (2021)

Random walks and forbidden minors III: poly$(d\varepsilon^{-1})$-time partition oracles for minor-free graph classes

Akash Kumar[*] C. Seshadhri[†] Andrew Stolman[‡]

Abstract

Consider the family of bounded degree graphs in any minor-closed family (such as planar graphs). Let d be the degree bound and n be the number of vertices of such a graph. Graphs in these classes have hyperfinite decompositions, where, for a sufficiently small $\varepsilon > 0$, one removes εdn edges to get connected components of size independent of n. An important tool for sublinear algorithms and property testing for such classes is the *partition oracle*, introduced by the seminal work of Hassidim-Kelner-Nguyen-Onak (FOCS 2009). A partition oracle is a local procedure that gives consistent access to a hyperfinite decomposition, without any preprocessing. Given a query vertex v, the partition oracle outputs the component containing v in time independent of n. All the answers are consistent with a single hyperfinite decomposition.

The partition oracle of Hassidim et al. runs in time $d^{\mathrm{poly}(d\varepsilon^{-1})}$ per query. They pose the open problem of whether poly$(d\varepsilon^{-1})$-time partition oracles exist. Levi-Ron (ICALP 2013) give a refinement of the previous approach, to get a partition oracle that runs in time $d^{\log(d\varepsilon^{-1})}$-per query.

In this paper, we resolve this open problem and give poly$(d\varepsilon^{-1})$-time partition oracles for bounded degree graphs in any minor-closed family. Unlike the previous line of work based on combinatorial methods, we employ techniques from spectral graph theory. We build on a recent spectral graph theoretical toolkit for minor-closed graph families, introduced by the authors to develop efficient property testers. A consequence of our result is a poly$(d\varepsilon^{-1})$-query tester for *any* property of minor-closed families (such as bipartite planar graphs). Our result also gives poly$(d\varepsilon^{-1})$-query algorithms for additive εn-approximations for problems such as maximum matching, minimum vertex cover, maximum independent set, and minimum dominating set for these graph families.

[*] Department of Computer Science, EPFL. akash.kumar@epfl.ch

[†] Department of Computer Science, University of California, Santa Cruz. sesh@ucsc.edu

[‡] Department of Computer Science, University of California, Santa Cruz. astolman@ucsc.edu

CS and AS acknowledge the support of NSF grants CCF-1740850, CCF-1813165, CCF-1909790, CCF-2023495, and ARO Award W911NF1910294.

ISSN 1433-8092

1 Introduction

The algorithmic study of planar graphs is a fundamental direction in theoretical computer science and graph theory. Classic results like the Kuratowski-Wagner characterization [Kur30, Wag37], linear time planarity algorithms [HT74], and the Lipton-Tarjan separator theorem underscore the significance of planar graphs [LT80]. The celebrated theory of Robertson-Seymour give a grand generalization of planar graphs through minor-closed families [RS95a, RS95b, RS04]. This has led to many deep results in graph algorithms, and an important toolkit is provided by separator theorems and associated decompositions [AST94].

Over the past decade, there have been many advances in *sublinear* algorithms for planar graphs and minor-closed families. We focus on the model of random access to bounded degree adjacency lists, introduced by Goldreich-Ron [GR02]. Let $G = (V, E)$ be a graph with vertex set $V = [n]$ and degree bound d. The graph is accessed through *neighbor queries*: there is an oracle that on input $v \in V$ and $i \in [d]$, returns the ith neighbor of v. (If none exist, it returns $\perp$.)

One of the key properties of bounded-degree graphs in minor-closed families is that they exhibit hyperfinite decompositions. A graph G is hyperfinite if $\forall\ 0 < \varepsilon < 1$, one can remove εdn edges from G and obtain connected components of size independent of n (we refer to these as pieces). For minor-closed families, one can remove εdn edges and get pieces of size $O(\varepsilon^{-2})$.

The seminal result of Hassidim-Kelner-Nguyen-Onak (HKNO) [HKNO09] introduced the notion of *partition oracles*. This is a local procedure that provides "constant-time" access to a hyperfinite decomposition. The oracle takes a query vertex v and outputs the piece containing v. Each piece is of size independent of n, and at most εdn edges go between pieces. Furthermore, all the answers are consistent with a single hyperfinite decomposition, despite there being no preprocessing or explicit coordination. (All queries uses the same random seed, to ensure consistency.) Partition oracles are extremely powerful as they allow a constant time procedure to directly access a hyperfinite decomposition. As observed in previous work, partition oracles lead to a plethora of property testing results and sublinear time approximation algorithms for minor-closed graph families [HKNO09, NS13]. In some sense, one can think of partition oracles as a moral analogue of Széмeredi's regularity lemma for dense graph property testing: it is a decomposition tool that immediately yields a litany of constant time (or constant query) algorithms.

We give a formal definition of partition oracles. (We deviate somewhat from the definition in Chap. 9.5 of Goldreich's book [Gol17] by including the running time as a parameter, instead of the set size.)

Definition 1.1. *Let $\mathcal{P}$ be a family of graphs with degree bound d and $T : (0, 1) \to \mathbb{N}$ be a function. A procedure $\boldsymbol{A}$ is an $(\varepsilon, T(\varepsilon))$*-partition oracle *for $\mathcal{P}$ if it satisfies the following properties. The deterministic procedure takes as input random access to $G = (V, E)$ in $\mathcal{P}$, random access to a random seed r (of length polynomial in graph size), a proximity parameter $\varepsilon > 0$, and a vertex v of G. (We will think of fixing G, r, ε, so we use the notation $\boldsymbol{A}_{G,r,\varepsilon}$. All probabilities are with respect to r.) The procedure $\boldsymbol{A}_{G,r,\varepsilon}(v)$ outputs a set of vertices and satisfies the following properties.*

1. *(Consistency) The sets $\{\boldsymbol{A}_{G,r,\varepsilon}(v)\}$, over all v, form a partition of V. Also, these sets $\boldsymbol{A}_{G,r,\varepsilon}(v)$ induce connected graphs for all $v \in V$.*
2. *(Cut bound) With probability (over r) at least $2/3$, the number of edges between the sets $\boldsymbol{A}_{G,r,\varepsilon}(v)$ is at most εdn.*
3. *(Running time) For every v, $\boldsymbol{A}_{G,r,\varepsilon}(v)$ runs in time $T(\varepsilon)$.*

We stress that there is no explicit "coordination" or sharing of state between calls to $\boldsymbol{A}_{G,r,\varepsilon}(v)$ and $\boldsymbol{A}_{G,r,\varepsilon}(v')$ (for $v \neq v'$). There is no global preprocessing step once the random seed is fixed. The consistency guarantee holds with probability 1. Note that the running time $T(\varepsilon)$ is clearly an upper bound on the size of the sets $\boldsymbol{A}_{G,r,\varepsilon}(v)$. For minor-closed families, one can convert any partition oracle to one that output sets of size $O(\varepsilon^{-2})$ with a constant factor increase in the cut bound. (refer to the end of Sec. 9.5 in [Gol17]).

The challenge in partition oracles is to bound the running time $T(\varepsilon)$. HKNO gave a partition oracle with running time $(d\varepsilon^{-1})^{\text{poly}(d\varepsilon^{-1})}$. Levi-Ron [LR15] built on the ideas from HKNO and dramatically improved the bound to $(d\varepsilon^{-1})^{\log(d\varepsilon^{-1})}$. Yet, for all minor-closed families, one can (in linear time) remove εdn edges to get connected components of size $O(\varepsilon^{-2})$. HKNO raise the natural open question as to whether $(\varepsilon, \text{poly}(d\varepsilon^{-1}))$-partition oracles exist.

In this paper, we resolve this open problem.

Theorem 1.2. *Let $\mathcal{P}$ be the set of d-bounded degree graphs in a minor-closed family. There is an $(\varepsilon, \text{poly}(d\varepsilon^{-1}))$-partition oracle for $\mathcal{P}$.*

1.1 Consequences

As observed by HKNO and Newman-Sohler [NS13], partition oracles have many consequences for property testing and sublinear algorithms.

Recall the definition of property testers. Let $\mathcal{Q}$ be a property of graphs with degree bound d. The distance of G to $\mathcal{Q}$ is the minimum number of edge additions/removals required to make G have $\mathcal{Q}$, divided by dn. A property tester for $\mathcal{P}$ is a randomized procedure that takes query access to an input graph G and a proximity parameter, $\varepsilon > 0$. If $G \in \mathcal{P}$, the tester accepts with probability at least 2/3. If the distance of G to $\mathcal{Q}$ is at least ε, the tester rejects with probability at least 2/3. We often measure the query complexity as well as time complexity of the tester.

A direct consequence of Theorem 1.2 is an "efficient" analogue of a theorem of Newman-Sohler stating that all properties of hyperfinite graphs are testable.

Theorem 1.3. *Let $\mathcal{Q}$ be any property of bounded degree graphs of a minor-closed family. There exists a $\text{poly}(d\varepsilon^{-1})$-query tester for $\mathcal{Q}$.*

If membership in $\mathcal{Q}$ can be determined exactly in polynomial (in input size) time, then $\mathcal{Q}$ has $\text{poly}(d\varepsilon^{-1})$-time testers.

An appealing consequence of Theorem 1.3 is that the property of bipartite planar graphs can be tested in $\text{poly}(d\varepsilon^{-1})$ time. For any fixed subgraph H, the property of H-free planar graphs can be tested in the same time. And all of these bounds hold for any minor-closed family.

As observed by Newman-Sohler, partition oracles give sublinear query algorithms for any graph parameter that is "robust" to edge changes. Again, Theorem 1.2 implies an efficient version for minor-closed families.

Theorem 1.4. *Let f be a real-valued function on graphs that changes by $O(1)$ on edge addition/removals, and has the property that $f(G_1 \cup G_2) = f(G_1) + f(G_2)$ for graphs G_1, G_2 that are not connected to each other.*

For any minor-closed family $\mathcal{P}$, there is a randomized algorithm that, given $\varepsilon > 0$ and $G \in \mathcal{P}$, outputs an additive εn-approximation to $f(G)$ and makes $\text{poly}(d\varepsilon^{-1})$ queries. If f can be computed exactly in polynomial time, then the above algorithm runs in $\text{poly}(d\varepsilon^{-1})$ time.

The functions captured by Theorem 1.4 are quite general. Functions such as maximum matching, minimum vertex cover, maximum independent set, minimum dominating set, maxcut, etc. all have the robustness property. As a compelling application of Theorem 1.4, we can get $(1+\varepsilon)$-approximations[1] for the maximum matching in planar (or any minor-closed family) graphs in $\text{poly}(d\varepsilon^{-1})$ time.

These theorems are easy consequences of Theorem 1.2. Using the partition oracle, an algorithm can essentially assume that the input is a collection of connected components of size

[1] The maximum matching is $\Omega(n/d)$ for a connected bounded degree graph. One simply sets $\varepsilon \ll 1/d$ in Theorem 1.4.

poly$(d\varepsilon^{-1})$, and run an exact algorithm on a collection of randomly sampled components. We give formal proofs in Sec. 8.

1.2 Related work

The subject of property testing and sublinear algorithms in bounded degree graphs is a vast topic. We refer the reader to Chapters 9 and 10 of Goldreich's textbook [Gol17]. We focus on the literature relevant to sublinear algorithms for minor-closed families.

The first step towards a characterization of testable properties in the bounded-degree model was given by Czumaj-Sohler-Shapira, who showed hereditary properties in non-expanding graphs are testable [CSS09]. This was an indication that notions like hyperfiniteness are connected to property testing. Benjamini-Schramm-Shapira achieved a breakthrough by showing that all minor-closed properties are testable, in time triply-exponential in $d\varepsilon^{-1}$ [BSS08]. Hassidim-Kelner-Nguyen-Onak introduced partition oracles, and designed one running in time exp$(d\varepsilon^{-1})$. Levi-Ron improved this bound to quasipolynomial in $d\varepsilon^{-1}$, using a clever analysis inspired by algorithms for minimum spanning trees [LR15]. Newman-Sohler built on partition oracles for minor-close families to show that all properties of hyperfinite graphs are testable [NS13]. Fichtenberger-Peng-Sohler showed any testable property contains a hyperfinite property [FPS19].

There are two dominant combinatorial ideas in this line of work. The first is using subgraph frequencies in neighborhood of radius poly(ε^{-1}) to characterize properties. This naturally leads to exponential dependencies in poly(ε^{-1}). The second idea is to use random edge contractions to reduce the graph size. Recursive applications lead to hyperfinite decompositions, and the partition oracles of HKNO and Levi-Ron simulate this recursive procedure. This is extremely non-trivial, and leads to a recursive local procedure with a depth dependent of ε. Levi-Ron do a careful simulation, ensuring that the recursion depth is at most $\log(d\varepsilon^{-1})$, but this simulation requires looking at neighborhoods of radius $\log(d\varepsilon^{-1})$. Following this approach, there is little hope of getting a recursion depth independent of ε, which is required for a poly$(d\varepsilon^{-1})$-time procedure.

Much of the driving force behind this work was the quest for a poly$(d\varepsilon^{-1})$-time tester for planarity. This question was resolved recently using a different approach from spectral graph theory, which was itself developed for sublinear time algorithms for finding minors [KSS18, KSS19]. A major inspiration is the random walk based one-sided bipartiteness tester of Goldreich-Ron [GR99]. This paper is a continuation of that line of work, and is a further demonstration of the power of spectral techniques for sublinear algorithms. The tools build on local graph partitioning techniques pioneered by Spielman-Teng [ST12], which is itself based on classic mixing time results of Lovász-Simonovits [LS90]. In this paper, we develop new diffusion-based local partitioning tools that form the core of partition oracles.

We also mention other key results in the context of sublinear algorithms for minor-closed families, notably the Czumaj et al [CGR$^+$14] upper bound of $O(\sqrt{n})$ for testing cycle minor-freeness, the Fichtenberger et al [FLVW17] upper bound of $O(n^{2/3})$ for testing $K_{2,r}$-minor-freeness, and poly$(d\varepsilon^{-1})$ testers for outerplanarity and bounded treewidth graphs [YI15, EHNO11].

2 Main Ideas

The starting point for this work are the spectral methods used in [KSS18, KSS19]. These methods discover cut properties within a neighborhood of radius poly$(d\varepsilon^{-1})$, without explicitly constructing the entire neighborhood.

One of the key tools used in these results in a local partitioning algorithm, based on techniques of Spielman-Teng [ST12]. The algorithm takes a seed vertex s, performs a diffusion from s (equivalently, performs many random walks) of length poly$(d\varepsilon^{-1})$, and tracks the diffusion vector to detect a low conductance cut around s in poly$(d\varepsilon^{-1})$ time. We will use the term *diffusions*, instead of random walks, because we prefer the deterministic picture of a unit of "ink" spreading through the graph. A key lemma in previous results states that, for graphs in minor-closed families, this procedure succeeds from more than $(1-\varepsilon)n$ seed vertices. This yields a global algorithm to construct a hyperfinite decomposition with components of $\text{poly}(d\varepsilon^{-1})$ size. Pick a vertex s at random, run the local partitioning procedure to get a low conductance cut, remove and recurse. Can there be a local implementation of this algorithm?

Let us introduce some setup. We will think of a global algorithm that processes seed vertices in some order. Given each seed vertex s, a local partitioning algorithm generates a low conductance set $C(s)$ containing s (this is called a cluster). The final output is the collection of these clusters. For any vertex v, let the *anchor* of v be the vertex s such that $v \in C(s)$. A local implementation boils down to finding the anchor of query vertex v.

Observe that at any point of the global procedure, some vertices have been clustered, while the remaining are still *free*. The global procedure described above seems hopeless for a local implementation. The cluster $C(s)$ is generated by diffusion in some subgraph G' of G, which was the set of free vertices when seed s was processed. Consider a local procedure trying to discover the anchor of v. It would need to figure out the free set corresponding to every potential anchor s, so that it can faithfully simulate the diffusion used to cluster v. From an implementation standpoint, it seems that the natural local algorithm is to use diffusions from v in G to discover the anchor. But diffusion in a subgraph G' is markedly different from G and difficult to simulate locally. Our first goal is to design a partitioning method using diffusions directly in G.

Finding low conductance cuts in subsets, by diffusion in supersets: Let us now modify the global algorithm with this constraint in mind. At some stage of the global algorithm, there is a set F of free vertices. We need to find a low conductance cut contained in F, while running random walks in G. Note that we must be able to deal with F as small as $O(\varepsilon n)$. Thus, random walks (even starting from F) will leave F quite often; so how can these walks/diffusions find cuts in F?

One of our main insights is that these challenges can be dealt with, even for diffusions of $\text{poly}(d\varepsilon^{-1})$ length. We show that, for a uniform random vertex $s \in F$, a spectral partitioning algorithm that performs diffusion from s in G can detect low conductance cuts contained in F. Diffusion in the superset (all of V) provides information about the subset F. This is a technical and non-trivial result, and crucially uses the spectral properties of minor-closed families. Note that diffusions from F can spread very rapidly in short random walks, even in planar graphs. Consider a graph G, where F is a path on εn vertices, and there is a tree of size $1/\varepsilon$ rooted at every vertex of F. Diffusions from any vertex in F will initially be dominated by the trees, and one has to diffuse for at least $1/\varepsilon$ timesteps before structure within F can be detected. Thus, the proof of our theorem has to look at average behavior over a sufficiently large time horizon before low conductance cuts in F are "visible". Remarkably, it suffices to look at $\text{poly}(d\varepsilon^{-1})$ timesteps to find structure in F, because of the behavior of diffusions in minor-closed families.

The main technical tool used is the Lovász-Simonovits curve technique [LS90], whose use was pioneered by Spielman-Teng [ST12]. We also use the truncated probability vector technique from Spielman-Teng to give cleaner implementations and proofs. A benefit of using diffusion (instead of random walks) on truncated vectors is that the clustering becomes deterministic.

The problem of ordering the seeds: With one technical hurdle out of the way, we end up at another gnarly problem. The above procedure only succeeds if the seed is in F. Quite naturally, one does not expect to get any cuts in F by diffusing from a random vertex in G. From the perspective of the global algorithm, this means that we need some careful ordering of the seeds, so that low conductance cuts are discovered. Unfortunately, we also need local implementations of this ordering. The authors struggled with carrying out this approach, but to no avail.

To rid ourselves of the ordering problem, let us consider the following, almost naive global algorithm. First, order the vertices according to a uniform random permutation. At any stage, there is a free set F. We process the next seed vertex s by running some spectral partitioning procedure, to get a low conductance cut $C(s)$. Simply output $C(s) \cap F$ (instead of $C(s)$) as the new cluster, and update F to $F \setminus C(s)$. It is easy to locally implement this procedure. To find the anchor of v, perform a diffusion of $\text{poly}(\varepsilon^{-1})$ timesteps from v. For every vertex s with high enough value in the diffusion vector, determine if $C(s) \ni v$. The vertex s that is lowest according to the random ordering is the anchor of v. Unfortunately, there is little hope of bounding the number of edges cut by the clustering. When s is processed, it may be that $s \notin F$, and there is no guarantee of $C(s) \cap F$. Can we modify the procedure to bound the number of cut edges, but still maintain its ease of local implementability?

The amortization argument: Consider the scenario when $F = \Theta(\varepsilon n)$. Most of the subsequent seeds processed are not in F and there is no guarantee on the cluster conductance. But every $\Theta(1/\varepsilon)$ seeds (in expectation), we will get a "good" seed s contained in F, such that $C(s) \cap F$ is a low conductance set. (This is promised by the diffusion algorithm that we develop in this paper, as discussed earlier.) Our aim is to perform some amortization, to argue that $|C(s) \cap F|$ is so large, that we can "charge" away the edges cut by the previous $\Theta(1/\varepsilon)$ seeds.

This amortization is possible because our spectral tools give us much flexibility in the (low) conductances obtained. Put differently, we essentially prove that existence of many cuts of extremely low conductance, and show that it is "easy" for a diffusion-based algorithm to find such cuts. (This is connected to the spectral behavior of minor-closed families.) As a consequence, we can actually pre-specify the size of the low conductance cuts obtained. We show that as long as $|F| = \Omega(\varepsilon n)$, we can find a *size threshold* $k = \text{poly}(\varepsilon^{-1})$ such that for at least $\Omega(\varepsilon^2 n)$ vertices $s \in F$, a spectral partitioning procedure seeded at s can find a cut of size $\Theta(k)$ and conductance at most ε^c. Moreover, this cut is guaranteed to contain at least $\varepsilon^{c'}k$ vertices in F, despite the procedure being oblivious to F. The parameter c can be easily tuned, so we can increase c arbitrarily while keeping c' fixed, at the cost of polynomial increases in running time. This tunability is crucial to our amortization argument. We also show that given query access to F, a size threshold k can be computed in $\text{poly}(d\varepsilon^{-1})$ time.

So when the global algorithm processes seed s, it runs the above spectral procedure to try to obtain a set of size $\Theta(k)$ with conductance at most ε^c. (If the procedure fails, the global algorithm simply set $C(s) = \{s\}$.) Thus, we cut $O(\varepsilon^c kd)$ edges for each seed processed. But after every $O(1/\varepsilon)$ seeds, we choose a "good" seed such that $|C(s) \cap F| > \varepsilon^{c'}k$. The total number of edges cut is $O(\varepsilon^c kd \times \varepsilon^{-1}) = O(\varepsilon^{c-1}kd)$. The total number of new vertices clustered is at least $\varepsilon^{c'}k$. Because we can tune parameters with much flexibility, we can set $c \gg c'$. So the total number of edges cut is $O(\varepsilon^{c-c'-1}d)$ times the number of vertices clustered, where $c - c' - 1 > 1$. Overall, we will cut only $O(\varepsilon nd)$ edges.

Making it work through phases: Unfortunately, as the process described above continues, F shrinks. Thus, the original choice of k might not work, and the guarantees on $|C(s) \cap F|$ for good seeds no longer hold. So we need to periodically recompute the value of k. In a careful analysis, we show that this recomputation is only required $\text{poly}(\varepsilon^{-1})$ times. Formally, we implement the recomputation through *phases*. Each vertex is independently assigned to one of $\text{poly}(\varepsilon^{-1})$ phases. (Technically, we choose the phase of a vertex by sampling an independent geometric random variable. We heavily use the memoryless property of the geometric distribution.)

For each phase, the value of k is fixed. The local partition oracle will compute these size thresholds for all phases, as a $\text{poly}(d\varepsilon^{-1})$ time preprocessing step. The oracle (for v) runs a diffusion from v to get a collection of candidate anchors. For each candidate s, the oracle determines its phase, runs the spectral partitioning algorithm with correct phase parameters, and determines if the candidate's low conductance cut contains v. The anchor is simply such a candidate of minimum phase, with ties broken by vertex id.

2.1 Outline of sections

The algorithm description and proof has many moving parts, encapsulated by different sections. Sec. 3 begins by discussing the truncated diffusion process, the main algorithmic tool for partitioning. We then describe the global partitioning algorithm `globalPartition` (modulo a preprocessing step called `findr`), which is far more convenient to analyze. It will be readily apparent that this global procedure outputs a partition of G into connected components; the main challenge is to bound the number of edges cut.

Within Sec. 3, we discuss how to implement `globalPartition` by a local procedure. By ensuring that the output of the local procedure is identical to `globalPartition`, we prove the consistency property of Definition 1.1. We then perform a fairly straightforward running time analysis, which proves the running time property of Definition 1.1.

The real heavy lifting begins in Sec. 4, where we describe the procedure `findr` that computes the size thresholds. This section is devoted to proving salient properties of the size thresholds output by `findr`. The analysis hinges on the diffusion and cut properties stated in Theorem 4.1, which is the main tool connecting minor-freeness, diffusions, and local partitioning. Sec. 5 uses all these tools to prove the cut bound of `globalPartition`. At this stage, the complete description and guarantees of the partition oracle are complete, modulo the proof of Theorem 4.1.

The proof of Theorem 4.1 is split into sections. In Sec. 6, we use the hyperfiniteness of minor-closed families to prove properties of truncated diffusions on minor-free families. Sec. 7 has the key spectral calculations, where the Lovász-Simonovits curve technique is used to find low conductance cuts. This section has the crucial insights that allow for partitioning in the free set, using diffusions in the overall graph.

Sec. 8 has short proofs of the applications Theorem 1.3 and Theorem 1.4. These are provided for completeness, since identical calculations appear in the proof of Theorem 9.28 in [Gol17].

3 Global partitioning and its local implementation

There are a number of parameters that are used in the algorithm. We list them out here for reference. It is convenient to fix the value of ε in advance, so that all the values of the following parameters are fixed. Note that all these parameters are polynomial is ε. We will express all running times as polynomials in these parameters, ensuring all running time are $\text{poly}(\varepsilon^{-1})$.

- $\rho = d^{-60}\varepsilon^{3000}$: Minimum probability for truncation.
- $\ell = d^6\varepsilon^{-30}$: Maximum random walk length.
- $\beta = \varepsilon/10$: Unclustered fraction cutoff.
- $\delta = d^{-70}\varepsilon^{3100}$: Phase probability.
- $\alpha = \frac{\varepsilon^{4/3}}{300{,}000}$: Heavy bucket parameter.
- $\phi = \varepsilon^{10}$: Conductance parameter.

3.1 Truncated diffusion

The main process used to find sets of the partition is a *truncated diffusion*. We assume that the input graph G is connected, has n vertices, and degree bound d. Define the symmetric random walk matrix M as follows. For every edge (u, v), $M_{u,v} = M_{v,u} = 1/2d$. For every vertex v, $M_{v,v} = 1 - d(v)/2d$, where $d(v)$ is the degree of v. The matrix M is doubly stochastic, symmetric, and the (unique) stationary distribution is the uniform distribution.

Given a vector $\vec{x} \in (\mathbb{R}^+)^n$, diffusion is the evolution $M^t\vec{x}$. We define a truncated version, where after every step, small values are removed. For any vector $\vec{x}$, let $\text{supp}(\vec{x})$ denote the support of the vector.

Definition 3.1. *Define the operator $\widehat{M} \colon (\mathbb{R}^+)^n \to (\mathbb{R}^+)^n$ as follows. For $\vec{x} \in (\mathbb{R}^+)^n$, the vector $\widehat{M}\vec{x}$ is obtained by zeroing out all coordinates in $M\vec{x}$ whose value is at most ρ.*

For $t > 1$, the operator $\widehat{M}^t$ is the t-step truncated diffusion, and is recursively defined as $\widehat{M}(\widehat{M}^{t-1}\vec{x})$.

Define $\widehat{p}_{v,t}(w)$ to be the coordinate corresponding to vertex w in the t-step truncated diffusion starting from vertex v.

We stress that the t-step truncated diffusion is obtained from a standard diffusion by truncating low values at *every* step of the diffusion. Note that as the truncated diffusion progresses, the l_1-norm of the vector may decrease at each step. Importantly, for any distribution vector $\vec{x}$, $\mathrm{supp}(\widehat{M}^t\vec{x})$ has size at most ρ^{-1}. We heavily use this property in our running time analysis.

We define *level sets*, a standard concept in spectral partitioning algorithms. Somewhat abusing notation, for vertex $v \in V$, we use $\vec{v}$ to denote the unit vector in $(\mathbb{R}^+)^n$ corresponding to the vertex v. (We never use to vector notation for any other kind of vectors.)

Definition 3.2. *For vertex $v \in V$, length t, and threshold k, let $L_{v,t,k}$ be the set of vertices corresponding to the k largest coordinates in $\widehat{M}^t\vec{v}$ (ties are broken by vertex id).*

For any set S of vertices, the conductance of S is $\Phi(S) := E(S,\overline{S})/[2\min(|S|,|\overline{S}|)d]$. (We use $E(S,\overline{S})$ to denote the number of edges between S and its complement.)

We describe the key subroutine that finds low conductance cuts. It performs a sweep cut over the truncated diffusion vector.

```
cluster($(v, t, k)$)
  1. Determine $\widehat{M}^t\vec{v}$
  2. For all $k' \in [k, 2k]$ calculate $\Phi(L_{v,t,k'})$.
  3. Find the largest $k' \in [k, 2k]$ (if any) with the following properties: $\Phi(L_{v,t,k'} \cup \{v\}) \leq \phi$ and
     $L_{v,t,k'} \in \mathrm{supp}(\widehat{M}^t\vec{v})$.
  4. If such a $k'$ exists, set $C := L_{v,t,k'} \cup \{v\}$, else $C := \{v\}$.
  5. Return $C$.
```

Claim 3.3. *The procedure `cluster`(v,t,k) runs in time $O(\rho^{-1}td\log(\rho^{-1}td) + kd\log k)$. The output set C has the following properties. (i) $v \in C$. (ii) If C is not a singleton, then $|C| \in [k, 2k]$, $\Phi(C) \leq \phi$, and $C \subseteq \mathrm{supp}(\widehat{M}^t\vec{v})$.*

Proof. The latter properties are apparent from the description of `cluster`.

We analyze the running time. The convenience of the truncated diffusion is that it can computed exactly by a deterministic process. First, for any $b \geq 1$, we show that the running time to compute $\widehat{M}^b\vec{v}$ is $O(\rho^{-1}bd)$. Note that for any t, $\mathrm{supp}(\widehat{M}^b\vec{v})$ has size at most ρ^{-1}, since M is a stochastic matrix and all non-zero entries in $\widehat{M}^b\vec{v}$ have value at least ρ. Given the vector $\widehat{M}^b\vec{v}$, the vector $\widehat{M}^{b+1}\vec{v}$ can be computed by determining $M\widehat{M}^b\vec{v}$ and then zeroing out coordinates that are less than ρ. This process can be done in $O(d|\mathrm{supp}(\widehat{M}^b\vec{v})|) = O(\rho^{-1}d)$. By summing this running time over all timesteps, we get that the total time is $O(\rho^{-1}bd)$.

Thus, $\widehat{M}^t\vec{v}$ can be computed exactly in $O(\rho^{-1}td)$ time. To compute the level sets, one can sort the coordinates of this vector (breaking ties by id), and process them in decreasing order. One can iteratively store $L_{v,t,k}$ in a dictionary data structure. Given $\Phi(L_{v,t,k})$, one can compute $\Phi(L_{v,t,k+1})$ by $O(d)$ lookups into the dictionary. The total running time of this step is $O(kd\log k)$. □

3.2 The global partitioning procedure

The global partitioning procedure `globalPartition` will output a partition of the vertices satisfying the conditions in Definition 1.1. This global procedure will run in linear time. In the next subsection, we show how the output of the global procedure can be generated locally in $\mathrm{poly}(\varepsilon^{-1})$ time, thereby giving us the desired partition oracle. It will be significantly easier to understand and analyze the partition properties of the global procedure.

The key ingredient in `globalPartition` that allows for a local implementation is a *preprocessing* step. The preprocessing allows for the "coordination" required for consistency of various local partitioning steps. All the randomness is used in the preprocessing, after which the actual partitioning is deterministic. The job of the preprocessing is to find the following sets of values, which are used for two goals: (i) ordering vertices, (ii) setting parameters for calls to `cluster`.

The preprocessing generates, for all vertices v, the following values.

- h_v: The *phase* of v.
- k_v: The size threshold of v.
- t_v: The walk length of v.

Before giving the procedure description, we explain how these values are generated.

Phases: For each v, h_v is set to $\max(X, \overline{h})$, where X is independently sampled from $Geo(\delta)$, the geometric distribution with parameter δ. Moreover $\overline{h} := 2\delta^{-1}\log(\delta^{-1})$, so the maximum phase value is capped.

Size thresholds: The computation of these thresholds is the most complex part of our algorithm (and analysis), and is the "magic ingredient" that makes the partition oracle possible. We first run a procedure `findr` that runs in $\mathrm{poly}(\varepsilon^{-1})$ time and outputs a set of *phase size thresholds* $k_1, k_2, \ldots, k_{\overline{h}}$. All the thresholds have value at most ρ^{-1} and $k_{\overline{h}}$ will be zero. The (involved) description of `findr` and its properties are in Sec. 4. For now, it suffices to say that its running time is $\mathrm{poly}(\varepsilon^{-1})$, and that it outputs phase size thresholds. The size threshold for a vertex v is simply k_{h_v}, corresponding to the phase it belongs to.

Walk lengths: These are simply chosen independently and uniformly in $[1, \ell]$.

The analysis is more transparent when we assume that all the randomness used by the algorithm is in a random seed $\boldsymbol{R}$, of $O(n \cdot \mathrm{poly}(\varepsilon^{-1}))$ length. The seed $\boldsymbol{R}$ is passed as an argument to the partitioning procedure, which uses $\boldsymbol{R}$ to generate all the values described above. (For convenience, we will assume random access to the adjacency list of G, without passing the graph as a parameter.)

It is convenient to define an ordering on the vertices, given these values. For cleaner notation, we drop the dependence on $\boldsymbol{R}$.

Definition 3.4. *For vertex $u, v \in V$, we say that $u \prec v$ if: $h_u < h_v$ or if $h_u = h_v$, the id of v is less than that of v.*

```
globalPartition($\boldsymbol{R}$)
Preprocessing:
    1. For every $v \in V$:
        (a) Use $\boldsymbol{R}$ to set $h_v := \max(X, \overline{h})$ ($X \sim Geo(\delta)$).
        (b) Use $\boldsymbol{R}$ to set $t_v$ uniform random in $[1, \ell]$.
    2. Call findr($\boldsymbol{R}$) to generate values $k_1, k_2, \ldots, k_{\overline{h}}$. For every $v \in V$, set $k_v = k_{h_v}$.
Partitioning:
    1. Initialize the partition $\boldsymbol{P}$ as an empty collection. Initialize the free set $F := V$.
    2. For all vertices in $V$ in increasing order of $\prec$:
        (a) Compute $C =$ cluster($v, t_v, k_v$).
        (b) Add the connected components of $C \cap F$ to the partition $\boldsymbol{P}$.
        (c) Reset $F = F \setminus C$.
    3. Output $\boldsymbol{P}$.
```

Since all of our subsequent discussions are about `globalPartition`, we abuse notation assuming that the preprocessing is fixed. We refer to `cluster`(v) to denote `cluster`(v, t_v, k_v).

These are the only calls to cluster that are ever discussed, so it is convenient to just parametrize by the vertex argument. Furthermore, for ease of notation, we sometimes refer to the output of the procedure as $\texttt{cluster}(v)$.

We observe that the output $\boldsymbol{P}$ is indeed a partition of V into connected components. At any intermediate step, the free set F is precisely the set of vertices that have not been assigned to a cluster. Note that $\texttt{cluster}(v)$ always contains v (Claim 3.3), so all vertices eventually enter (the sets of) $\boldsymbol{P}$.

We note that v might not be in F when $\texttt{cluster}(v)$ is called. This may lead to new components in $\boldsymbol{P}$ that do not involve v, which may actually not be low conductance cuts. This may seem like an oversight: why initiate diffusion clusters from vertices that are already partitioned? Many challenges in our analysis arise from such clusters. On the other hand, such an "oblivious" partitioning scheme leads to a simple local implementation.

3.3 The local implementation

A useful definition in the local implementation is that of *anchors* of vertices. As mentioned earlier, we fix the output of the preprocessing (which is equivalent to fixing $\boldsymbol{R}$).

Definition 3.5. *Consider the running of* $\texttt{globalPartition}(\boldsymbol{R})$. *The* anchor *of w is the (unique) vertex w such that the component in $\boldsymbol{P}$ containing v was created by the call to* $\texttt{cluster}(v)$.

Suppose we label every vertex by its anchor. We can easily determine the sets of $\boldsymbol{P}$ locally.

Claim 3.6. *The sets of $\boldsymbol{P}$ are exactly the maximal connected components of vertices with the same anchor.*

Proof. We prove by induction over the $\prec$ ordering of vertices. The base case is vacuously true. Suppose, just before v is considered, all current sets in $\boldsymbol{P}$ are maximal connected components with the same anchor, which cannot be v. No vertex in F can have an anchor yet; otherwise, it would be clustered and part of (a set in) $\boldsymbol{P}$. All the new vertices clustered have v as anchor. Moreover, the sets added to $\boldsymbol{P}$ are precisely the maximal connected components with v as anchor. □

We come to a critical definition that allows for searching for anchors. We define the "inverse ball" of a vertex: this is the set of all vertices that reach v through truncated diffusions. We note that reachability is not symmetric, because the diffusion is truncated at every step.

Definition 3.7. *For $v \in V$, let $IB(v) = \{w \mid \exists t \in [0, \ell], v \in \text{supp}(\widehat{M}^t \vec{w})\}$.*

Claim 3.8. $|IB(v)| \leq \ell\rho^{-1}$.

Proof. All vertices $w \in IB(v)$ have the property that (for some $t \leq \ell$) $\widehat{p}_{w,t}(v) \neq 0$. That implies that $p_{w,t}(v) \geq \rho$. By the symmetry of the random walk, $p_{v,t}(w) \geq \rho$. For any fixed t, there are at most ρ^{-1} such vertices w. Overall, there can be at most $\ell\rho^{-1}$ vertices in $IB(v)$. □

Now we have a simple characterization of the anchor that allows for local implementations.

Lemma 3.9. *The anchor of v is the smallest vertex (according to $\prec$) in the set $\{s|s \in IB(v)$ and $v \in$ $\texttt{cluster}(s)\}$.*

Proof. Let the anchor of v be the vertex u. We first argue that u in the given set. Clearly, $v \in \texttt{cluster}(u)$. If $u = v$, then $u = v \in IB(v)$ and we are done. Suppsoe $u \neq v$. Then $\texttt{cluster}(u)$ is not a singleton (since it contains v). By Claim 3.3, $\texttt{cluster}(u)$ is contained in the support of $\widehat{M}^{t_v}\vec{u}$, implying that $v \in \text{supp}(\widehat{M}^{t_v}\vec{u})$. Thus, $u \in IB(v)$ and the anchor u is present in the given set.

It remains to argue that u is the smallest such vertex. Suppose there exists $u' \prec u$ in this set. In `globalPartition`, `cluster`(u') is called before `cluster`(u). At the end of this call, v is partitioned and would have u' as its anchor. Contradiction. □

We are set for the local implementation. For a vertex v, we compute $IB(v)$ and run `cluster`(u) for all $u \in IB(v)$. By Lemma 3.9, we can compute the anchor of v, and by Claim 3.6, we can perform a BFS to find all connected vertices with the same anchor.

We begin by a procedure that computes $IB(v)$. Since the truncated diffusion is not symmetric, this requires a little care. We use $N(u)$ to denote the neighborhood of vertex u.

```
findIB($v$)
    1. Initialize $S = \{v\}$.
    2. For every $t = 1, \ldots, \ell$:
        (a) For every $w \in S \cup N(S)$, compute $\widehat{M}^t \vec{w}$. If $v \in \text{supp}(\widehat{M}^t \vec{w})$, add $v$ to $S$.
    3. Return $S$.
```

Claim 3.10. *The output of* `findIB`(v) *is* $IB(v)$. *The running time is* $O(d^2 \ell^3 \rho^{-2})$.

Proof. We prove by induction on t, that after t iterations of the loop, S is the set $\{w \mid \exists t' \in [0,t], v \in \text{supp}(\widehat{M}^{t'} \vec{w})\}$. The base case $t = 0$ holds because S is initialized to $\{v\}$. Now for the induction. Consider some w such that $v \in \text{supp}(\widehat{M}^{t+1} \vec{w})$. This means that $(1-d(w)/2d)\widehat{p}_{w,t}(v) + (1/2d)\sum_{w' \in N(w)} \widehat{p}_{w',t}(v) \geq \rho$. Since the LHS is an average, for some $w' \in N(w) \cap \{w\}$, $\widehat{p}_{w',t}(v) \geq \rho$. Hence, $v \in \text{supp}(\widehat{M}^t \vec{w'})$, and by induction $w' \in S$ at the beginning of the $(t+1)$th iteration. The inner loop will consider w (as it is either w' or a neighbor of w'), correctly determine that $v \in \text{supp}(\widehat{M}^{t+1} \vec{w})$, and add it to S. By construction, every (new) vertex w added to S has the property that $v \in \text{supp}(\widehat{M}^{t+1} \vec{w})$. This completes the induction and the output property.

For the running time, observe that for all iterations, $S \subseteq IB(v)$. By Claim 3.8, $|S| \leq \ell \rho^{-1}$. Hence, $|S \cup N(S)|$ has size $O(d\ell\rho^{-1})$. The computation of each $\widehat{M}^t \vec{w}$ can be done in $O(d\ell\rho^{-1})$ time, since the distribution vector after each step has support size at most ρ^{-1}. The total running time of each iteration is $O(d^2\ell^2\rho^{-2})$. There are at most ℓ iterations, leading to a total running time of $O(d^2\ell^3\rho^{-2})$.

□

We can now describe the local partitioning oracle (modulo the description of `findr`).

```
findAnchor($v, \boldsymbol{R}$)
    1. Run findr($\boldsymbol{R}$) to get the set $K = \{k_1, k_2, \ldots, k_{\overline{h}}\}$.
    2. Run findIB($v$) to compute $IB(v)$.
    3. Initialize $A = \emptyset$.
    4. For every $s \in IB(v)$:
        (a) Using $\boldsymbol{R}$ determine $h_s, t_s$. Using $K$, determine $k_s$.
        (b) Compute $C$ = cluster($s, t_s, k_s$).
        (c) If $C \ni v$, then add $s$ to $A$.
    5. Output the smallest vertex according to $\prec$ in $A$.
```

```
findPartition($v, \boldsymbol{R}$)
    1. Call findAnchor($v, \boldsymbol{R}$) to get the anchor $s$.
    2. Perform BFS from $v$. For every vertex $w$ encountered, first call findAnchor($w, \boldsymbol{R}$). If the
       anchor is $s$, add $w$ to the BFS queue (else, ignore $w$).
    3. Output the set of vertices that entered the BFS queue.
```

The following claim is a direct consequence of Lemma 3.9 and Claim 3.10.

Claim 3.11. *The procedure* `findAnchor`$(v, \boldsymbol{R})$ *outputs the anchor of* v *and runs in time* $O((d\ell\rho^{-1})^3)$ *plus the running time of* `findr`*.*

Proof. Observe that `findAnchor`$(v, \boldsymbol{R})$ finds $IB(v)$, computes `cluster`(s) for each $s \in IB(v)$, and outputs the smallest (by $\prec$) s such that $v \in$ `cluster`(s). By Lemma 3.9, the output is the anchor of v.

By Claim 3.10, the running time of `findIB`(v) is $O(d^2\ell^3\rho^{-2})$. The number of calls to `cluster` is $|IB(v)|$, which is at most $\ell\rho^{-1}$ (Claim 3.8). Each call to `cluster` runs in time $O(d\ell\rho^{-2})$, by Claim 3.3 and the fact that $k_s \leq \rho^{-1}$). Ignoring the call to `findr`, the total running time is $O(d^2\ell^3\rho^{-3})$. □

Theorem 3.12. *The output of* `findPartition`$(v, \boldsymbol{R})$ *is precisely the set in* $\boldsymbol{P}$ *containing* v*, where* $\boldsymbol{P}$ *is the partition output by* `globalPartition`$(\boldsymbol{R})$*. The running time of* `findPartition`$(v, \boldsymbol{R})$ *is* $O((d\ell\rho^{-1})^4)$ *plus the running time of* `findr`*.*

Proof. By Claim 3.11, `findAnchor` correctly outputs the anchor. By Claim 3.6, the set S in $\boldsymbol{P}$ containing v is exactly the maximal connected component of vertices sharing the same anchor (as v). The set S in $\boldsymbol{P}$ is generated in `globalPartition`$(\boldsymbol{R})$ by a call to `cluster`, whose output is a set of size at most ρ^{-1}. The total number of calls to `findAnchor` made by `findPartition`$(v, \boldsymbol{R})$ is at most $d\rho^{-1}$, since a call is made to either a vertex in the set S or a neighbor of S. Overall, the total running time is $O((d\ell\rho^{-1})^5)$ plus the running time of `findr`. (Instead of calling `findr` in each call to `findAnchor`, one can simply store its output.) □

4 Coordination through the size thresholds: the procedure `findr`

We now come to the heart of our algorithm; coordination through `findr`. This section gives the crucial ingredient in arguing that the partitioning scheme does not cut too many edges. The ordering of vertices (to form clusters) is chosen independent of the graph structure. It is highly likely that, as the partitioning proceeds, newer `cluster`(v) sets overlap heavily with the existing partition. Such clusters may cut many new edges, without clustering enough vertices. Note that `cluster`(v) is a low conductance cut only in the original graph; it might have high conductance restricted to F (the current free set).

To deal with such "bad" clusters, we need to prove that every so often, `cluster`(v) will successfully partition enough new vertices. Such "good" clusters allow the partitioning scheme to suffer many bad clusters. This argument is finally carried about by a careful charging argument. First, we need to argue that such good clusters exist. The key tool is given by the following theorem, which is proved using spectral graph theoretic methods. We state the theorem as an independent statement.

Theorem 4.1. *Let* G *be a bounded degree graph in a minor-closed family. Let* F *be an arbitrary set of vertices of size at least* βn*. There exists a size threshold* $k \leq \rho^{-1}$ *such that the following holds. For at least* $(\beta^2/\log^2\beta^{-1})n$ *vertices* $s \in F$*, there are at least* $(\beta/\log^2\beta^{-1})\ell$ *timesteps* $t \leq \ell$ *such that: there exists* $k' \in [k, 2k]$ *such that (i)* $L_{s,t,k'} \subseteq \operatorname{supp}(\widehat{M}^t\vec{s})$*, (ii)* $\Phi(L_{s,t,k'} \cup \{s\}) < \phi$*, and (iii)* $|L_{s,t,k'} \cap F| \geq \beta^3 k$*.*

The proof of this theorem is deferred to Sec. 7. In this section, we apply this theorem to complete the description of the partition oracle and prove its guarantees.

We discuss the significance of this theorem. The diffusion used to define $L_{s,t,k'}$ occurs in G, but we are promised a low conductance cut with non-trivial intersection with F (since $\phi \ll \beta^3$). Moreover, such cuts are obtained for a non-trivial fraction of timesteps, so we can choice one uar. Given oracle access to membership in F, it is fairly easy to find such a size threshold by random sampling.

The importance of phases: Recall the global partitioning procedure globalPartition. We can think of the partitioning process as divided into phases, where the hth phase involves calling $\texttt{cluster}(v, t_v, k_v)$ for all vertices v whose phase value is h. Consider the free set at the beginning of a phase h, denoting it F_h. We apply Theorem 4.1 to determine the size threshold k_h. Since all k_v values in this phases are precisely k_h, this size threshold "coordinates" all clusters in this phase. As the phase proceeds, the free set shrinks, and the size threshold k_h stops satisfying the properties of Theorem 4.1. Roughly speaking, at this point, we start a new phase $h+1$, and recompute the size threshold. The frequency of recomputation is chosen carefully to ensure that the total running time remains $\text{poly}(\varepsilon^{-1})$.

We now discuss the randomness involved in selecting phases and why geometric random variables are used. Recall that h_v is independently (for all v) set to be $\min(X, \overline{h})$, where $X \sim Geo(\delta)$. We first introduce some notation regarding phases.

Definition 4.2. *The* phase h seeds*, denoted V_h, are the vertices whose phase value is h. Formally, $V_h = \{v \mid h_v = h\}$. We use $V_{<h}$ to denote $\bigcup_{h'<h} V_h$. (We analogously define $V_{\leq h}, V_{\geq h}$.)*

The free set at phase h*, denoted F_h, is the free set F in* globalPartition*, just before the first phase h vertex is processed. Formally, $F_h = V \setminus \bigcup_{v \in V_{<h}} \texttt{cluster}(v)$.*

One can think of the V_hs being generated iteratively. Assume that we have fixed the vertices in $V_1, \ldots, V_{h-1}$. All other vertices are in $V_{\geq h}$, implying that $h_v \geq h$ for such vertices. By the properties of the geometric random variables, $\Pr[h_v = h+1 | h_v > h] = \delta$. Thus, we can imagine that V_{h+1} is generated by independently sampling each element in $V_{\geq h}$ with δ probability. We restate this observation as Claim 4.4. Claim 4.5 is a simple Chernoff bound argument.

Before proceeding, we state some standard Chernoff bounds (Theorem 1.1 of [DP09]).

Theorem 4.3. *Let $X_1, X_2, \ldots, X_r$ be independent variables in $[0,1]$. Let $\mu := \mathbf{E}[\sum_i X_i]$.*

- $\Pr[X \geq 3\mu/2] \leq \exp(-\mu/12)$.
- $\Pr[X \leq \mu/2] \leq \exp(-\mu/8)$.
- *For $t \geq 6\mu$, $\Pr[X \geq t] \leq 2^{-t}$.*

Claim 4.4. *For all $v \in V$ and $1 < h < \overline{h}$, $\Pr[v \in V_h \mid v \in V_{\geq h}] = \delta$.*

Claim 4.5. *Let $h < \overline{h}$. Condition on the randomness used to specify $V_1, V_2, \ldots, V_{h-1}$. Let S be an arbitrary subset of $V_{\geq h}$. With probability at least $1 - 2\exp(-\delta|S|/12)$ over the choice of V_h, $|S \cap V_h| \in [\delta|S|/2, 2\delta|S|]$.*

Proof. For every $s \in S$, let X_s be the indicator random variable for $s \in V_h$. By Claim 4.4 and independent phase choices for each vertex, the X_s are independent Bernoullis with δ probability. By the Chernoff lower tail of Theorem 4.3, $\Pr[\sum_{s \in S} X_s \leq \delta|S|/2] \leq \exp(-\delta|S|/8)$ and $\Pr[\sum_{s \in S} X_s \geq 2\delta|S|] \leq \exp(\delta|S|/12)$. A union bound completes the proof. □

Claim 4.6. *With probability at least $1 - 2^{-\delta n}$, $|V_{\overline{h}}| \leq \delta n$.*

Proof. Recall that $\overline{h}$ is the last phase and $\overline{h} = 2\delta^{-1}\log(\delta^{-1})$. The probability that $X \sim Geo(\delta)$ is at least $2\delta^{-1}\log(\delta^{-1})$ is $(1-\delta)^{2\delta^{-1}\log(\delta^{-1})-1} < \delta/6$. Hence, the probability that any vertex lies in $V_{\overline{h}}$ is at most $\delta/6$ and the expectation of $V_{\overline{h}}$ is at most $\delta n/6$. . By the Chernoff bound of Theorem 4.3, $\Pr[|V_{\overline{h}}| \geq \delta n] \leq 2^{-\delta n}$. □

With this preamble, we proceed to the description of findr and the main properties of its output.

4.1 The procedure `findr`

It is convenient to assume that for all v, h_v and t_v have been chosen. These quantities are chosen independently for each vertex using simple distributions, so we will not carry as arguments the randomness used to decide these quantities. Recall that the output of `findr` is the set of size thresholds $\{k_1, k_2, \ldots, k_{\overline{h}}\}$. It is convenient to use K_h to denote $\{k_1, k_2, \ldots, k_h\}$. Before describing `findr`, we define a procedure that is a membership oracle for F_h.

```
IsFree($u, h, K_{h-1}$)
  1. If $h = 1$, output YES.
  2. Run findIB($u$) to determine $IB(u)$. Let $C$ be $IB(u) \cap V_{<h}$.
  3. Using $K_{h-1}$, determine $k_v$ for all $v \in C$.
  4. For all $v \in C$, compute cluster($v, t_v, k_v$). If the union contains $u$, output NO. Else, output
     YES.
```

Claim 4.7. *Assume that K_{h-1} is provided correctly. Then* `IsFree`(v, h, K_{h-1}) *outputs YES iff $v \in F_h$. The running time is $O((d\ell\rho^{-1})^3)$.*

Proof. If $h = 1$, then all vertices are free (this is the free set before `globalPartition` begins any partitioning). Assume $h > 1$. So $F_h = V \setminus \bigcup_{v \in V_{<h}} \texttt{cluster}(v)$.

If $u \notin F_h$, then there exists $v \in V_{<h}$ such that $u \in \texttt{cluster}(v)$. By construction $\texttt{cluster}(v)$ is contained in $\mathrm{supp}(\widehat{M}^t\vec{v})$ for some $t \leq \ell$. Thus, $v \in IB(u)$ and $h_v < h$. Hence, v will be considered in Step 4 and the union will contain u. The output is NO. For the converse, observe that if the output is NO, then there is a $v \in V_{<h}$ such that $u \in \texttt{cluster}(v)$. Hence, $u \notin F_h$.

Now for the running time analysis. The running time of $\texttt{findIB}(v)$ is $O(d^2\ell^3\rho^{-2})$ (Claim 3.10) and $|C| \leq \ell\rho^{-1}$ (Claim 3.8). Each call to `cluster` takes $O(d\ell\rho^{-2})$ (Claim 3.3). The total running time is $O((d\ell\rho^{-1})^3)$.

□

We have the necessary tools to define the procedure `findr`. We will need the following definition in our description and analysis of `findr`.

Definition 4.8. *Assume $F_h \geq \beta n$. A vertex $s \in V_{\geq h}$ is called (h, k)-*viable *if $C := \texttt{cluster}(s, t_s, k)$ is not a singleton and $|C \cap F_h| \geq \beta^3 k$. (If $F_h < \beta n$, no vertex is (h, k)-viable.)*

Let us motivate this definition. When $C := \texttt{cluster}(s, t_s, k)$ is not a singleton, it is a low conductance cut of $\Theta(k)$ vertices. The vertex s is (h, k)-viable if C contains a non-trivial fraction of free vertices available in the hth phase. The viable vertices are those from which clustering will make significant "progress" in the hth phase. For each h, the procedure `findr` searches for values of k that lead to many (h, k)-viable vertices. In the next section, we prove that having sufficiently many clusters come from viable vertices ensures the cut bound of Definition 1.1.

```
findr($\boldsymbol{R}$)
  1. For $h = 1$ to $\overline{h}$:
     (a) Sample $\beta^{-10}$ uar vertices independently. Let $S_h$ be the multiset of sampled vertices that are in phase $\geq h$.
     (b) If $|S_h| \leq \beta^{-9}/2$, set $k_h = 0$ and continue for loop. Else, reset $S_h$ to the multiset of the first $\beta^{-8}$ vertices sampled.
     (c) For $k \in [\rho^{-1}]$ and for every $s \in S_h$:
         i. Compute $C :=$ cluster($s, t_s, k$).
         ii. For all $u \in C$, call IsFree($u, h, K_{h-1}$) to determine if $u \in F_{h-1}$.
         iii. If $C$ is not a singleton and $|C \cap F_{h-1}| \geq \beta^3 k$, mark $s$ as being $(h,k)$-viable.
     (d) If there exists some $k$ such that there are at least $12\beta^4|S_h|$ $(h,k)$-viable vertices, assign an arbitrary such $k$ as $k_h$. Else, assign $k_h := 0$.
  2. Output $K_{\overline{h}} = \{k_1, k_2, \ldots, k_{\overline{h}}\}$.
```

Claim 4.9. *The running time of* `findr` *is* $O((d\ell\delta^{-1}\rho^{-1})^5)$.

Proof. There are $\overline{h} = 2\delta^{-1}\log(\delta^{-1})$ iterations. We compute the running time of each iteration. There are at most $\rho^{-1}\beta^{-8}$ calls to `cluster`, each of which takes $O(d\ell\rho^{-2})$ time by Claim 3.3. For each call to `cluster`, there are at most ρ^{-1} calls to `IsFree`. Each call to `IsFree` takes $O((d\ell\rho^{-1})^3)$ time (Claim 4.7). The running time of each iteration is $O(\beta^{-10} + d\ell\rho^{-3}\beta^{-8} + d^3\ell^3\rho^{-5}\beta^{-8})$. By the parameter settings, since $\ell^2 \geq \varepsilon^{2\cdot 30} \geq (\varepsilon/10)^{-8} = \beta^{-8}$, the running time of each iteration $O((d\ell\rho^{-1})^5)$. The total running time is $O((d\ell\delta^{-1}\rho^{-1})^5)$. □

The following theorem gives the main guarantee of `findr`. The proof is a fairly straightforward Chernoff bound on top of an application of Theorem 4.1. Quite simply, the proof just says the following. Theorem 4.1 shows the existence of (h,k) pairs for which many vertices are viable. The `findr` procedure finds such pairs by random sampling.

Theorem 4.10. *The following property of the values* $K_{\overline{h}}$ *of* `findr`($\boldsymbol{R}$) *and the preprocessing choices holds with probability at least* $1 - \exp(-1/\varepsilon)$ *over all the randomness in* $\boldsymbol{R}$. *For all* $h \leq \overline{h}$, *if* $|F_h| \geq \beta n$, *at least* $\beta^5\delta n$ *vertices in* V_h *are* (h, k_h)*-viable.*

Proof. The proof has two parts. In the first part, we argue that whp, if $|F_h| \geq \beta n$, then a non-zero k_h is output. This part is an application of Theorem 4.1. In the second part, we prove that (whp), if a non-zero k_h is output, then it satisfies the desired properties. This part is proven using a simple Chernoff bound argument.

Fix an h. Condition on any choice of $V_1, V_2, \ldots, V_{h-1}$ such that $|F_h| \geq \beta n$. Note that $V_{\geq h} \supseteq F_h$, since all vertices in $V_{<h}$ are necessarily clustered by the hth phase. (Recall that `cluster`(v) always contains v.) Hence, $|V_{\geq h}| \geq \beta n$. There will be numerous low probability "bad" events that we need to track. We will describe these bad events, and refer to their probabilities as "Error 1", "Error 2", etc.

Error 1, $\exp(-\beta^{-8})$**.** The probability that a uar vertex is in $V_{\geq h}$ is at least β, and the expected size of S_h is at least $\beta \times \beta^{-10} = \beta^{-9}$. By the Chernoff bound of Theorem 4.3, $\Pr[|S_h| \leq \beta^{-9}/2] \leq \exp(-\beta^{-9}/12) \leq \exp(-\beta^{-8})$. Thus, with probability at least $1 - \exp(-\beta^{-8})$, Step 1c is reached and S_h is a multiset of iid uar β^{-8} elements in $V_{\geq h}$.

Let us assume that S_h is such a multiset, and prove that a non-zero k_h is output whp. We bring out the main tool, Theorem 4.1. Since $|F_h| \geq \beta n$, there exists a size threshold $k \leq \rho$ such that the following holds. For at least $(\beta^2/\log^2\beta^{-1})n$ vertices $s \in F_h$, there are at least $(\beta/\log^2\beta^{-1})\ell$ timesteps t such that: there exists $k' \in [k, 2k]$ such that (i) $L_{s,t,k'} \subseteq \mathrm{supp}(\widehat{M^t\vec{s}})$, (ii) $\Phi(L_{s,t,k'} \cup \{s\}) < \phi$, and (iii) $|L_{s,t,k'} \cap F| \geq \beta^3 k$. For any such (s,t,k) triple, consider a call to `cluster`(s,t,k). Observe that the call will output the largest level set of size in $[k, 2k]$ satisfying (i) and (ii). Hence, it will output (non-singleton) $L_{s,t,k''}$ such that $k' \leq k'' \leq 2k$ and

(i) and (ii) hold. Note that $L_{s,t,k''} \supseteq L_{s,t,k'}$, so the third item will also hold. Thus, *if* t_s *is set to one of these* $(\beta/\log^2\beta^{-1})\ell$ *timesteps* t, then s will be (h,k)-viable.

Error 2, $\exp(-\beta^{-1})$**.** Let us fix a size threshold k promised by Theorem 4.1. The probability that a uar element on $V_{\geq h}$ is marked as (k,h)-viable is at least the product of probability of choosing an appropriate s with the probability that t_s is chosen appropriately. Thus, the probability of find an (h,k)-viable vertex is at least $(\beta^2/\log^2\beta^{-1})\times(\beta/\log^2\beta^{-1}) = \beta^3/\log^4\beta^{-1}$. This probability is independent for all vertices in $V_{\geq h}$. By the Chernoff bound in Theorem 4.3, with probability at least $1-\exp(-\beta^4|S_h|/12)$, at least $\beta^3|S_h|/2\log^4\beta^{-1} \geq 12\beta^4|S_h|$ (h,k)-viable vertices are discovered in `findr`. In this case, in Step 1d, k_h is set to a non-zero value. The probability of this event happening is at least $1-\exp(-\beta^{-8})-\exp(-\beta^4|S_h|/8) \geq 1-\exp(\beta^{-1})$. (Recall that whp S_h is a multiset of iid uar β^{-8} vertices. In the union bound above, the first "bad event" is S_h *not* having β^{-8} vertices and the second "bad event" is discovering too few viable vertices.) We have concluded that whp, if $|F_h| \geq \beta n$, then k_h is non-zero.

We move to the second part of the proof, which asserts that (with high probability), an output non-zero k_h has the desired properties. Condition on any choice of the preprocessing. Note that the randomness is only over the choice of S_h. Fix any $k \leq \rho^{-1}$. Suppose that the number of (h,k)-viable vertices in $V_{\geq h}$ is at most $2\beta^5 n$. Then, the expected number of such vertices in S_h is at most $2\beta^5 n/|V_{\geq h}| \times |S_h| \leq 2\beta^4|S_h|$. (We use the lower bound $|V_{\geq h}| \geq |F_h| \geq \beta n$.)

Error 3, $2^{-12\beta^{-4}}$**.** Let X_k denote the random variable of the number of (h,k)-viable vertices in S_h. Since X_k is distributed as a binomial, by the Chernoff bound of Theorem 4.3, $\Pr[X_k > 12\beta^4|S_h|] \leq 2^{-12\beta^4|S_h|}$. Note than when $X_k < 12\beta^4|S_h|$, then k_h cannot be k. All in all, for any h, any choice of the t_vs, and any choice of k, if Step 1c is reached and the number of (h,k)-viable vertices in $V_{\geq h}$ is at most $2\beta^5 n$, then $k_h \neq k$ with probability at least $1-2^{-12\beta^{-4}}$. Taking the contrapositive, if $k_h \neq 0$ (Step 1c must have been reached), then the number of (h,k_h)-viable vertices in $V_{\geq h}$ is at least $2\beta^5 n$.

Error 4, $2\exp(-\delta\beta^5 n/12)$**.** Suppose the number of (h,k_h)-viable vertices in $V_{\geq h}$ is at least $2\beta^5 n$. By Claim 4.5 applied on the set of (h,k_h)-viable vertices in $V_{\geq h}$, with probability at least $1-2\exp(-\delta\beta^5 n/12)$, the number of such viable vertices in V_h is at least $\delta\beta^5 n$.

We take a union bound over the $2\delta^{-1}\log(\delta^{-1})$ values of h, the ρ^{-1} values of k, and all errors encountered thus far. The total error probability is at most $2\delta^{-1}\log(\delta^{-1})\cdot\rho^{-1}(\exp(-\beta^{-8})+\exp(\beta^{-1})+2^{-12\beta^{-4}}+2\exp(-\delta\beta^5 n/12))$. Note that $2\delta^{-1}\log(\delta^{-1}), \beta, \rho^{-1}$ are $\mathrm{poly}(\varepsilon^{-1})$, and thus the total error probability is at most $\exp(-\varepsilon^{-1})$. With the remaining probability, the following holds. For all phases h, if $|F_h| \geq \beta n$, a non-zero k_h is output. If a non-zero k_h is output, the number of (h,k_h)-viable vertices in V_h is at least $\delta\beta^5 n$. □

5 Proving the cut bound: the amortization argument

We come to the final piece of proving the guarantees of Theorem 1.2. We need to prove that the number of edges cut by the partition of `globalPartition` is at most εnd. This requires an amortization argument explained below. For the sake of exposition, we will ignore constant factors in this high-level description. One of the important takeaways is how various parameters are chosen to prove the cut bound.

Consider phase h where $|F_h| \geq \beta n$. Let us upper bound the number of edges cut by the clustering done on this phase. Roughly speaking, $|V_h| = \delta n$, so there are δn clusters created in this phase. Each cluster in this phase has at most $2k_h$ vertices. The number of edges cut by each such cluster is at most $2\phi k_h d$ (since `cluster` outputs a low conductance cut; ignore singleton outputs). So the total number of edges cut is at most $2\phi\delta k_h nd$.

Let us now lower bound the number of new vertices that are partitioned in phase h; this is the set $F_{h+1} \setminus F_h$. For each (h,k_h)-viable v in V_h, `cluster`(v) contains at least $\beta^3 k_h$ vertices in F_h. These will be newly partitioned vertices. Here comes the primary difficulty: the clusters for the different such v might not be disjoint. We need to lower bound the union of the clustered vertices in F_h. An alternate description of the challenge is as follows. We are only guaranteed that clusters from viable vertices v contains many vertices in F_h, the free set at the *beginning* of phase h. What we really need is for the cluster from v to contain many free vertices *at the time* that v is processed. Phases were introduced to solve this problem. By reducing δ, we can limit the size of V_h, thereby limiting the intersection between the clusters produced in this phase.

We now explain the math behind this argument. Consider some $w \in F_h$ and let c_w be the number of vertices in $V_{\geq h}$ that cluster v (call these seeds). Thus, $c_w = |\{s \mid s \in V_{\geq h}, v \in \mathtt{cluster}(s)\}$. The vertex w is clustered in phase h iff one of these c_w seeds is selected in V_h. By Claim 4.4, each such seed is independently selected in V_h with probability δ. The probability that w is clustered in this phases is precisely $1-(1-\delta)^{c_w}$. Crucially, $c_w \leq |IB(w)| \leq \ell\rho^{-1}$. We chose $\delta \ll \ell\rho^{-1}$, so $1-(1-\delta)^{c_w} \approx \delta c_w$.

Thus, the expected number of newly clustered vertices is at least $\sum_{w\in F_h} \delta c_w$. By rearranging summations, $\sum_{w\in F_h} c_w = \sum_{v\in V_{\geq h}} |\mathtt{cluster}(v) \cap F_h|$. For every (h,k_h)-viable vertex v in $V_{\geq h}$, $|\mathtt{cluster}(v)\cap F_h| \geq \beta^3 k_h$. The arguments in the proof of Theorem 4.10 shows that there are $\beta^5 n$ such vertices in $V_{\geq h}$ whp. Hence, we can lower bound (in expectation) the new number of newly clustered vertices as follows:

$$\sum_{w\in F_h} \delta c_w \geq \delta\cdot(\beta^5 n)\cdot(\beta^3 k_h) = \delta\beta^8 k_h n$$

We upper bounded the number of edges cut by $2\phi\delta k_h nd$. The ratio of edges cut to vertices clustered is $8\phi\beta^{-8}d$. The parameters are set to ensure that $8\phi\beta^{-8} \ll \varepsilon$, so the total number of edges cut is εnd.

The formal analysis requires some care to deal with conditional probabilities and dependencies between various phases. Also, Theorem 4.10 talks about V_h and not $V_{\geq h}$, which necessitates some changes. But the essence of the argument is the same.

Our main theorem is a cut bound for `globalPartition`.

Theorem 5.1. *The expected number of edges cut by the partitioning of* `globalPartition`$(\boldsymbol{R})$ *is at most εnd.*

We will break up the proof into two technical claims. Somewhat abusing notation, we say a vertex in $V_{\geq h}$ is h-viable if it is (h,k_h)-viable.

Claim 5.2.

$$\mathbf{E}[\#\ \textit{edges cut by}\ \mathtt{globalPartition}(\boldsymbol{R})] \leq 32\phi\beta^{-8}d^2\Big(\sum_{h<\overline{h}}\mathbf{E}[\sum_{v\in V_h}|\mathtt{cluster}(v)\cap F_h|)]\Big) + 2\beta nd$$

Proof. The proof goes phase by phase. We call a phase significant if $|F_h| \geq \beta n$. Edges cut in a significant phase are also called significant. Observe that the total number of edges cut is at most the number of significant edges cut plus βnd. (This contributes to the extra additive term in the claim statement.) Below, we will bound the total number of significant edges cut.

By Claim 4.6, with probability at least $1-2^{-\delta n}$, $|V_{\overline{h}}| \leq \delta n$. Note that $|F_h| \leq V_{\geq \overline{h}} = |V_{\overline{h}}|$. (The equality is because this is the last phase.) Since $\delta n < \beta n$, the expected number of significant edges cut in the last phase is at most $2^{-\delta n} nd < 1$.

Now assume that $h < \overline{h}$. Consider the edges cut in the hth phase. Consider any choice of $V_1, V_2, \ldots, V_{h-1}$ and $k_1, k_2, \ldots, k_h$. If $|F_h| < \beta n$, no significant edges are cut. Let us assume that $|F_h| \geq \beta n$. Each set `cluster`(v) output in this phase is either a singleton or a set of size at most $2k_h$ and conductance at most ϕ. In either case, the number of edges cut by removing `cluster`$(v) \cap F$ (in `globalPartition`) is at most $2\phi k_h d + d$. Note that $2\phi k_h d \geq 1$ (otherwise, by the connectedness of G, there can never be a set of size at most $2k_h$ of conductance $\leq \phi$). Hence, the number of significant edges cut by a single cluster is at most $2\phi k_h(d+d^2) \leq 4\phi k_h d^2$.

Note that $|V_{\geq h}| \geq |F_h| \geq \beta n$ and $|V_{\geq h}|$ is obviously at most n. By Claim 4.5 with $S = V_{\geq h}$, with probability at least $1 - 2\exp(-\delta\beta n/12)$ over the choice of V_h, $|V_h| \leq 2\delta n$. Hence, the total number of significant edges cut is at most $4\phi k_h d^2 \times 2\delta n = 8\phi\delta k_h d^2 n$.

By Theorem 4.10, with probability at least $1 - \exp(\varepsilon^{-1})$, if $|F_h| \geq \beta n$, at least $\beta^5 \delta n$ vertices in V_h are h-viable. Call this event $\mathcal{E}$. For every h-viable vertex in V_h, $|\texttt{cluster}(v) \cap F_h| \geq \beta^3 k_h$. For convenience, let $X_h := \sum_{v \in V_h} |\texttt{cluster}(v) \cap F_h|)$. Conditioned on $\mathcal{E}$, $X_h \geq \beta^8(\delta k_h n)$. Recall that with probability at least $1 - 2\exp(-\delta\beta n/12)$, the number of significant edges cut in this phase is at most $8\phi d^2(\delta k_h n)$. If $\mathcal{E}$ occurs, we can apply the bound $\beta^{-8} X_h \geq \delta k_h n$ and upper bound the number of significant edges cut in this phase by $8\phi\beta^{-8} d^2 X_h$,

Thus, with probability at least $1 - \exp(\varepsilon^{-1}) - 2\exp(-\delta\beta n/12)$, the number of significant edges cut in phase h is at most $(8\phi\beta^{-8}d^2)X_h$. In other words, there is an event $\mathcal{F}_h$ conditioned on which the above bound happens, and $\Pr[\mathcal{F}_h] \geq 1 - \exp(\varepsilon^{-1}) - 2\exp(-\delta\beta n/12)$. In the calculation below, we break into conditional expectations and use the fact that $\delta = \text{poly}(\varepsilon)$, $\beta = \Theta(\varepsilon)$, and that the number of phases is at most $2\delta^{-1}\log(\delta^{-1})$. We also use the fact that X_h is non-negative.

$$\sum_h \mathbf{E}[\#\text{ significant edges cut in phase } h] \leq \sum_h (\Pr[\mathcal{F}]\mathbf{E}[X_h|\mathcal{F}] + \Pr[\overline{\mathcal{F}}]nd) \tag{1}$$

$$\leq \sum_h \mathbf{E}[X_h] + 2\delta^{-1}\log(\delta^{-1})(\exp(\varepsilon^{-1}) + 2\exp(-\delta\beta n/12))nd \leq \sum_h \mathbf{E}[X_h] + \beta nd/2 \tag{2}$$

To this bound, we add the expected number of edges cut in the last phase (at most 1) and the number of non-significant edges cut (at most βn). This completes the proof. □

Claim 5.3.

$$\sum_{h<\overline{h}} \mathbf{E}[\sum_{v \in V_h} |\textit{cluster}(v) \cap F_h|)] \leq 4n$$

Proof. We will apply the following charging argument. When a vertex v is processed in $\texttt{globalPartition}(\boldsymbol{R})$, we will add one unit of charge to every vertex in $\texttt{cluster}(v) \cap F_h$. Note that the total amount of charge is exactly the quantity we wish to bound. Crucially, note that any vertex w receives charge in at most one phase; the phase where it leaves the free set.

We will prove that the expected charge that any vertex receives is at most 4 units, which will prove the claim. Fix a vertex w. Let χ be the random variable denoting the charge that w receives, and $\mathcal{E}_h$ be the event that w receives charge in phase h. Since w receives charge in exactly one phase, $\mathbf{E}[\chi] = \sum_h \mathbf{E}[\chi|\mathcal{E}_h]\Pr[\mathcal{E}_h]$. We will prove that, for all h, $\mathbf{E}[\chi|\mathcal{E}_h] \leq 4$, which implies that $\mathbf{E}[\chi] \leq 4$ as desired.

To analyze $\mathbf{E}[\chi|\mathcal{E}_h]$, first condition on a setting of $V_1, V_2, \ldots, V_{h-1}$ (such that $w \in F_h$) and all other preprocessing for all vertices. We refer to this setting as the event $\mathcal{C}$. The randomness for specifying V_h has not been set. The event $\mathcal{E}_h$ occurs if there is a $v \in V_h$ such that $w \in \texttt{cluster}(v)$. The charge χ is the number of vertices $v \in V_h$ such that $w \in \texttt{cluster}(v)$. Let c be the number of such vertices in $V_{\geq h}$. Note that $v \in IB(w)$, and by Claim 3.8, $c \leq \ell\rho^{-1}$.

By Claim 4.4, every vertex in $V_{\geq h}$ is in V_h with probability δ. Hence, $\Pr[\mathcal{E}_h|\mathcal{C}] = 1-(1-\delta)^c$. Note that $\delta c \leq \delta\ell\rho^{-1} = (d^{-70+6+60}\varepsilon^{3100} \cdot \varepsilon^{30} \cdot \varepsilon^{-3000} < 1/2$. Hence $(1-\delta)^c \leq 1 - \delta c + (\delta c)^2 \leq 1 - \delta c/2$ and $\Pr[\mathcal{E}_h|\mathcal{C}] \geq \delta c/2$. Note that $\mathbf{E}[\chi|\mathcal{C}] = \sum_{b>0}\binom{c}{b}\delta^b \leq \sum_{b>0}(\delta c)^b \leq 2\delta c$. Observe that $\mathbf{E}[(\chi|\mathcal{E}_h)|\mathcal{C}] \leq (2\delta c)/(\delta c/2) = 4$.

Note that the event $\mathcal{E}_h$ can be partitioned according to the different $\mathcal{C}$ events. Hence $\mathbf{E}[\chi|\mathcal{E}_h] = \sum_{\mathcal{C}} \mathbf{E}[(\chi|\mathcal{E}_h)|\mathcal{C}]\Pr[\mathcal{C}] \leq 4$. Thus, the proof is completed. □

Theorem 5.1 follows by a direct application of these claims and plugging in the parameter values.

Proof. (of Theorem 5.1) By Claim 5.2 and Claim 5.3, the expected number of edges cut by $\texttt{globalPartition}(\boldsymbol{R})$ is at most $128\phi\beta^{-8}d \cdot nd + 2\beta nd$. Plugging in the parameters $\phi = d^{-1}\varepsilon^{10}$, $\beta = \varepsilon/10$, and noting that ε is sufficiently small, the expectation is at most εnd. □

We can now wrap up the proof of Theorem 1.2, showing the existence of $(\varepsilon, \text{poly}(d\varepsilon^{-1}))$-partition oracles for minor-closed families.

Proof. (of Theorem 1.2) The procedure for the partition oracle is $\texttt{findPartition}(v, \boldsymbol{R})$. Let us prove each property of Definition 1.1.

Consistency: By Theorem 3.12, the partition created by calls to $\texttt{findPartition}(v, \boldsymbol{R})$ is precisely the same as the partition created by $\texttt{globalPartition}(\boldsymbol{R})$.

Cut bound: By Theorem 5.1, the expected number of edges cut is at most εnd.

Running time: The running time of $\texttt{findPartition}(v, \boldsymbol{R})$ is $O((d\ell\rho^{-1})^5)$ plus the running time of $\texttt{findr}$. The running time of $\texttt{findr}$ is $O((d\ell\delta^{-1}\rho^{-1})^5)$, by Claim 4.9. By the parameter settings, $\ell, \delta^{-1}, \rho^{-1}$ are all $\text{poly}(d\varepsilon^{-1})$. Hence, the total running time of $\texttt{findPartition}(v, \boldsymbol{R})$ is also $\text{poly}(d\varepsilon^{-1})$. □

6 Diffusion Behavior on Minor-Free Families

In this section, we state and prove the main theorem about diffusions on minor-free graph classes. This is the (only) part of the paper where the property minor-freeness makes an appearance. Theorem 6.1 is used in the proof of Sec. 7. For convenience, we recall the parameters involved.

- $\rho = d^{-60}\varepsilon^{3000}$: Minimum probability for truncation.
- $\ell = d^6\varepsilon^{-30}$: Maximum random walk length.
- $\beta = \varepsilon/10$: Unclustered fraction cutoff.
- $\delta = d^{-70}\varepsilon^{3100}$: Phase probability.
- $\alpha = \frac{\varepsilon^{4/3}}{300{,}000}$: Heavy bucket parameter.
- $\phi = \varepsilon^{10}$: Conductance parameter.

Theorem 6.1. *Let G be a bounded degree graph in minor-closed family. Let F be an arbitrary subset of at least βn vertices. There are at least $\beta^2 n/8$ vertices $s \in F$ such that: for at least $\beta\ell/8$ timesteps $t \in [\ell]$, $\widehat{M}^t\vec{s}(F) \geq \beta/16$.*

We note that this theorem holds for all graphs, if we replace the truncated walk $\widehat{M}$ by the standard random walk M. The main insight is that, for G in a minor-closed family, "polynomial" truncation of the walk distribution does not significantly affect the behavior.

The main property of bounded degree minor-free graphs we require is hyperfiniteness, as expressed by Proposition 4.1 of [AST90] (also used as Lemma 3.3 of [KSS19]).

Theorem 6.2. *There is an absolute constant γ such that the following holds. Let H be a graph on r vertices. Suppose G is an H-minor-free graph. Then, for all $b \in \mathbb{N}$, there exists a set of at most $\gamma r^{3/2}n/\sqrt{b}$ vertices whose removal leaves G with all connected components of size at most k.*

The key stepping stone to proving Theorem 6.1 is Lemma 6.4, which shows that truncation does not affect walk distributions from many vertices. Let us first state a simple fact on l_1-norms.

Fact 6.3. *Let $\vec{x}$ and $\vec{y}$ be vectors with non-negative entries, such that for all coordinates i, $\vec{x}(i) \geq \vec{y}(i)$. Then $\|\vec{x} - \vec{y}\|_1 = \|\vec{x}\|_1 - \|\vec{y}\|_1$.*

Proof. $\|\vec{x} - \vec{y}\|_1 \geq \sum_i |\vec{x}(i) - \vec{y}(i)| = \sum_i (\vec{x}(i) - \vec{y}(i)) = \|\vec{x}\|_1 - \|\vec{y}\|_1$. □

This fact bears relevance for us, since truncations of walk distribution vectors only reduce coordinates.

Lemma 6.4. *For at least $(1-\rho^{1/8})n$ vertices v, the following holds. For every $t \leq \ell$, $\|M^t\vec{v} - \widehat{M}^t\vec{v}\|_1 \leq \ell\rho^{1/9}$.*

Proof. Let H be an arbitrary forbidden minor for the minor-closed family of interest. We first apply Theorem 6.2 with $k = \lceil 1/\sqrt{\rho} \rceil$. There exists a set C of at most $\gamma r^{3/2}\rho^{1/4}dn$ edges who removal leads to connected components of size at most $\lceil 1/\sqrt{\rho} \rceil \leq 2/\sqrt{\rho}$. For convenience, set the constant $\gamma' := \gamma r^{3/2}$. We will need the following claim.

Claim 6.5. *For at least $(1-\rho^{1/8})n$ vertices v, the probability that an ℓ-length random walk encounters an edge of R is at most $\gamma'\ell\rho^{1/8}$.*

Proof. The proof is a Markov bound argument. Suppose not; so there exist strictly more than $\rho^{1/8}n$ vertices v such that an ℓ-length random walk encounters an edge of C with at least $\gamma'\ell\rho^{1/8}$ probability. Consider an ℓ-length random walk that starts from the uniform (also stationary) distribution. The above assumption implies that the expected number of C edges encountered is $> \rho^{1/8} \cdot \gamma'\ell\rho^{1/8} = \gamma'\ell\rho^{1/4}$. On the other hand, since the walk remains in the stationary distribution, for all $t \leq \ell$, the probability of encountering an edge in C at the tth step is precisely $|C|/2dn$. (Recall that the lazy random walk has $1/2dn$ of taking any edge.) By linearity of expectation, the expected number of C edges encountered is $\ell|C|/2dn$. By the bound of Theorem 6.2, $\ell|C|/2dn \leq \gamma'\ell\rho^{1/4}$ contradicting the bound obtained from the assumption. □

Consider such a vertex v, as promised by the previous paragraph. Let S be the connected component over vertices that contains v, after removing the edge cut C. Let q_t be the probability that the walk from v leaves S at the tth step; by the property of the previous parameter, $\sum_{t\leq\ell} p_t \leq \gamma'\ell\rho^{1/8}$. Let M_S be the transition matrix of the random walk M *restricted to* S. Note that M_S is not necessarily stochastic. We will use the truncated walk $\widehat{M_S}$. Observe that $\|\widehat{M}^t\vec{v}\|_1 \geq \|\widehat{M}^t_S\vec{v}\|_1$.

Since all coordinates of $\widehat{M}^t\vec{v}$ are at most those of $M^t\vec{v}$, by Fact 6.3, $\|M^t\vec{v} - \widehat{M}^t\vec{v}\|_1 = \|M^t\vec{v}\|_1 - \|\widehat{M}^t\vec{v}\|_1$. Since $\|M^t\vec{v}\|_1 = 1 = \|\vec{v}\|_1$ and $\|\widehat{M}^t\vec{v}\|_1 \geq \|\widehat{M}^t_S\vec{v}\|_1$, we can upper bound as follows by a telescoping sum.

$$\|M^t\vec{v} - \widehat{M}^t\vec{v}\|_1 \leq \sum_{l=1}^{t} \left(\|\widehat{M}^{l-1}_S\vec{v}\|_1 - \|\widehat{M}^l_S\vec{v}\|_1 \right) \tag{3}$$

$$= \sum_{l=1}^{t} \left(\|\widehat{M}^{l-1}_S\vec{v}\|_1 - \|M_S\widehat{M}^{l-1}_S\vec{v}\|_1 + \|M_S\widehat{M}^{l-1}_S\vec{v}\|_1 - \|\widehat{M}^l_S\vec{v}\|_1 \right) \tag{4}$$

The quantity $\|\widehat{M}^{l-1}_S\vec{v}\|_1 - \|M_S\widehat{M}^{l-1}_S\vec{v}\|_1$ is exactly the probability that a single step (according to M) from $\widehat{M}^{l-1}_S\vec{v}$ leaves S. Since all coordinates in $\widehat{M}^{l-1}_S\vec{v}$ are at most those of $M^{l-1}\vec{v}$, this probability is at most q_l. The quantity $\|M_S\widehat{M}^{l-1}_S\vec{v}\|_1 - \|\widehat{M}^l_S\vec{v}\|_1$ is the probability mass lost by truncation of $M_S\widehat{M}^{l-1}_S\vec{v}$. We apply the trivial bound $\rho|S|$. This is where the hyperfiniteness plays a role; since $|S| \leq 2/\sqrt{\rho}$, $\|\widehat{M}_S\vec{x}_{l-1} - M_S\vec{x}_{l-1}\|_1 \leq \rho \cdot /2\sqrt{\rho} = 2\sqrt{\rho}$.

We sum all the these bounds over $l \leq t$, and plug into (4). We bound $\|M^t\vec{v} - \widehat{M}^t\vec{v}\|_1 \leq \sum_{l\leq t} p_l + 2t\sqrt{\rho}$. By the properties of v, this is at most $\gamma'\ell\rho^{1/8} + 2\ell\sqrt{\rho} \leq \ell\rho^{1/9}$ (for sufficiently small ρ). □

We are now ready to prove Theorem 6.1. We will need the following simple "reverse Markov" inequality for bounded random variables.

Fact 6.6. *Let X be a random variable taking values in $[0,1]$ such that $\mathbf{E}[X] \geq \delta$. Then $\Pr[X \geq \delta/2] \geq \delta/2$.*

Proof. Let p be the probability that $\Pr[X \geq \delta/2]$.

$$\begin{aligned} \delta \leq \mathbf{E}[X] &= \Pr[X \geq \delta/2]\mathbf{E}[X|X \geq \delta/2] + \Pr[X < \delta/2]\mathbf{E}[X|X < \delta/2] \\ &\leq p + (1-p)(\delta/2) \leq p + \delta/2 \end{aligned}$$

□

Proof. (of Theorem 6.1) Define $\theta_{s,t}$ as follows. For $s \in F$ and $t \in [\ell]$: if t is odd, $\theta_{s,t} = 0$. If t is even, then $\theta_{s,t}$ is the probability that the t-length random walk starting from s ends in F.

Let us pick a uar source vertex in $s \in F$, pick a uar length $t \in [\ell]$. We use the fact that M is a symmetric matrix. We use $\mathbf{1}_F$ to denote the all 1s vector on F.

$$\mathbf{E}_{s,t}[\theta_{s,t}] = \mathbf{1}_F^T \sum_{i=1}^{\ell/2} (M^{2i}/\ell)(\mathbf{1}_F/|F|) = (\ell|F|)^{-1} \sum_{i \leq \ell/2} \mathbf{1}_F^T M^{2i} \mathbf{1}_F = (\ell|F|)^{-1} \sum_{i \leq \ell/2} \|M^i \mathbf{1}_F\|_2^2 \tag{5}$$

Note that $\|M^i \mathbf{1}_F\|_1 = |F|$, so by Jensen's inequality, $\|M^i \mathbf{1}_F\|_2^2 \geq |F|^2/n$. Plugging in (5), $\mathbf{E}_{s,t}[\theta_{s,t}] \geq \ell^{-1} \times (\ell/2)|F|/n \geq \beta/2$. For any s, $\mathbf{E}_t[\theta_{s,t}] \leq 1$. By Fact 6.6, there are at least $\beta|F|/4$ vertices $s \in F$ such that $\mathbf{E}_t[\theta_{s,t}] \geq \beta/4$. Again applying Fact 6.6, for at least $\beta|F|/4$ vertices $s \in F$, there are at least $\beta\ell/8$ timesteps $t \in [\ell]$ such that $\theta_{s,t} \geq \beta/8$, implying that $M^t \vec{s}(F) \geq \beta/8$.

By Lemma 6.4, there are at least $(1-\rho^{1/8})n$ vertices s such that for all $t \leq \ell$, $\|M^t\vec{s} - \widehat{M^t}\vec{s}\|_1 \leq \ell\rho^{1/9} = d^{6-60/9}\varepsilon^{-30+3000/9} \leq \beta/16$. By the parameters settings, $\rho^{1/8} < \varepsilon^{3000/8} \leq \beta|F|/8$. Invoking the bound from the previous paragraph, there are at least $\beta|F|/8$ satisfying the property of Lemma 6.4 and the condition at the end of the previous paragraph. For all such vertices s, for all $t \leq \ell$, $\widehat{M^t}\vec{s}(F) \geq M^t\vec{s}(F) - \beta/16$. Thus, for all such s, there are at least $\beta\ell/8$ timesteps $t \in [\ell]$ such that $\widehat{M^t}\vec{s}(F) \geq \beta/16$.

□

7 The proof of Theorem 4.10: local partitioning within F

We repeat the parameter values for convenience.

- $\rho = d^{-60}\varepsilon^{3000}$: Minimum probability for truncation.
- $\ell = d^6\varepsilon^{-30}$: Maximum random walk length.
- $\beta = \varepsilon/10$: Unclustered fraction cutoff.
- $\delta = d^{-70}\varepsilon^{3100}$: Phase probability.
- $\alpha = \frac{\varepsilon^{4/3}}{300{,}000}$: Heavy bucket parameter.
- $\phi = \varepsilon^{10}$: Conductance parameter.

Recall that Theorem 4.1 shows that there are many $s \in F$ from which (level sets of) diffusions in G discover low conductances cuts in F. We use the Lovász-Simonovits curve to represent the truncated diffusion vector, and keep track of the vertices of F wrt to the curve. This is done via a careful adaptation of Lovász-Simonovits method, as presented in Lemma 7.4.

The main technical tool which we will use in our analysis is the Lovász-Simonovits method, defined in [LS90], whose use for clustering was pioneered by [ST12].

Definition 7.1. *For a non-negative vector* $\mathbf{p}$ *over* V*, the function* $I : \mathbb{R}^n \times [n] \to [0,1]$ *is defined as*

$$I(\mathbf{p}, x) = \max_{\substack{\mathbf{w}\in[0,1]^n \\ \sum \mathbf{w}(u)=x}} \sum_{u\in V} \mathbf{p}(u)\mathbf{w}(u)$$

This is equivalent to summing over the x *heaviest elements of* $\mathbf{p}$ *when* x *is an integer, and linearly interpolating between these points otherwise.*

For notational convenience, we define:

$$I_{s,t}(x) = I(\widehat{M}^t\vec{s}, x).$$

Note that $I_{s,t}$ is a concave curve.

7.1 The Lovász-Simonovits lemma

The fundamental lemma of Lovász-Simonovits is the following (Lemma 1.4 of [LS90], also refer to Theorem 7.3.3 of Lecture 7 of [Spi]).

Lemma 7.2. *Let* $\overline{x} = \min(x, n-x)$*. Consider any non-negative vector* $\vec{p}$*, and let* S_x *denote the level set of* $M\vec{p}$ *with* x *vertices.*

$$I(M\vec{p}, x) \leq (1/2)(I(\vec{p}, x - 2\overline{x}\Phi(S_x)) + I(\vec{p}, x - 2\overline{x}\Phi(S_x)))$$

The concavity of the curves implies monotonicity, $I(M\vec{p}) \leq I(\vec{p})$. The application of this lemma to our setting leads to the following statement.

Lemma 7.3. *For all* $t \leq \ell$ *and* $x \leq 1/\rho$*,*

$$I_{s,t}(x) \leq (1/2)(I_{s,t-1}(x(1-\Phi(L_{s,t,x}))) + I_{s,t-1}(x(1+\Phi(L_{s,t,x}))))$$

Let $f_{t,w,y}$ be the straight line between the points $(w, I_{s,t}(w))$ and $(y, I_{s,t}(y))$.

Lemma 7.4. *Let* $t_0 < t_1 < \ldots < t_h$ *be time steps. Suppose* $\forall i \leq h$ *and* $x \in [w,y]$*:* $L_{s,t_i,x} \subseteq \mathrm{supp}(\widehat{M}^t\vec{s}) \implies \Phi(L_{s,t_i,x}) \geq \psi$*. Then,* $\forall i \leq h, \forall x \in [w,y]$

$$I_{s,t_i}(x) \leq f_{t_0-1,w,y}(x) + \sqrt{\min(x-w, y-x)}(1-\psi^2/128)^i$$

Proof. For convenience, let $\Delta_x = \min(x-w, y-x)$. We prove by induction over i.

For showing the base case take $i = 0$. Now consider the following cases.

- Suppose $x = w$ or $x = y$. By monotonicity, $I_{s,t_0}(x) \leq I_{s,t_0-1}(x)$. Since $x \in \{w,y\}$, the latter is exactly $f_{t_0,w,y}(x)$.
- Suppose $x \in [w+1, y-1]$. Then $\Delta_x \geq 1$ and $I_{s,t_0}(x) \leq 1 \leq \sqrt{\Delta_x}$.
- Suppose $x \in (w, w+1)$. Note that $\Delta_x = w - x < 1$. By the definition of the LS curve, $I_{s,t_0}(x) = I_{s,t_0}(w) + (w-x)(I_{s,t_0}(w+1) - I_{s,t_0}(w)) \leq I_{s,t_0-1}(w) + \sqrt{w-x} \leq f_{t_0-1,w,y}(x) + \sqrt{\Delta_x}$.
- Suppose $x \in (y-1, y)$. An identical argument to the above holds.

Now for the induction. Suppose the premise holds at step t_i. Namely for $x \in [w,y]$, for all level sets $L_{s,t_i,x}$ contained inside $\mathrm{supp}(\widehat{M}^t\vec{s})$, $\Phi(L_{s,t_i,x}) \geq \psi \geq \psi$. We would like to upperbound $I_{s,t_i}(x)$. To this end, let us consider some $x \in [w,y]$. By Lemma 7.3,

$$\begin{aligned} I_{s,t_i}(x) &\leq (1/2)[I_{s,t_i-1}(x(1-\Phi(L_{s,t_i,x}))) + I_{s,t_i-1}(x(1+\Phi(L_{s,t_i,x})))] && (6) \\ &\leq (1/2)[I_{s,t_{i-1}}(x(1-\Phi(L_{s,t_i,x}))) + I_{s,t_{i-1}}(x(1+\Phi(L_{s,t_i,x})))] && (7) \end{aligned}$$

The second inequality follows by monotonicity, since $t_{i-1} \leq t_i - 1$. Note that $\Delta_x = \min(x - w, y-x) \leq x$ for all $x \in [w,y]$. Claim 7.5 (which we prove after the current lemma) shows the following.

Claim 7.5. *For all $1 \leq i \leq h$, for all $x \in [w, y]$, the following holds*

$$I_{s,t_i}(x) \quad \leq \quad (1/2)[I_{s,t_{i-1}}(x - \Delta_x \psi/4)) + I_{s,t_{i-1}}(x + \Delta_x \psi/4))] \tag{8}$$

Now, let $x_L = x - \Delta_x \psi/4$ and $x_R = x + \Delta_x \psi/4$. Using Claim 7.5 we get

$$\begin{aligned} I_{s,t_i}(x) \quad &\leq \quad (1/2)[f_{t_0-1,w,y}(x_L) + \sqrt{\Delta_{x_L}}(1 - \psi^2/128)^{i-1} \\ &\quad\; + f_{t_0-1,w,y}(x_R) + \sqrt{\Delta_{x_R}}(1 - \psi^2/128)^{i-1}] \qquad (9) \\ &= \quad (1/2)[f_{t_0-1,w,y}(x_L) + f_{t_0-1,w,y}(x_R)] \\ &\quad\; + (1/2)[\sqrt{\Delta_{x_L}})(1 - \psi^2/8)^{i-1} + \sqrt{\Delta_{x_R}}(1 - \psi^2/128)^{i-1}] \qquad (10) \end{aligned}$$

Here, (10) follows from the induction hypothesis. Since $f_{t_0-1,w,y}$ is a linear function, the first term is exactly $f_{t_0-1,w,y}(x)$. We analyze the second term.

We first assume that $\Delta_x = x - w$ (instead of $y - x$).

$$\begin{aligned} \Delta_{x_L} \quad &= \quad \min(x - \psi\Delta_x/4 - w, y - x + \psi\Delta_x/4) \qquad (11) \\ &= \quad \min((1 - \psi/4)\Delta_x, y - x + \psi/4\Delta_x) \leq (1 - \psi/4)\Delta_x \qquad (12) \end{aligned}$$

Analogously,

$$\begin{aligned} \Delta_{x_R} \quad &= \quad \min(x + \psi\Delta_x/4 - w, y - x - \psi\Delta_x/4) \qquad (13) \\ &= \quad \min((1 + \psi/4)\Delta_x, y - x - \psi\Delta_x/4) \leq (1 + \psi/4)\Delta_x \qquad (14) \end{aligned}$$

Thus, the second term of (10) is at most $(1/2)(1 - \psi^2/128)^{i-1}\sqrt{\Delta_x}(\sqrt{1 - \psi/4} + \sqrt{1 + \psi/4})$.

Now, we consider $\Delta_x = y - x$.

$$\begin{aligned} \Delta_{x_L} \quad &= \quad \min(x - \psi\Delta_x/4 - w, y - x + \psi\Delta_x/4) \qquad (15) \\ &= \quad \min(x - \psi\Delta_x/4 - w, (1 + \psi/4)\Delta_x) \leq (1 + \psi/4)\Delta_x \qquad (16) \end{aligned}$$

Analogously,

$$\begin{aligned} \Delta_{x_R} \quad &= \quad \min(x + \psi\Delta_x/4 - w, y - x - \psi\Delta_x/4) \qquad (17) \\ &= \quad \min(x + \psi\Delta_x/4 - w, (1 - \psi/4)\Delta_x) \leq (1 - \psi/4)\Delta_x \qquad (18) \end{aligned}$$

In this case as well, the second term of (10) is at most $(1/2)(1 - \psi^2/128)^{i-1}\sqrt{\Delta_x}(\sqrt{1 - \psi/4} + \sqrt{1 + \psi/4})$.

In both cases, we can upper bound (10) as follows. (We use the inequality $\frac{\sqrt{1-z}+\sqrt{1+z}}{2} \leq 1 - z^2/8$.

$$I_{s,t_i}(x) \leq f_{t_0-1,w,y}(x) + (1 - \psi^2/128)^{i-1}\sqrt{\Delta_x}\,\frac{\sqrt{1 - \psi/4} + \sqrt{1 + \psi/4}}{2} \leq f_{t_0-1,w,y}(x) + (1 - \psi^2/128)^{i}\sqrt{\Delta_x}$$

□

Now, we establish Claim 7.5, the missing piece in the above proof.

Proof. (of Claim 7.5) Suppose $x_{max} \in [w, y]$ is the maximum value of $x \in [w, y]$ for which $L_{s,t_i,x}$ is still inside the support of the truncated diffusion at the t_i-th step. We split into three cases: $x \leq x_{max}$, $x \in (x_{max}, x_{max} + \Delta_{x_{max}}\psi/2]$, $x > x_{max} + \Delta_{x_{max}}\psi/2$. Note that in the latter two cases, $L_{s,t_i,x}$ is not contained in $\mathrm{supp}(\widehat{M^{t_i}\vec{s}})$.

Case 1, $x \leq x_{max}$: Note that (8) holds by concavity of the Lovász-Simonovits curve when $L_{s,t_i,x} \subseteq \mathrm{supp}(\widehat{M^{t_i}\vec{s}})$ (because then this level set has conductance at least ψ).

Case 2, $x \in (x_{max}, x_{max} + \Delta_{x_{max}}\psi/2]$**:** Let $S = L_{s,t_i,x_{max}}$ and let $T = L_{s,t_i,x}$. Observe that

$$\Phi(T) = \frac{|E(T,\overline{T})|}{d|T|} \overset{(\mathbf{1})}{\geq} \frac{|E(S,\overline{S})| - \psi/2 \cdot d|S|}{d|S| + \psi/2 \cdot d|S|} \overset{(\mathbf{2})}{\geq} \frac{\psi d|S|/2}{2d|S|} \geq \frac{\psi}{4} \tag{19}$$

Here, (**1**) follows because T could contain at most $\psi|S|/2$ neighbors of S which could cost us at most $\psi d|S|/2$ edges in the cut $(S, \overline{S})$. (**2**) follows by upperbounding ψ by 1. Again the claim in (8) follows by concavity of the Lovázs-Simonovits curve.

Case 3, $x > x_{max} + \Delta_{x_{max}}\psi/2$**:** Now let $x_r = x_{max} + \Delta_{x_{max}}\psi/2$. Write $x = x_{max} + \Delta_{x_{max}}\psi/2 + s$. Recall $\Delta_x = \min(x - w, y - x)$. We claim that $x - \Delta_x\psi/4 \geq x_{max}$. First let us see how to establish (8) assuming this claim holds. Assuming this claim, we have

$$I_{s,t_i}(x - \Delta_x\psi/4) = I_{s,t_i}(x_{max}) = I_{s,t_i}(x + \Delta_x\psi/4) = \|\widehat{M}^{t_i}\vec{s}\|_1.$$

And therefore,

$$\begin{aligned} I_{s,t_i}(x) &= \frac{1}{2} \cdot [I_{s,t_i}(x - \Delta_x\psi/4) + I_{s,t_i}(x + \Delta_x\psi/4)] \\ &\leq \frac{1}{2} \cdot \left[I_{s,t_{i-1}}(x - \Delta_x\psi/4) + I_{s,t_{i-1}}(x + \Delta_x\psi/4)\right] \end{aligned}$$

Now, all that remains to establish (8) is to show $x - \Delta_x\psi/4 \geq x_{max}$. For simplicity, write $\Delta_m = \Delta_{x_{max}}$. Now consider two cases depending on the value of Δ_m

1. **Case 1** $\Delta_m = x_{max} - w$.
   In this case note that

$$\begin{aligned} x - \Delta_x\psi/4 &= x_{max} + \Delta_m\psi/2 + s - (x - w)\psi/4 \\ &\geq x_{max} + \Delta_m\psi/2 + s - (x_{max} + \Delta_m\psi/2 + s - w)\psi/4 \\ &\geq x_{max} + \Delta_m\psi/4 - \Delta_m\psi^2/8 + s - s\psi/4 \\ &\geq x_{max} + \Delta_m\psi/8 + s(1 - \psi/4) \geq x_{max} \end{aligned}$$

   which establishes the claim above as desired.

2. **Case 2** $\Delta_m = y - x_{max}$.
   In this case note that

$$\begin{aligned} x - \Delta_x\psi/4 &= x_{max} + \Delta_m\psi/2 + s - (y - x)\psi/4 \\ &\geq x_{max} + \Delta_m\psi/2 + s - (y - x_{max} - \Delta_m\psi/2 - s)\psi/4 \\ &\geq x_{max} + \Delta_m\psi/4 + \Delta_m\psi^2/8 + s + s\psi/4 \\ &\geq x_{max} \end{aligned}$$

Thus, in both cases, the claim from above holds. This means that (8) holds as long as the premise holds for the t_i-th step. □

7.2 From leaking timesteps to the dropping of the LS curve

We fix a source vertex s, and consider the evolution of $\widehat{M}^t\vec{s}$. Therefore, we drop the dependence of s from much of the notation.

We use $\widehat{p}_t$ to denote $\widehat{M}^t\vec{s}$. We begin with a few definitions.

Definition 7.6. *A timestep t is called* leaking for source s *if, for all $k \leq \rho^{-1}$: if $L_{s,t,k} \subseteq \operatorname{supp}(\widehat{M}^t \vec{s})$ and $|L_{s,t,k} \cap F| \geq \alpha^2 k/400$, then $\Phi(L_{s,t,k}) \geq 1/d\ell^{1/3}$.*

*If timestep t is not leaking for s, there exists $k \leq \rho^{-1}$ such that $L_{s,t,k} \subseteq \operatorname{supp}(\widehat{M}^t \vec{s})$, $|L_{s,t,k} \cap F| \geq \alpha^2 k/400$, and $\phi(L_{s,t,k}) < 1/d\ell^{1/3}$. Such a k is denoted as an (s,t)-*certificate of non-leakiness.

We set $\alpha = \varepsilon^{4/3}/300,000$.

Following the construction of the LS curve $I_{s,t}$, we will order each vector $\widehat{p}_t$ in decreasing order, breaking ties by id. The *rank* of a vertex is its position in (the sorted version of) $\widehat{p}_t$.

Definition 7.7. *Let the* bucket $B_{t,r}$ *denote the set of vertices whose rank in $\widehat{p}_t$ is in the range $[2^r, 2^{r+1})$.*

A bucket $B_{t,r}$ is called heavy *if $\sum_{v \in B_{t,r} \cap F} \widehat{p}_t(v) \geq \alpha$. (The bucket restricted to F has large probability.)*

The following lemma says that if there are many leaking timesteps, then the LS curve drops at heavy buckets.

Lemma 7.8. *Fix $r \geq 0$. Suppose for some $s \in F$, there exist $\ell' \geq \beta^3 \ell/8$ leaking timesteps $t_0 < t_1 < \ldots < t_{\ell'}$ such that for all $0 \leq i \leq \ell'$, $B_{t_i,r}$ is heavy. Then, $I_{s,t_{\ell'}}(2^{r+1}) < I_{s,t_0}(2^{r+1}) - \alpha/4$.*

The main tool used in our proof is our adaptation of Lovász-Simonovits lemma done in Lemma 7.4. We first make a definition.

Definition 7.9. *Fix $r \geq 0$, a source s and a timestep t. A vertex $w \in [2^r, 2^{r+1}]$ is called a balanced split for t if $|L_{t,w} \cap F| \geq \alpha 2^r/3$ and $\sum_{v \in B_{t,r} \setminus L_{t,w}} \widehat{p}_t(v) \geq \alpha/3$.*

We will first prove the following claim which essentially follows by averaging arguments.

Claim 7.10. *Fix $r \geq 0$ and suppose for some source vertex $s \in F$, there exist ℓ' leaking timesteps $t_0 < t_1 < \ldots < t_{\ell'}$ such that for all $0 \leq i \leq \ell'$, $B_{t_i,r}$ is heavy. Then, there exists a vertex w that is a balanced split for at least an $\alpha/3$-fraction of timesteps in $T = \{t_0, t_1, \ldots t_{\ell'}\}$.*

Proof. Since $B_{t_0,r}$ is heavy, $I_{s,t_0}(2^r) < 1$. Since the support of $\widehat{p}_t$ is at most ρ^{-1}, this implies that $2^r < \rho^{-1}$ and $r \leq -\lg \rho$ (and this holds by the choice of parameters).

For all $v \in B_{t,r}$, $\widehat{p}_t(v) \leq 1/2^r$. Since $\sum_{v \in B_{t,r} \cap F} \widehat{p}_t(v) \geq \alpha$, $|B_{t,r} \cap F| \geq \alpha 2^r$.

For convenince, let $T = \{t_0, t_1, \ldots t_{\ell'}\}$. Pick w uar in $[2^r, 2^{r+1})$. Let X_i be the indicator for w being a balanced split for t_i. Recall that $|B_{t_i,r} \cap F| \geq \alpha 2^r$. Sort the vertices of $B_{t_i,r} \cap F$ by increasing rank and consider the vertices in positions $\alpha 2^r/3$ and $2\alpha 2^r/3]$. Let the rank corresponding to these vertices by u_1 and u_2. We first argue that any rank $w \in [u_1, u_2]$ is a balanced split. We have $|L_{t,w} \cap F| \geq \alpha 2^r/3$ because $w \geq u_1$. For all $v \in B_{t_i,r}$, $\widehat{p}_{t_i}(v) \leq 1/2^r$. Thus, $\sum_{v \in L_{t_i,u_2} \cap B_{t_i,r}} \widehat{p}_{t_i}(v) \leq (1/2^r)(2\alpha 2^r/3) = 2\alpha/3$. Note that $\sum_{v \in B_{t_i,r}} \widehat{p}_t(v) \geq \alpha$, since the bucket is heavy Hence, for any $w \leq u_2$, $\sum_{v \in B_{t,r} \setminus L_{t,w}} \widehat{p}_t(v) \geq \alpha - 2\alpha/3 = \alpha/3$.

As a consequence, for any t_i, there are at least $\alpha 2^r/3$ values of w that are balanced splits. In other words, $\mathbf{E}[X_i] \geq \alpha/3$. By linearity of expectation, $\mathbf{E}[\sum_{i \leq \ell'} X_i] \geq \alpha \ell'/3$. Thus, there must exist some $w \in [2^r, 2^{r+1})$ that is a balanced split for at least $\alpha \ell'/3$ timesteps. □

Next, we show the following claim which essentially uses leakiness of a timestep $t \in T$ and the balanced split vertex w promised by Claim 7.10 to spell out a set with enough free vertices with large conductance.

Claim 7.11. *Fix $r \geq 0$ and let $w \in [2^r, 2^{r+1})$ be a split vertex as promised by Claim 7.10 and let $t_{i_1} < t_{i_2} < \ldots < t_{i_{\alpha\ell'/3}}$ denote the timesteps for which w is a balanced split. Let $y = \min(2^{r+6+\lceil \lg(1/\alpha) \rceil}, \rho^{-1})$. Then, for all $x \in [w, y]$ and for all $t \in \{t_{i_1}, t_{i_2}, \cdots, t_{i_{\alpha\ell'/3}}\}$, whenever $L_{t,x} \subseteq \operatorname{supp}(\widehat{M}^t \vec{s})$, then $\Phi(L_{t,x}) \geq 1/d\ell^{1/3}$.*

Proof. Take $x \in [w, y]$ and a leaking timestep $t \in \{t_{i_1}, t_{i_2}, \cdots, t_{i_{\alpha\ell'/3}}\}$. Note that $x \le y \le \rho^{-1}$ clearly holds. Now, to establish the lower bound on conductance claimed, we first unpack what it means for t to be a leaking timestep Definition 7.6. It says: If $L_{t,x} \subseteq \mathrm{supp}(\widehat{M^t}\vec{s})$ and $|L_{t,x} \cap F| \ge \alpha^2 k/400$, then it better hold that $\phi(L_{t,x}) \ge 1/d\ell^{1/3}$.

Note that $y \le 2^{r+6+\lceil \lg(1/\alpha) \rceil} \in [2^r(64/\alpha), 2^{r+1}(64/\alpha)]$. Since $r \le -\lg\rho$, $y \le 128(\rho\alpha)^{-1}$.

Note that for all $t \in \{t_{i_1}, t_{i_2}, \cdots, t_{i_{\alpha\ell'/3}}\}$ and $x \in [w, y]$, $L_{t,x}$ contains at least $\alpha 2^r/3$ vertices of F. Thus, at least a $(\alpha 2^r/3)/(2^{r+1} \cdot 64/\alpha) \ge \alpha^2/400$-fraction of $L_{t,x}$ is in F. Now note that since t is leaking, we see that one of the following will hold. Either

- $L_{t,x} \subseteq \mathrm{supp}(\widehat{M^t}\vec{s})$ and $\Phi(L_{t,x}) \ge 1/d\ell^{1/3}$, Or
- $L_{t,x} \not\subseteq \mathrm{supp}(\widehat{M^t}\vec{s})$.

And this establishes the claim. □

Now, we have all the ingredients to prove Lemma 7.8. The key step which remains is an application of Lemma 7.4.

Proof. (Of Lemma 7.8) Suppose $w \in [2^r, 2^{r+1})$ is a balanced split at $\alpha\ell'/3$ timesteps as promised by Claim 7.10. Let $y = \min(2^{r+6+\lceil \lg(1/\alpha) \rceil}, \rho^{-1})$ and as observed in Claim 7.11, note that for $x \in [w, y]$ if $L_{t,x} \subseteq \mathrm{supp}(\widehat{M^t}\vec{s})$, it holds that $\phi(L_{t,x}) \ge 1/d\ell^{1/3}$. Now, we apply Lemma 7.4. For all $x \in [w, y]$, we have $I_{s,t_{\ell'}}(x) \le I_{s,t_{i_{\alpha\ell'/3}}}(x) \le f_{t_{i_1-1},w,y}(x) + \sqrt{x}(1 - 1/128d^2\ell^{2/3})^{\alpha\ell'/3}$. By the premise, $\ell' \ge \beta^3\ell/8$ and therefore we have

$$(1-1/128d^2\ell^{2/3})^{\alpha\ell'/3} \le (1-1/128d^2\ell^{2/3})^{\alpha\beta^3\ell/3} = (1-1/128d^2\ell^{2/3})^{128d^2\ell^{2/3}\cdot\frac{\alpha\beta^3\ell^{1/3}}{3\cdot 128 d^2}} \le \exp(-1/\alpha)$$

which holds because, for sufficiently small $\varepsilon > 0$, we have

$$\ell^{1/3} = \frac{d^2}{\varepsilon^{10}} \ge \frac{d^2 \cdot 10^{20}}{\varepsilon^7} \ge \frac{d^2}{\alpha^3\beta^3}.$$

Further, by the monotonicity of LS curves, $I_{s,t_{\ell'}}(x) \le f_{t_{i_1-1},w,y}(x) + \exp(-1/\alpha) \le f_{t_{i_0},w,y}(x) + \exp(-1/\alpha)$. Specifically, we get

$$I_{s,t_{\ell'}}(2^{r+1}) \le f_{t_{i_0},w,y}(2^{r+1}) + \exp(-1/\alpha). \tag{20}$$

Since w is a good split, $I_{s,t_{i_0}}(2^{r+1}) \ge I_{s,t_{i_0}}(w) + \alpha/3$. Note that

$$\begin{aligned} f_{t_{i_0},w,y}(2^{r+1}) &= I_{s,t_{i_0}}(w) + (2^{r+1} - w)\left(\frac{I_{s,t_{i_0}}(y) - I_{s,t_{i_0}}(w)}{y - w}\right) \\ &\le I_{s,t_{i_0}}(w) + 2^{r+1}/(y/2) & (21) \\ &\le I_{s,t_{i_0}}(w) + 2^{r+1} \times \left(\frac{2\alpha}{2^r \cdot 64}\right) = I_{s,t_{i_0}}(w) + \alpha/16 & (22) \end{aligned}$$

The first inequality above follows by upper bounding $I_{s,t_{i_0}}(y) - I_{s,t_{i_0}}(w)$ by 1, dropping the negative term and noting that $y - w \ge y/2$ for a sufficiently small α. Together with (20), we get

$$\begin{aligned} I_{s,t_{\ell'}}(2^{r+1}) \le f_{t_{i_0},w,y}(2^{r+1}) + \exp(-1/\alpha) &\le I_{s,t_{i_0}}(w) + \alpha/16 + \exp(-1/\alpha) \\ &\le I_{s,t_{i_0}}(2^{r+1}) - \alpha/3 + \alpha/16 + \exp(-1/\alpha) & (23) \end{aligned}$$

By monotonicity of the LS curve, $I_{s,t_{\ell'}}(2^{r+1}) < I_{s,t_0}(2^{r+1}) - \alpha/4$.

□

Now, we state a key lemma. It says that a fixed bucket (parameterized by r) satisfies the following at most timesteps: (i) either it does not contain enough free vertices, or (ii) if it contains many free vertices at a particular timestep, then most of the corresponding timesteps are not leaky.

Lemma 7.12. *Fix $r \geq 0$ and take any $s \in F$. There are at most $\beta^3\ell/\alpha$ leaking timesteps t (with respect to s) where $B_{t,r}$ is heavy.*

Proof. We prove by contradiction. Suppose there are more than $\beta^3\ell/\alpha$ leaking timesteps t where $B_{t,r}$ is heavy. We break these up into $4/\alpha$ contiguous blocks of $\beta^3\ell/4$ leaking timesteps. By Lemma 7.8, after every such block of timesteps, $I_{s,t}(2^{r+1})$ reduces by more than $\alpha/4$. Note that $I_{s,0}(2^{r+1}) \leq 1$, and thus, after $4/\alpha$ blocks, $I_{s,t}(2^{r+1})$ becomes negative. Contradiction to the non-negativity of $I_{s,t}(2^{r+1})$. □

7.3 Proof of Theorem 4.1

We finally prove Theorem 4.1. In particular, recall that this theorem claims that for an arbitrary set $F \subseteq V$ with $|F| \geq \beta n$, there exists a size threshold k such that one can find enough source vertices $s \in F$ such that ℓ-step diffusions from s contain enough non-leaky timesteps. Moreover, these non-leaky timesteps can be used to obtain a low conductance cut restricted to F. We begin by showing that indeed many sources $s \in F$ have the desired behavior.

Lemma 7.13. *There are at least $\beta^2 n/8$ vertices $s \in F$, such that: there are at least $\beta\ell/16$ timesteps t in $[\ell]$ that are not leaking for s.*

Proof. We fix any vertex s satisfying the conditions of Theorem 6.1. Let us recall what this means. This means that for at least $\beta\ell/8$ timesteps t, it holds that $\widehat{M^t \vec{s}}(F) \geq \beta/16$. We will show that conclusion in Lemma 7.13 above holds for s which will establish the lemma. We prove by contradiction.

To this end, let us suppose for any vertex s satisfying the conditions of Theorem 6.1, there are at most $\beta\ell/16$ non-leaky timesteps. There are at least $\beta\ell/8 - \beta\ell/16 = \beta\ell/16$ timesteps t that are leaking for s, such that $\widehat{M^t \vec{s}}(F) \geq \beta/16$. Fix any such timestep t and consider the buckets $B_{t,r}$. There are at most $-\lg\rho$ buckets with non-zero probability mass, and by averaging, there exists $r \leq -\lg\rho$ such that

$$\sum_{v \in F \cap B_{t,r}} \widehat{p}_t(v) \geq \beta/(-16\lg\rho) = \frac{\varepsilon}{160 \cdot 3000 \lg(1/\varepsilon)} \geq \frac{\varepsilon^{4/3}}{300,000} = \alpha$$

where the last step holds for sufficiently small ε and therefore, $B_{t,r}$ is heavy.

Thus, for each of the $\beta\ell/16$ leaking timesteps t above, there exists some $r \leq -\lg\rho$ such that $B_{t,r}$ is heavy. By averaging, there exists some $r \leq -\lg\rho$ such that for $\beta\ell/(-16\lg\rho)$ leaking timesteps t, $B_{t,r}$ is heavy. However, for sufficiently small ε ($\varepsilon < 2^{-30}$), we have

$$\frac{\beta\ell}{-16\lg\rho} = \frac{\varepsilon \cdot \ell}{160 \cdot 3000 \log(1/\varepsilon)} \geq 1000\varepsilon^{3-4/3}\ell \geq \frac{\beta^3\ell}{\alpha}$$

which contradicts Lemma 7.12. □

Lemma 7.14. *Let $|F| \geq \beta n$. There exists a $r \leq \lg(1/\rho)$ such that for $\geq \beta^2 n/(8\lg^2(\rho^{-1}))$ vertices $s \in F$, the following holds. For at least $\beta\ell/(\lg^2(\rho^{-1}))$ timesteps t, there exists $k \in [2^r, 2^{r+1}]$ that is an (s,t)-certificate of non-leakiness.*

Proof. This is an averaging argument. Apply Lemma 7.13. For each of the $\beta^2 n/8$ vertices $s \in F$, there are at least $\beta\ell/16$ timesteps t that are not leaking for s. Thus, for every such (s,t) pair, there exists $k_{s,t} \leq \rho^{-1}$ that is an (s,t)-certificate of non-leakiness. We basically bin the logarithm of the certificates. Thus, to every pair (s,t) (of the above form), we associate $r_{s,t} = \lfloor \lg k_{s,t} \rfloor$. By averaging, for each relevant s, there is a value r_s such that for at least $\beta\ell/(16\lg(\rho^{-1}))$ timesteps t, there is an (s,t)-certificate in $[2^{r_s}, 2^{r_s+1}]$. Again, by averaging there exists $r \leq \lg(\rho - 1)$ such that there are at least $\beta^2 n/(8\lg(\rho^{-1})) \geq \beta^2 n/(\lg^2(\rho^{-1}))$ vertices $s \in F$ for which there exist at least $\beta\ell/(16\lg(\rho^{-1})) \geq \beta\ell/\lg^2(\rho^{-1})$ timesteps t, such that there is an (s,t)-certificate for non-leakiness in $[2^r, 2^{r+1}]$. □

Theorem 4.1 follows as a corollary of Lemma 7.14. We now present the proof.

Proof. (Of Theorem 4.1) As seen from Lemma 7.14, there exists some $r \leq -\lg(\rho)$ such that there are at least $\Omega(\beta^2/\lg(\beta^{-1})) \cdot n$ vertices $s \in F$ each of which in turn has (s,t)-certificates of non-leakiness for at least $\Omega(\beta/16\lg^2(\beta^{-1})) \cdot \ell$ different values of t. We simply choose $k = 2^r$.

Let $S \subseteq F$ denote the collection of these relevant sources. And for $s \in S$, define

$$C_s = \{t \leq \ell : \text{ there exists a } (s,t) - \text{ certificate of non-leakiness}\}.$$

Take $s \in S$, $t \in C_s$. We will show that there exists $k' = k'(s,t) \in [k, 2k]$ such that the level set $L_{s,t,k'}$ satisfies the following.

- $L_{s,t,k} \subseteq \text{supp}(\widehat{M}^t \vec{s})$.
- $\phi(L_{s,t,k'} \cup \{s\}) \leq 1/\ell^{1/3}$.
- $|L_{s,t,k'} \cap F| \geq \alpha^2 k'/400 \geq \beta^3 k$.

The first item above follows from the conclusion of Lemma 7.14, Definition 7.6 and taking contrapositive in Lemma 7.4. Unpacking, this means that since $t \in C_s$ is a non-leaking timestep for s, it follows that there exists $k' = k'(s,t) \in [k, 2k]$ for which $L_{s,t,k'} \subseteq \text{supp}(\widehat{M}^t \vec{s})$. The last item above holds for this choice of k' from the conclusion of Lemma 7.14. For item 2 above, again note that our choice of k' and Lemma 7.14 imply that

$$\phi(L_{s,t,k'}) \leq 1/d\ell^{1/3} = 1/d \cdot \frac{\varepsilon^{10}}{d^2} = \varepsilon^{10}/d^3 = \phi/d^3$$

and therefore $\phi(L_{s,t,k'} \cup \{s\}) \leq \phi$ also follows as by (possibly) including a single vertex in the set, the number of cut-edges can only increase by d. □

8 Proofs of applications

The proofs here are quite straightforward and appear (in some form) in previous work. We sketch the proofs, and do not give out the specifics of the Chernoff bound calculations. Specifically, we mention Theorem 9.28 and its proof in [Gol17], which contains these calculations.

Proof. (of Theorem 1.3) For input graph G, we set up the partition oracle. Note that we can estimate the number of edges cut by random sampling. We pick a vertex u uniformly at random, pick a uar neighbor v, and call the partition oracle on u and v. If these lie in different components, the edge (u,v) is cut. By sampling $\Theta(1/\varepsilon)$, we can determine with high probability if more than εnd edges are cut by the partitioning (Chernoff bound). If so, we simply reject, since G is far from being in a minor-closed family.

Otherwise, we sample poly(ε^{-1}) uar vertices, and determine the component that each vertex belongs to. For each component, we directly determine if it belongs to $\mathcal{Q}$. (If there is an efficient algorithm, we can run that algorithm.) By a Chernoff bound, if G was ε-far from $\mathcal{Q}$, with high probability, one of the components would not be in $\mathcal{Q}$. Overall, the query complexity is poly$(d\varepsilon^{-1})$. □

Proof. (of Theorem 1.4) As with the previous proof, we set up the partition oracle. With high probability, at most $\varepsilon dn/c$ edges are cut by the partitioning given by the oracle, where c is the largest amount by which an edge addition/deletion changes f. We sample poly$(d\varepsilon^{-1})$ uar vertices and determine the component that each vertex belongs to. For each component, we compute f exactly. We take the sum of f-values, and rescale appropriately to get an additive εnd estimate for f. □

References

[AST90] Noga Alon, Paul Seymour, and Robin Thomas. A separator theorem for nonplanar graphs. *Journal of the American Mathematical Society*, 3(4):801–808, 1990. 18

[AST94] Noga Alon, Paul D. Seymour, and Robin Thomas. Planar separators. *SIAM J. Discrete Math.*, 7(2):184–193, 1994. 1

[BSS08] I. Benjamini, O. Schramm, and A. Shapira. Every minor-closed property of sparse graphs is testable. In *Symposium on the Theory of Computing (STOC)*, pages 393–402, 2008. 3

[CGR$^+$14] Artur Czumaj, Oded Goldreich, Dana Ron, C Seshadhri, Asaf Shapira, and Christian Sohler. Finding cycles and trees in sublinear time. *Random Structures & Algorithms*, 45(2):139–184, 2014. 3

[CSS09] Artur Czumaj, Asaf Shapira, and Christian Sohler. Testing hereditary properties of nonexpanding bounded-degree graphs. *SIAM Journal on Computing*, 38(6):2499–2510, 2009. 3

[DP09] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms.* Cambridge, 2009. 12

[EHNO11] Alan Edelman, Avinatan Hassidim, Huy N. Nguyen, and Krzysztof Onak. An efficient partitioning oracle for bounded-treewidth graphs. In *Workshop on Randomization and Computation (RANDOM)*, pages 530–541, 2011. 3

[FLVW17] Hendrik Fichtenberger, Reut Levi, Yadu Vasudev, and Maximilian Wötzel. On testing minor-freeness in bounded degree graphs with one-sided error. *CoRR*, abs/1707.06126, 2017. 3

[FPS19] H. Fichtenberger, P. Peng, and C. Sohler. Every testable (infinite) property of bounded-degree graphs contains an infinite hyperfinite subproperty,. In *Symposium on Discrete Algorithms (SODA)*, page 714–726, 2019. 3

[Gol17] O. Goldreich. *Introduction to Property Testing.* Cambridge University Press, 2017. 1, 2, 3, 6, 27

[GR99] O. Goldreich and D. Ron. A sublinear bipartite tester for bounded degree graphs. *Combinatorica*, 19(3):335–373, 1999. 3

[GR02] O. Goldreich and D. Ron. Property testing in bounded degree graphs. *Algorithmica*, 32(2):302–343, 2002. 1

[HKNO09] A. Hassidim, J. Kelner, H. Nguyen, and K. Onak. Local graph partitions for approximation and testing. In *Foundations of Computer Science (FOCS)*, pages 22–31, 2009. 1

[HT74] John Hopcroft and Robert Tarjan. Efficient planarity testing. *Journal of the ACM (JACM)*, 21(4):549–568, 1974. 1

[KSS18] Akash Kumar, C. Seshadhri, and Andrew Stolman. Finding forbidden minors in sublinear time: A $\mathrm{o}(\mathrm{n}^{1/2\,+\,\mathrm{o}(1)})$-query one-sided tester for minor closed properties on bounded degree graphs. In *Foundations of Computer Science (FOCS)*, pages 509–520, 2018. 3

[KSS19] Akash Kumar, C. Seshadhri, and Andrew Stolman. Random walks and forbidden minors II: a poly($d\,\epsilon^{-1}$)-query tester for minor-closed properties of bounded degree graphs. In *STOC 2019, Phoenix, AZ, USA, June 23-26, 2019.*, pages 559–567, 2019. 3, 18

[Kur30] K. Kuratowski. Sur le problème des courbes gauches en topologie. *Fundamenta Mathematica*, 15:271–283, 1930. 1

[LR15] Reut Levi and Dana Ron. A quasi-polynomial time partition oracle for graphs with an excluded minor. *ACM Transactions on Algorithms (TALG)*, 11(3):24, 2015. 2, 3

[LS90] László Lovász and Miklós Simonovits. The mixing rate of markov chains, an isoperimetric inequality, and computing the volume. In *Foundations of Computer Science (FOCS)*, pages 346–354, 1990. 3, 4, 20, 21

[LT80] Richard J. Lipton and Robert Endre Tarjan. Applications of a planar separator theorem. *SIAM J. Comput.*, 9(3):615–627, 1980. 1

[NS13] Ilan Newman and Christian Sohler. Every property of hyperfinite graphs is testable. *SIAM Journal on Computing*, 42(3):1095–1112, 2013. 1, 2, 3

[RS95a] N. Robertson and P. D. Seymour. Graph minors. XII. Distance on a surface. *Journal of Combinatorial Theory Series B*, 64(2):240–272, 1995. 1

[RS95b] N. Robertson and P. D. Seymour. Graph minors. XIII. The disjoint paths problem. *Journal of Combinatorial Theory Series B*, 63(1):65–110, 1995. 1

[RS04] N. Robertson and P. D. Seymour. Graph minors. XX. Wagner's conjecture. *Journal of Combinatorial Theory Series B*, 92(1):325–357, 2004. 1

[Spi] D. Spielman. Lecture notes on spectral graph theory. http://www.cs.yale.edu/homes/spielman/eigs/. 21

[ST12] D. Spielman and S.-H. Teng. A local clustering algorithm for massive graphs and its application to nearly-linear time graph partitioning. *SIAM Journal on Computing*, 42(1):1–26, 2012. 3, 4, 20

[Wag37] K. Wagner. Über eine eigenschaft der ebenen komplexe. *Mathematische Annalen*, 114:570–590, 1937. 1

[YI15] Yuichi Yoshida and Hiro Ito. Testing outerplanarity of bounded degree graphs. *Algorithmica*, 73(1):1–20, 2015. 3