Almost Optimal Local Graph Clustering Using Evolving Sets

REID ANDERSEN, Microsoft Corp, Redmond, Washington
SHAYAN OVEIS GHARAN, University of Washington, Seattle, Washington
YUVAL PERES, Microsoft Research, Redmond, Washington
LUCA TREVISAN, UC Berkeley, Berkeley, California

Spectral partitioning is a simple, nearly linear time algorithm to find sparse cuts, and the Cheeger inequalities provide a worst-case guarantee for the quality of the approximation found by the algorithm. A *local graph partitioning algorithm* finds a set of vertices with small conductance (i.e., a sparse cut) by adaptively exploring part of a large graph G, starting from a specified vertex. For the algorithm to be local, its complexity must be bounded in terms of the size of the set that it outputs, with at most a weak dependence on the number n of vertices in G. Previous local partitioning algorithms find sparse cuts using random walks and personalized PageRank [Spielman and Teng 2013; Andersen et al. 2006].

In this article, we introduce a simple randomized local partitioning algorithm that finds a sparse cut by simulating the *volume-biased evolving set process*, which is a Markov chain on sets of vertices. We prove that for any $\epsilon > 0$, and any set of vertices A that has conductance at most ϕ, for at least half of the starting vertices in A our algorithm will output (with constant probability) a set of conductance $O(\sqrt{\phi/\epsilon})$. We prove that for a given run of the algorithm, the expected ratio between its computational complexity and the volume of the set that it outputs is $\mathrm{vol}(A)^{\epsilon}\phi^{-1/2}\,\mathrm{polylog}(n)$, where $\mathrm{vol}(A) = \sum_{v\in A} d(v)$ is the volume of the set A. This gives an algorithm with the same guarantee (up to a constant factor) as the Cheeger's inequality that runs in time slightly superlinear in the size of the output. This is the first sublinear (in the size of the input) time algorithm with almost the same guarantee as the Cheeger's inequality. In comparison, the best previous local partitioning algorithm, by Andersen et al. [2006], has a worse approximation guarantee of $O(\sqrt{\phi \log n})$ and a larger ratio of $\phi^{-1}\,\mathrm{polylog}(n)$ between the complexity and output volume.

As a by-product of our results, we prove a bicriteria approximation algorithm for the expansion profile of any graph. For $0 < k \leq \mathrm{vol}(V)/2$, let $\phi(k) := \min_{S:\mathrm{vol}(S)\leq k} \phi(S)$. There is a polynomial time algorithm that, for any $k, \epsilon > 0$, finds a set S of volume $\mathrm{vol}(S) \leq O(k^{1+\epsilon})$ and expansion $\phi(S) \leq O(\sqrt{\phi(k)/\epsilon})$. As a new technical tool, we show that for any set S of vertices of a graph, a lazy t-step random walk started from a randomly chosen vertex of S will remain entirely inside S with probability at least $(1 - \phi(S)/2)^t$. This itself provides a new lower bound to the uniform mixing time of any finite state reversible Markov chain.

Categories and Subject Descriptors: F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems; G.1.8 [**Partial Differential Equations**]: Spectral Methods; G.3 [**Probability and Statistics**]: Markov Processes

General Terms: Algorithms, Theory

Additional Key Words and Phrases: Cheeger's inequality, local graph clustering, random walks, sparsest cut

The work of L. Trevisan was supported by the National Science Foundation under Grant No. CCF-1017403.
Author's addresses: R. Andersen, One Microsoft Way, Redmond, WA 98052; email: reidan@microsoft.com; S. O. Gharan, Paul Allen Center for Computer Science and Engineering, Box 352350, University of Washington, WA 98195-2350; email: shayan@cs.washington.edu; Y. Peres, Microsoft Research, One Microsoft Way, Redmond, WA 98052; email: peres@microsoft.com; L. Trevisan, Department of Electrical Engineering and Computer Sciences, U.C¿ Berkeley, CA 94720; email: luca@berkeley.edu.
Preliminary versions of this article appeared in Proceedings of the ACM Symposium on Theory of Computing, ACM, New York, 2009, and the IEEE Symposium on Foundations of Computer Science, 2012.

DOI: http://dx.doi.org/10.1145/2856030

ACM Reference Format:
Reid Andersen, Shayan Oveis Gharan, Yuval Peres, and Luca Trevisan. 2016. Almost optimal local graph clustering using evolving sets. J. ACM 63, 2, Article 15 (May 2016), 31 pages.
DOI: http://dx.doi.org/10.1145/2856030

1. INTRODUCTION

Let $G = (V, E)$ be an undirected graph, with $n := |V|$ vertices, and $m := |E|$ edges, and let $d(v)$ denote the degree of vertex $v \in V$. The *volume* of a set $S \subseteq V$ is defined as the sum of the degree of vertices in S,

$$\mathrm{vol}(S) := \sum_{v \in S} d(v).$$

The *conductance* of a set S is defined as

$$\phi(S) := \partial(S)/\mathrm{vol}(S),$$

where $\partial(S)$ denotes the number of edges that leaves S. Let

$$\phi(G) := \min_{S:\mathrm{vol}(S) \leq \mathrm{vol}(V)/2} \phi(S)$$

be the conductance (uniform sparsest cut) of G. The Cheeger inequalities [Alon and Milman 1985; Alon 1986; Kwok et al. 2013] prove that the Spectral Partitioning algorithm finds, in nearly linear time, a $O(1/\sqrt{\phi(G)})$ approximation to the uniform sparsest cut problem. Most notably, the approximation factor does not depend of the size of the graph; in particular, the Cheeger inequalities imply a constant factor approximation if $\phi(G)$ is constant. Variants of the spectral partitioning algorithm are widely used in practice [Kleinberg 1999; Shi and Malik 2000; Tolliver and Miller 2006].

Often, one is interested in applying a sparsest cut approximation algorithm *iteratively*, that is, first find an approximate sparsest cut in the graph and then recurse on one or both of the subgraphs induced by the set found by the algorithm and by its complement. Such an iteration might be used to find a *balanced* sparse cut if one exists (cf. Orecchia and Vishnoi [2011] and Orecchia et al. [2012]) or to find a good *clustering* of the graph, an approach that leads to approximate clusterings with good worst-case guarantees, as shown by Kannan, Vempala, and Vetta [Kannan et al. 2004]. Even though each application of the spectral partitioning algorithm runs in nearly linear time, iterated applications of the algorithm can result in a quadratic running time.

Spielman and Teng [2004], and subsequently [Andersen et al. 2006], studied local graph partitioning algorithms that find a set S of small conductance, in time nearly linear in the size of the output set S, by adaptively examining a small subset of the input graph near a specified starting vertex. Note that the running time can be sublinear in the size of the input graph if the algorithm finds a small output set S. When iterated, such an algorithm finds a balanced sparse cut in nearly linear time in the size of the graph and can be used to find a good clustering in nearly linear time as well. Such algorithms are useful for finding target clusters in large graphs and for quickly finding collections of small clusters. They have been applied in practice to probe the community structure of social and information networks [Flake et al. 2000; Andersen and Lang 2006; Leskovec et al. 2008] and have been used as subroutines to design fast algorithms for other partitioning problems [Spielman and Teng 2004, 2013].

Another advantage of such "local" algorithms is that if there are both large and small sets of near-optimal conductance, the algorithm is more likely to find the smaller sets. Thus, such algorithms can be used to approximate the "small-set expander" problem, which is related to the unique games conjecture [Raghavendra and Steurer 2010] and

the expansion profile of a graph (that is, the cut of the smallest conductance among all sets of a given volume).

Finding small, low-conductance sets is also interesting in clustering applications. In a social network, for example, a low-conductance set of users in the "friendship" graph represents a "community" of users who are significantly more likely to be friends with other members of the community than with non-members, and discovering such communities has several applications. While large communities might correspond to large-scale, known factors, such as the fact that American users are more likely to have other Americans as friends or that people are more likely to have friends around their age, small communities generally contain more interesting and substantial information. Leskovec et al. [2009, 2010] observed that, in large networks, the "best" communities, that is, sets with the smallest conductance, have a size of roughly 100 nodes; as the size increases, the best possible communities become more and more blended into the remainder of the network. There is also experimental evidence that a significant fraction of vertices in networks belong to small communities [Li et al. 2011; Li and Peng 2011].

1.1. Almost Optimal Local Graph Clustering

A *local graph clustering* algorithm is a *local* graph algorithm that finds a non-expanding set in the local neighborhood of a given vertex v in time proportional to the size of the output set. The *work/volume ratio* of such an algorithm, which is the ratio of the computational time of the algorithm in a single run, and the volume of the output set, may depend only poly logarithmically to the size of the graph. The main result of this article is a new local graph clustering algorithm with improved work/volume ratio and an improved bound on the conductance of the output set.

The problem was first studied in the remarkable work of Spielman and Teng [2004]. They designed an algorithm, `Nibble`, such that for any set $A \subseteq V$, if the initial vertex, v, is sampled randomly according to the degree of vertices in A, with a constant probability, then `Nibble` finds a set of conductance $O(\sqrt{\phi(A)\log^3 n})$, with a work/volume ratio of $O(\phi^{-2}(A)\,\mathrm{polylog}(n))$. `Nibble` finds the desired set by looking at the *threshold sets* of the probability distribution of a t-step random walk started at v. To achieve the desired computational time, they keep the support of the probability distribution small by removing a small portion of the probability mass at each step.

Andersen et al. [2006] used the approximate PageRank vector rather than approximate random-walk distribution, and they managed to improve the conductance of the output set to $O(\sqrt{\phi(A)\log n})$ and the work/volume ratio to $O(\phi^{-1}(A)\,\mathrm{polylog}\,n)$.

Note that the approximation guarantee of both of the above algorithms depends poly-logarithmically on the size of the graph. It has been an open problem to design a local variant of the Cheeger's inequality, that is, to provide a local graph clustering algorithm with an approximation guarantee that does not depend on the size of G.

In this article, we design a local partitioning algorithm, called `EvoPar`, based on evolving sets. Our algorithm is conceptually simpler than the previous local graph partitioning algorithms in the sense that it depends on fewer number of parameters. It simply simulates a Markov chain (or independent copies of a Markov chain) called "evolving set process" on the subsets of vertices of G until a certain stopping time is reached and then outputs the resulting set.

Apart from simplicity, it has two main advantages over the previous results. The first and the third authors [Andersen and Peres 2009] showed that the evolving sets Markov chain can be *simulated* very efficiently by only looking at the boundary of the current state, thus reducing the dependency on $\phi(A)$ in the work-per-volume ratio to $\phi^{-1/2}(A)$. The second and the last authors [Oveis Gharan and Trevisan 2012] employed

a more careful analysis of the evolving set process together with a new upper bound on the escape probability of random walks and managed to eliminate the dependency on n in the conductance of the output set. Putting these together, we obtain the first sublinear (in the size of the input) time algorithm with almost the same guarantee as the Cheeger's inequality.

The following theorem summarizes the contributions of both articles.

THEOREM 1.1 (MAIN). EvoPar(v, k, ϕ, ϵ) *takes as input a starting vertex* $v \in V$*, a target conductance* $\phi \in (0, 1)$*, a target volume* k*, and* $0 < \epsilon < 1$ *and outputs a set of vertices. For a given run of the algorithm, let* S *be the output and let* W *be the computational complexity of the algorithm. Then,* S *and* W *depend on the randomness of the algorithm, and*

(1) *The work per volume ratio* $W/\mathrm{vol}(S)$ *satisfies*

$$\mathbb{E}[W/\mathrm{vol}(S)] = O(k^{\epsilon}\phi^{-1/2}\log^2 n).$$

(2) *If* $A \subseteq V$ *is a set of vertices that satisfy* $\phi(A) \leq \phi$*, and* $\mathrm{vol}(A) \leq k$*, then there is a subset* $A' \subseteq A$ *with volume at least* $\mathrm{vol}(A)/2$*, such that if* $v \in A'$*, then with a constant probability* S *satisfies*
   (*a*) $\phi(S) = O(\sqrt{\phi/\epsilon})$,
   (*b*) $\mathrm{vol}(S) = O(k^{1+\epsilon})$.

Letting $\epsilon = 1/\log(k)$ in the above theorem we obtain a local graph clustering algorithm that for any target set A finds a set of S such that $\mathrm{vol}(S) = O(k)$, $\phi(S) = O(\sqrt{\phi(A)\log(k)})$ with the work/volume ratio of $O(\phi^{-1/2}(A)\,\mathrm{polylog}(n))$. Therefore, the above theorem is a strict improvement on the previous local graph clustering algorithms. In addition, letting $\epsilon = \Omega(1)$, we can avoid any dependency on n in the conductance of the output set at the cost of a slight increase in the running time.

Our algorithm can be seen as an almost optimal local graph clustering algorithm in the sense that it approximates $\phi(A)$; this is because, on one hand, obtaining a better-than-square-root approximation for the balanced separator problem, in the regime where the optimum value is a constant independent of n, is SSE hard[1] [Raghavendra et al. 2010]; on the other hand, for smaller values of $\phi(A)$, using Cheeger's inequality, it seems impossible to approximate conductance by a factor better than square root with a random-walk-based algorithm. This is because the performance of random walks is closely tied to the spectrum of the random-walk matrix of G, but, using Cheeger's inequality, there is a square-root gap between the conductance of cuts and the spectral gap [Jerrum and Sinclair 1989].

1.2. Approximating the Expansion Profile

As a by-product of the above result, we give an approximation algorithm for the expansion profile of G. Lovász and Kannan [1999] defined the *expansion profile* of a graph G as follows:

$$\phi(k) := \min_{S:\mathrm{vol}(S)\leq k} \phi(S).$$

Lovász and Kannan used the expansion profile as a parameter to prove strong upper bounds on the mixing time of random walks. The notion of expansion profile recently received significant attention in the literature because of its close connection to the small set expansion problem and the unique games conjecture [Raghavendra and Steurer 2010].

[1]See Theorem 1.3 for the statement of the small set expansion conjecture.

Raghavendra and Steurer [2010] defined the small set expansion problem.

PROBLEM 1.2 (GAP SMALL SET EXPANSION sse(δ, ϕ)). *Given a graph G, and constants $\delta, \phi > 0$, distinguish whether*

$$\phi(\delta \cdot m) \geq 1 - \phi \quad or \quad \phi(\delta \cdot m) \leq \phi.$$

They conjectured that the above problem is hard for sufficiently small values of δ,

CONJECTURE 1.3 (GAP-SMALL-SET EXPANSION CONJECTURE). *For every $\phi > 0$, there exists $\delta > 0$ such that* sse(δ, ϕ) *is NP-hard.*

Most interestingly, they show that Gap-Small-Set-Expansion problem is easier than the Unique Games conjecture.

THEOREM 1.4 (RAGHAVENDRA AND STEURER [2010]). *The Gap-Small-Set Expansion conjecture implies the Unique Games Conjecture.*

The above theorem shows that if someone wants to design an algorithm to refute Unique Games conjecture, then she must start by designing an algorithm for the small set expansion problem.

Let A be the adjacency matrix of G and D be the diagonal matrix of vertex degrees. Let the *threshold rank* of G, denoted by $\text{rank}_{1-\eta}(D^{-1}A)$, be the number (with multiplicities) of eigenvalues λ of the random-walk matrix $D^{-1}A$, satisfying $\lambda > 1 - \eta$. Arora et al. [2010a] designed a polynomial time algorithm for the small set expansion problem on graphs where for some constant $\eta > 0$, $\text{rank}_{1-\eta}(D^{-1}A) \geq n^{\text{poly}(\eta/\phi)}$. Unfortunately, their algorithm does not provide any approximation of $\phi(\delta \cdot m)$ for general graphs, since they may have much fewer large eigenvalues.

Raghavendra et al. [2010] and Bansal et al. [2011] used semidefinite programming relaxations of the problem and designed algorithms that approximate $\phi(\delta \cdot m)$ within factors $O(\phi(\delta \cdot m)^{-1/2}\sqrt{\log 1/\delta})$ and $O(\sqrt{\log n \log 1/\delta})$ of the optimum, respectively. However, the quality of both approximation algorithms is not independent of δ, which is of interest to the small set expansion problem.

Here, we design an algorithm that approximates $\phi(k)$ as a function of $\phi(k^{1-\epsilon})$, without any dependency on k or the size of the graph; specifically, we prove the following corollary:

COROLLARY 1.5. *There is a polynomial time algorithm that takes as input a target volume k and $0 < \epsilon < 1$ and outputs a set S, s.t.* $\text{vol}(S) \leq O(k^{1+\epsilon})$ *and* $\phi(S) \leq \sqrt{\phi(k)/\epsilon}$.

For example, if G has a set A where $\text{vol}(A) = \sqrt{m}$, then the above algorithm returns a set S with $\text{vol}(S) \leq m^{1/2+\epsilon}$ and $\phi(S) \leq \sqrt{\phi(A)/\epsilon}$. Note that if $k^{1+\epsilon} \geq m$, then the above corollary is trivial because the algorithm can return $S = V$ in the output. The above corollary shows that the hard instances of the small set expansion problem are graphs where even though $\phi(\delta \cdot m) \ll 1$, for any $\epsilon = \Omega(1)$, we have $\phi(m^{1-\epsilon}) \approx 1$.

Independent of our work, Kwok and Lau [2012] have obtained a somewhat different proof of Theorem 1.5 by also showing an upper bound on the escape probability of random walks (see the remarks after Theorem 3.1 for more details). They use this result to construct a local graph clustering algorithm based on the work of Spielman and Teng [2013] with a weaker guarantees compared to Theorem 1.1.

1.3. Approximating Balanced Separator

One application of our local partitioning algorithm is a fast algorithm for finding balanced cuts. Spielman and Teng showed how to find a balanced cut in nearly linear time by repeatedly removing small sets from a graph using local partitioning

[Spielman and Teng 2004]. Applying their technique with our algorithm yields an algorithm `EvoPartition` with the following properties. The algorithm has complexity $m^{1+O(\epsilon)}\phi^{-1/2}$ polylog(n), and it outputs a set of vertices whose conductance is $O(\sqrt{\phi/\epsilon})$ and whose volume is at least half of any set with conductance at most ϕ and volume at most $m^{1-\epsilon}/c$, where ϕ, ϵ are inputs to the algorithm and $c > 0$ is an absolute constant.

Orecchia et al. [2012] very recently designed almost linear ($O(m\,\text{polylog}(n))$) time algorithm that gives a $\sqrt{\phi}$ approximation to the balanced separator problem. Their algorithm is the current fastest algorithm that provides a nontrivial approximation guarantee to the balanced cut problem (note that compared to our algorithm, the running time of Orecchia et al. [2012] does not depend on ϕ). There are also several algorithms that provide stronger approximation guarantees with a slower running time (see, e.g., Arora et al. [2010b] and Sherman [2009]). These algorithms produce cuts with conductance $O(\phi\,\text{polylog}(n))$, and their computational complexity is dominated by the cost of solving $O(n^{o(1)})$ many single-commodity flow problems. In Section 6, we give a more detailed description of `EvoPartition`.

1.4. Lower Bounds on Uniform Mixing Time of Random Walks

Using our techniques, we also prove improved lower bounds on the mixing time of reversible Markov chains. Since any reversible finite-state Markov chain can be realized as a random walk on a weighted undirected graph, we model the Markov chain as a random walk on a weighted graph G.

Let $\tau(\epsilon)$ be the ϵ-*uniform* mixing time of a Markov Chain (see Section 3.1 for the definition). We prove the following lower-bound on $\tau(\epsilon)$.

PROPOSITION 1.6. *For any (weighted) graph* $G = (V, E)$*, any* $S \subseteq V$ *with* $\text{vol}(S) \leq \text{vol}(V)/2$*, and* $0 < \epsilon < 1$*, if* $\phi(S) \leq 0.7$*, then*

$$\tau(\epsilon) \geq \frac{\log(\text{vol}(V)/(1+\epsilon)\text{vol}(S))}{2\phi(S)} - 1.$$

1.5. Techniques

Our local partitioning algorithm is based on the evolving set process. The *evolving set process* (ESP) is a Markov chain whose states are subsets of the vertex set of a graph. Its transition rule is a simple procedure that grows or shrinks the current set. Morris and the second author used the ESP, and the closely related *volume-biased evolving set process* (volume-biased ESP), to bound the mixing time of Markov chains in terms of their isoperimetric properties [Morris and Peres 2003]. The volume-biased ESP is equivalent to the *strong stationary dual* of a random walk, which was introduced earlier by Diaconis and Fill [1990]. Further applications of evolving sets were described in Montenegro [2007, 2009]. In all of these results, evolving sets were used as analytical tools rather than algorithms.

Our algorithm simulates independent copies of the volume-biased evolving set process until a certain stopping time is reached and then outputs the resulting set. To prove the local approximation guarantee, we bound the rate of growth of the sets in the volume-biased ESP. In particular, we prove a lower bound that depends on the conductance of the sets observed by the process and an upper bound that depends on the conductance of certain sets that contain the starting vertex. Say the process starts with a vertex v, and $S_0 = \{v\}, S_1, S_2, \ldots, S_\tau$ is a sample path of the process. In Theorem 4.2, we employ a non-trivial martingale argument to show that, with high probability, for any stopping time τ, at least one set S_t is such that $\phi(S_t) \leq O(\sqrt{\log \text{vol}(S_\tau)/\tau})$. Now, if one can show that up to a sufficiently large time T, the process constructs sets all

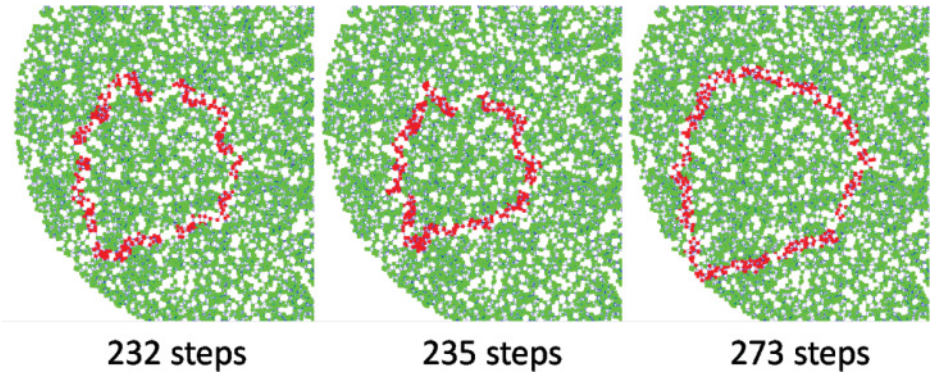

Fig. 1. The evolution of the volume-biased ESP in a huge graph. The red vertices represent the boundary of the current set in the process. Observe that the size of the boundary is significantly smaller than the size of the set.

of volume at most k, then we get a set of volume at most k and conductance at most $O(\sqrt{\log k/T})$.

So we just need to lower bound T. We prove a new technical result that for any set S of vertices, a t-step *lazy* random walk started at a random vertex of S is entirely contained in S with probability is at least $(1-\phi(S)/2)^t$ (see Theorem 3.1). Previously, only the lower bound $1-t\phi(S)/2$ was known. Using this lemma, we show that if the graph has a set A of conductance ϕ, then for any chosen $\epsilon < 1$ and $T = \Omega(\epsilon \log \mathrm{vol}(A)/\phi)$, with probability at least $1/\mathrm{vol}(A)^{\epsilon}$, all of the sets in the sample path of the process have volume at most $O(\mathrm{vol}(A)^{1+\epsilon})$ provided that the process starts from a random vertex of A. This is enough to guarantee that, with a probability of at least $1/\mathrm{vol}(A)^{\epsilon}$, the process constructs at least one set of conductance $O(\sqrt{\phi/\epsilon})$. To achieve a constant probability of success, we run $\mathrm{vol}(A)^{\epsilon}$-independent copies of the evolving set process simultaneously and stop as soon as one of the copies finds a small non-expanding set.

To bound the work/volume ratio, we combine a simple implementation trick with a nontrivial probabilistic analysis. We introduce an efficient method for simulating the volume-biased ESP that updates the vertices on the boundary of the current set and ignores the vertices in the interior. It turns out that the size of the boundary is "typically" significantly smaller than the size of the current set in the evolving set process (see Figure 1 for an example). Furthermore, the work required to generate a sample path using this method is proportional to the (two-sided) boundaries of the observed sets. Using a martingale argument, we prove that the expected ratio between the volume of the boundaries of the sets in a sample path and the volume of the last set of that path is $\mathrm{polylog}(n)\phi^{-1/2}$, which bounds the work/volume ratio of our algorithm.

2. PRELIMINARIES

2.1. Notation

Let $G = (V, E)$ be an undirected graph with $n := |V|$ vertices and $m := |E|$ edges. Let $d(v)$ be the degree of vertex $v \in V$. The *volume* of a subset $S \subseteq V$ is defined as the summation of the degree of vertices in S,

$$\mathrm{vol}(S) := \sum_{v \in S} d(v).$$

For $S, T \subseteq V$, let $E(S, V\backslash S) := \{(u, v) : u \in S, v \notin S\}$ be the set of the edges connecting S to $V\backslash S$, and we use $\partial(S)$ to denote the number of those edges. We also let $E(S) := \{\{u, v\} : u, v \in S\}$ be the set of edges inside S. The *conductance* of a set $S \subseteq V$ is defined

to be

$$\phi(S) := \partial(S)/\mathrm{vol}(S).$$

Observe that $\phi(V) = 0$. In the literature, the conductance of a set is sometimes defined to be $\partial(S)/\min(\mathrm{vol}(S), \mathrm{vol}(V \setminus S))$. Notice that the quantities are within a constant factor of each other if $\mathrm{vol}(S) = O(\mathrm{vol}(V \setminus S))$. Since here we are interested in finding *small* non-expanding sets, it is more convenient to work with the above definition.

We write $\ell^2(V)$ for the Hilbert space of functions $f : V \to \mathbb{R}$. For two functions $f, g : V \to \mathbb{C}$ we define the ordinary inner product:

$$\langle f, g \rangle := \sum_{v \in V} f(v) \cdot g(v).$$

We say f and g are orthogonal if $\langle f, g \rangle = 0$. The norm of a function $f \in \ell^2(V)$ is simply

$$\|f\| := \sqrt{\langle f, f \rangle}.$$

For a non-negative function $f : V \to \mathbb{R}$, we use $\sqrt{f}$ to denote the function function $\sqrt{f}(v) := \sqrt{f(v)}$. We use $\mathbf{1}$ to denote the all one function, and $\mathbf{1}_S$ as the indicator vector of a set $S \subseteq V$. We may abuse the notation and use $\mathbf{1}_v$ instead of $\mathbf{1}_{\{v\}}$ for a vertex $v \in V$.

For a matrix $M \in \mathbb{R}^{V \times V}$, we use $M^\intercal$ to denote the transpose of M. We use A to denote the adjacency matrix of G and D to denote the diagonal matrix of vertex degrees. We let I be the identity operator. For a set $S \subseteq V$, we use I_S to denote the identity operator on the set S, that is, for any function $f : V \to \mathbb{R}$, $I_S f(v) = f(v)$ if $v \in S$ and $I_S f(v) = 0$ otherwise. We use $\mathcal{L} := I - D^{-1/2} A D^{-1/2}$ to denote the normalized Laplacian matrix. It is well known that $\mathcal{L}$ is a positive semidefinite matrix with eigenvalues at most 2. Recall that a matrix $M \in \mathbb{R}^{V \times V}$ is positive semidefinite (PSD) if for any function $f : V \to \mathbb{R}$,

$$\langle Mf, f \rangle \geq 0.$$

Throughout the article, all logarithms are in base e.

2.2. Random Walks

We will consider the *lazy random walk* on G that each time step stays at the current vertex with probability $1/2$ and otherwise moves to the endpoint of a random edge attached to the current vertex. The transition kernel of the random walk is

$$p(u, v) = \begin{cases} 1/2d(u) & \text{if } (u, v) \in E \\ 1/2 & \text{if } u = v \\ 0 & \text{otherwise.} \end{cases}$$

Given a set S, we let $p(u, S)$ denote the probability of transitioning from u to some vertex in S,

$$p(u, S) := \sum_{v \in S} p(u, v) = \frac{1}{2}\left(\frac{e(u, S)}{d(u)} + \mathbb{I}[u \in S]\right),$$

where $e(u, S)$ denotes the number of edges between u and S, and $\mathbb{I}[.]$ is the indicator function of an event (see Figure 2 for an example).

We use $P := (D^{-1}A + I)/2$ as the transition probability matrix of this random walk. We define the following probability distribution function on a set $S \subseteq V$ of vertices:

$$\pi_S(v) := \begin{cases} d(v)/\mathrm{vol}(S) & \text{if } v \in S, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, $\pi(v) := \pi_V(v)$ is the stationary distribution of the walk, that is, $P^\intercal \pi = \pi$. We use $v \sim \pi_S$ to denote the distribution where v is sampled with probability $\pi_S(v)$.

$$p(u, S) = 2/8$$

$$p(v, S) = 3/8$$

$$p(w, S) = 6/8$$

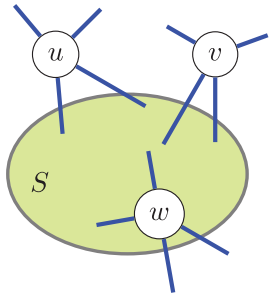

Fig. 2. An example of transition probabilities from several vertices to a set S.

2.3. The Evolving Set Process

The ESP is a Markov chain on subsets of the vertex set V. Given the current state S, the next state S_1 is chosen using the following rule: Pick a threshold Z uniformly at random from the interval $[0, 1]$, and let

$$S_1 = \{u : p(u, S) \geq Z\}. \tag{1}$$

Notice that $\emptyset$ and V are absorbing states for the process. Given a starting state $S_0 \subseteq V$, we write $\mathbf{P}_{S_0}[\cdot] := \mathbf{P}[\cdot \mid S_0]$ to denote the probability measure for the ESP Markov chain started from S_0. Similarly, we write $\mathbf{E}_{S_0}[\cdot]$ for the expectation. For a singleton set, we use the shorthand $\mathbf{P}_v[\cdot] = \mathbf{P}_{\{v\}}[\cdot]$. We define the transition kernel $\mathbf{K}(S, S') = \mathbf{P}_S[S_1 = S']$.

Morris and Peres [2003] used the evolving set process to prove upper bounds on the strong stationary time of Markov chains based on the conductance profile. They proved the following propositions to relate the conductance of a set in the ESP to the change in volume in the next step. The first proposition strengthens the fact that the sequence $(\mathrm{vol}(S_t))_{t\geq 0}$ is a martingale.

Proposition 2.1 (Morris and Peres [2003]). *Let Z be the uniform random variable used to generate S_1 from S in the ESP. Then,*

$$\mathbf{E}_S\left[\mathrm{vol}(S_1) \mid Z \leq \frac{1}{2}\right] = \mathrm{vol}(S) + \partial(S) = \mathrm{vol}(S)(1 + \phi(S)).$$

$$\mathbf{E}_S\left[\mathrm{vol}(S_1) \mid Z > \frac{1}{2}\right] = \mathrm{vol}(S) - \partial(S) = \mathrm{vol}(S)(1 - \phi(S)).$$

Proposition 2.2 (Morris and Peres [2003]). *The* growth gauge $\psi(S)$ *of a set S is defined by the following equation:*

$$1 - \psi(S) := \mathbf{E}_S\left[\sqrt{\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)}}\right].$$

For any set $S \subseteq V$, the growth gauge and conductance satisfy $\psi(S) \geq \phi(S)^2/8$.

PROOF. Using Theorem 2.1, for any set $S \subseteq V$, we have

$$\begin{aligned}\mathbf{E}_S\left[\sqrt{\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)}}\right] &= \frac{1}{2}\mathbf{E}_S\left[\sqrt{\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)}} \mid Z \le \frac{1}{2}\right] + \frac{1}{2}\mathbf{E}_S\left[\sqrt{\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)}} \mid Z > \frac{1}{2}\right] \\ &\le \frac{1}{2}\sqrt{\mathbf{E}_S\left[\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)} \mid Z \le \frac{1}{2}\right]} + \frac{1}{2}\sqrt{\mathbf{E}_S\left[\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)} \mid Z > \frac{1}{2}\right]} \\ &= \frac{1}{2}\sqrt{1+\phi(S)} + \frac{1}{2}\sqrt{1-\phi(S)} \le 1 - \phi^2(S)/8,\end{aligned}$$

where the first inequality follows from Jensen's inequality and the last inequality follows from the Taylor expansion of the square-root function. □

2.4. The Volume-Biased Evolving Set Process

The *volume-biased evolving set process* (volume-biased ESP) is a Markov chain on subsets of V with the following transition kernel:

$$\widehat{\mathbf{K}}(S, S') = \frac{\mathrm{vol}(S')}{\mathrm{vol}(S)}\mathbf{K}(S, S'), \tag{2}$$

where $\mathbf{K}(S, S')$ is the transition kernel for the ESP. We remark that $\widehat{\mathbf{K}}$ is the *Doob h-transform* of $\mathbf{K}$ with respect to vol (see chapter 17 of [Levin et al. 2006]) and that the volume-biased ESP is equivalent to the ESP conditioned to absorb in the state V. Given a starting state S_0, we write $\widehat{\mathbf{P}}_{S_0}[\cdot] := \widehat{\mathbf{P}}[\cdot \mid S_0]$ for the probability measure of the Markov chain. Similarly, we write $\widehat{\mathbf{E}}_{S_0}[\cdot]$ for the expectation.

The following proposition relates the volume-biased ESP and the ESP. This is a standard consequence of the Doob h-transform, but we include a proof for completeness.

PROPOSITION 2.3. *For any function f and any starting set $S_0 \neq \emptyset$,*

$$\widehat{\mathbf{E}}_{S_0}[f(S_0, \ldots, S_n)] = \mathbf{E}_{S_0}\left[\frac{\mathrm{vol}(S_n)}{\mathrm{vol}(S_0)} f(S_0, \ldots, S_n)\right]. \tag{3}$$

PROOF. Assume that $S_0 \neq \emptyset$. Let $\mathcal{C}$ be the collection of sample paths $(S_0, \ldots, S_t)$ such that $\widehat{\mathbf{P}}_{S_0}[S_1, \ldots, S_t] > 0$. If $(S_0, \ldots, S_t) \in \mathcal{C}$, then $\mathrm{vol}(S_i) > 0$ for all $i \in [0, t]$, so

$$\widehat{\mathbf{P}}_{S_0}[S_1, \ldots, S_t] = \prod_{i=0}^{t-1} \frac{\mathrm{vol}(S_{i+1})}{\mathrm{vol}(S_i)} \mathbf{P}_{S_i}[S_{i+1}] = \frac{\mathrm{vol}(S_t)}{\mathrm{vol}(S_0)} \mathbf{P}_{S_0}[S_1, \ldots, S_t].$$

Therefore,

$$\begin{aligned}\widehat{\mathbf{E}}_{S_0}[f(S_0, \ldots, S_t)] &= \sum_{(S_0, \ldots, S_t) \in \mathcal{C}} f(S_0, \ldots, S_t) \widehat{\mathbf{P}}_{S_0}[S_1, \ldots, S_t] \\ &= \sum_{(S_0, \ldots, S_t) \in \mathcal{C}} f(S_0, \ldots, S_t) \frac{\mathrm{vol}(S_t)}{\mathrm{vol}(S_0)} \mathbf{P}_{S_0}[S_1, \ldots, S_t] \\ &= \mathbf{E}_{S_0}\left[\frac{\mathrm{vol}(S_t)}{\mathrm{vol}(S_0)} f(S_0, \ldots, S_t)\right]. \quad \square\end{aligned}$$

2.5. The Diaconis-Fill Coupling

Diaconis and Fill [1990] introduced the following coupling between the random-walk process and the volume-biased ESP. Let (X_t, S_t) be a Markov chain, where X_t is a vertex

and $S_t \subseteq V$ is a subset of vertices. Let $\mathbf{P}^*[\cdot]$ be the probability measure for the Markov chain. Given a starting vertex v, let $X_0 = v$ and $S_0 = \{v\}$, and let $\mathbf{P}^*_v[\cdot] = \mathbf{P}^*[\cdot \mid X_0 = v, S_0 = \{v\}]$. Given the current state (X_t, S_t), the transition probabilities are defined as follows.

$$\mathbf{P}^*[X_{t+1} = v' \mid X_t = v, S_t = S] = p(v, v'),$$
$$\mathbf{P}^*[S_{t+1} = S' \mid S_t = S, X_{t+1} = v'] = \frac{\mathbf{K}(S, S')\mathbb{I}[v' \in S']}{\mathbf{P}[v' \in S_{t+1} \mid S_t = S]}.$$

In words, we first select X_{t+1} according to the random-walk transition kernel and then select S_{t+1} according to the ESP transition kernel restricted to sets that contain X_{t+1}. We define the transition kernel $\mathbf{K}^*((v, S), (v', S')) = \mathbf{P}^*[X_1 = v', S_1 = S' \mid X_0 = v, S_0 = S]$.

The following proposition shows that $\mathbf{P}^*[\cdot]$ is a coupling between the random-walk process and the volume-biased ESP, and, furthermore, that the distribution of X_t conditioned on $(S_0, \ldots, S_t)$ is the stationary distribution restricted to S_t. A proof of Theorem 2.4 is given in Levin et al. [2006, Chapter 17].

PROPOSITION 2.4 (DIACONIS AND FILL [1990]). *Let (X_t, S_t) be a Markov chain started from $(v, \{v\})$ with the transition kernel $\mathbf{K}^*$.*

(1) The sequence (X_t) is a Markov chain started from v with the transition kernel $p(.,.)$.
(2) The sequence (S_t) is a Markov chain started from $\{v\}$ with transition kernel $\widehat{\mathbf{K}}$.
(3) For any vertex u and time $t \geq 0$,

$$\mathbf{P}^*_v[X_t = u \mid S_1, \ldots, S_t] = \mathbb{I}[u \in S_t]\frac{d(u)}{\mathrm{vol}(S_t)}.$$

3. UPPER BOUNDS ON THE ESCAPE PROBABILITY OF RANDOM WALKS

In this section we establish new upper bounds on the escape probability of random walks. We show that for any $S \subseteq V$ a t-step lazy random walk started at a random (chosen proportional to degree) vertex of S remains entirely in S with probability at least $(1 - \phi(S)/2)^t$. Previously, only the lower bound $1 - t\phi(S)/2$ was known (see, e.g., Spielman and Teng [2013]).

For comparison, when $t = 1/\phi(S)$, the known bound would imply that the walk has probability at least $1/2$ of being entirely contained in S, with no guarantee being available in the case $t = 2/\phi(S)$, while our bound implies that for $t = (\epsilon \log n)/\phi$ the probability of being entirely contained in S is still at least $1/n^\epsilon$. Roughly speaking, the $\Omega(\log n)$ factor that we gain in the length of walks that we can study corresponds to our improvement in the conductance bound in Theorem 1.1, while the $1/n^\epsilon$ factor that we lose in the probability corresponds to the factor that we lose in the running time and the size of the output.

Let X_t be the random variable indicating the t-th step of the random walk started at v. Observe that the distribution of X_t is exactly $(P^\intercal)^t \mathbf{1}_v$. For a subset $S \subseteq V$, $v \in V$, and integer $t > 0$, we write

$$\mathrm{esc}(v, t, S) := \mathbb{P}_v\left[\cup_{i=0}^t X_i \notin S\right]$$

to denote the probability that the random walk started at v leaves S in the first t steps and $\mathrm{rem}(v, t, S) := 1 - \mathrm{esc}(v, t, S)$ as the probability that the walk stays entirely inside S. Now we are ready to describe the main result of this section.

PROPOSITION 3.1. *For any set $S \subseteq V$, and integer $t > 0$,*

$$\mathbb{E}_{v \sim \pi_S}[\mathrm{rem}(v, t, S)] \geq \left(1 - \frac{\phi(S)}{2}\right)\mathbb{E}_{v \sim \pi_S}[\mathrm{rem}(v, t-1, S)] \geq \cdots \geq \left(1 - \frac{\phi(S)}{2}\right)^t. \quad (4)$$

Furthermore, there is a universal constant $c_0 > 0$ *and a subset* $S' \subseteq S$*, such that* $\mathrm{vol}(S') \geq \mathrm{vol}(S)/2$*, and for all* $v \in S'$

$$\mathrm{rem}(v, t, S) \geq c_0 \left(1 - \frac{3\phi(S)}{2}\right)^t. \tag{5}$$

We remark that the second statement does not follow from a simple application of the Markov inequality to the first statement, as this is the case in Spielman and Teng [2013]. Whence, here both of the results incorporate non-trivial spectral arguments. Independent of our work, Kwok and Lau [2012] proved a weaker variant of the above statement. They showed that for any set S, there exists a vertex $v \in S$ such that

$$\mathrm{rem}(v, t, S) \geq (1 - \phi(S)/2)^t.$$

We refer interested readers to O'Donnell and Witmer [2012] for a different proof of the first part of the above proposition (i.e., Equation (4)).

As a corollary, we prove strong lower bounds on the uniform mixing time of random walks in Section 3.1. In the rest of this section we prove Theorem 3.1.

Proof. First, observe that

$$\mathrm{rem}(v, t, S) = \langle (I_S P^\intercal I_S)^t \mathbf{1}_v, \mathbf{1}_S \rangle = \langle \mathbf{1}_v, (I_S P I_S)^t \mathbf{1}_S \rangle.$$

So,

$$\mathbb{E}_{v \sim \pi_S}[\mathrm{rem}(v, t, S)] = \langle \pi_S, (I_S P I_S)^t \mathbf{1}_S \rangle. \tag{6}$$

Therefore, by a simple induction on t, Equation (4) is equivalent to the following equation:

$$\langle \pi_S, (I_S P I_S)^t \mathbf{1}_S \rangle \geq (1 - \phi(S)/2) \langle \pi_S, (I_S P I_S)^{t-1} \mathbf{1}_S \rangle. \tag{7}$$

Let $Q := D^{1/2} I_S P I_S D^{-1/2}$, and let $\sqrt{\pi_S} : V \to \mathbb{R}$, where $\sqrt{\pi_S}(v) = \sqrt{\pi_S(v)}$. First, we show that Equation (7) is equivalent to the following equation:

$$\langle \sqrt{\pi_S}, Q^t \sqrt{\pi_S} \rangle \geq \langle \sqrt{\pi_S}, Q \sqrt{\pi_S} \rangle \cdot \langle \sqrt{\pi_S}, Q^{t-1} \sqrt{\pi_S} \rangle. \tag{8}$$

Then we use Theorem 3.2 to prove the above equation. First, observe that by the definition of Q, for any $t > 0$,

$$\langle \pi_S, (I_S P I_S)^t \mathbf{1}_S \rangle = \langle \pi_S, D^{-1/2} Q^t D^{1/2} \mathbf{1}_S \rangle = \langle D^{-1/2} \pi_S, Q^t D^{1/2} \mathbf{1}_S \rangle = \langle \sqrt{\pi_S}, Q^t \sqrt{\pi_S} \rangle. \tag{9}$$

On the other hand,

$$\begin{aligned}
\langle \pi_S, (I_S P I_S) \mathbf{1}_S \rangle &= \left\langle \pi_S, \frac{1}{2}(D^{-1}A + I)\mathbf{1}_S \right\rangle \\
&= \frac{1}{2} \langle \pi_S, D^{-1} A \mathbf{1}_S \rangle + \frac{1}{2} \langle \pi_S, \mathbf{1}_S \rangle \\
&= \frac{1}{2\mathrm{vol}(S)} \langle \mathbf{1}_S, A \mathbf{1}_S \rangle + \frac{1}{2} \\
&= \frac{1}{2\mathrm{vol}(S)} 2|E(S)| + \frac{1}{2} \\
&= \frac{1}{2\mathrm{vol}(S)} (\mathrm{vol}(S) - \partial(S)) + \frac{1}{2} \\
&= 1 - \phi(S)/2,
\end{aligned} \tag{10}$$

where we used $E(S) := \{(u, v) \in E : u, v \in S\}$. Therefore, Equation (8) is equivalent to Equation (7) using Equation (9) and Equation (10). Next, we prove Equation (8) using

Theorem 3.2 when $f = \sqrt{\pi_S}$. We need to show $\|\sqrt{\pi_S}\| = 1$ and Q is PSD. First, observe that

$$\|\sqrt{\pi_S}\|^2 = \sum_{v \in V} \pi_S(v) = \sum_{v \in S} \frac{d(v)}{\mathrm{vol}(S)} = 1.$$

It remains to show that Q is PSD. We can rewrite Q as follows:

$$\begin{aligned} Q = D^{1/2} I_S P I_S D^{-1/2} &= \frac{1}{2} D^{1/2} I_S (D^{-1}A + I) I_S D^{-1/2} \\ &= \frac{1}{2} I_S (D^{-1/2} A D^{-1/2} + I) I_S = I_S (I - \mathcal{L}/2) I_S. \end{aligned} \tag{11}$$

Therefore, Q is PSD. This is because, on one hand, $I - \mathcal{L}/2$ is PSD as the maximum eigenvalue of $\mathcal{L}$ is at most 2, and, on the other hand, for any PSD matrix A and any symmetric matrix B, BAB is also PSD. This completes the proof of Equation (4).

It remains to prove Equation (5). We prove it by showing that for any set $T_1 \subseteq S$, of volume $\mathrm{vol}(T_1) \geq \mathrm{vol}(S)/2$, the random walk started at a randomly (proportional to degree) chosen vertex of T_1 remains in T_1 (and S), with probability at least $c_0(1 - 3\phi(S)/2)^t$, that is,

$$\langle \pi_{T_1}, (I_S P I_S)^t \mathbf{1}_{T_1} \rangle \geq c_0 \left(1 - \frac{3\phi(S)}{2}\right)^t. \tag{12}$$

Therefore, in any such set T_1, there is a vertex that satisfies Equation (5), so the volume of the set of vertices that satisfy Equation (5) is at least half of $\mathrm{vol}(S)$.

Using Equations (9) and (10), Equation (12) is equivalent to the following equation:

$$\langle \sqrt{\pi_{T_1}}, Q^t \sqrt{\pi_{T_1}} \rangle \geq c_0 \left(3\langle \sqrt{\pi_S}, Q\sqrt{\pi_S} \rangle - 2\right)^t. \tag{13}$$

We prove the above equation using Theorem 3.5. First, recall that from Equation (11), Q is PSD and its maximum eigenvalue is at most 1. Let $T_2 = S \backslash T_1$, and define

$$\begin{aligned} f_{T_1} &:= I_{T_1} \sqrt{\pi_S} = \sqrt{\mathrm{vol}(T_1)\pi_{T_1}/\mathrm{vol}(S)}, \\ f_{T_2} &:= I_{T_2} \sqrt{\pi_S} = \sqrt{\mathrm{vol}(T_2)\pi_{T_2}/\mathrm{vol}(S)}. \end{aligned} \tag{14}$$

Since $T_1 \cap T_2 = \emptyset$, $\langle f_{T_1}, f_{T_2} \rangle = 0$, and $\|f_{T_1} + f_{T_2}\| = \|\sqrt{\pi_S}\| = 1$. Furthermore, since $\mathrm{vol}(T_1) \geq \mathrm{vol}(S)/2 \geq \mathrm{vol}(T_2)$, $\|f_{T_1}\| \geq \|f_{T_2}\|$. Therefore, Q, f_{T_1}, f_{T_2} satisfy the requirements of Theorem 3.5. Finally, since

$$\langle \sqrt{\pi_{T_1}}, Q^t \sqrt{\pi_{T_1}} \rangle \geq \langle f_{T_1}, Q^t f_{T_1} \rangle,$$

Equation (12) follows from Theorem 3.5. This completes the proof of Theorem 3.1. □

Lemma 3.2. *Let $Q \in \mathbb{R}^{V \times V}$ be a PSD matrix. Then, for any $f : V \to \mathbb{R}$ with norm $\|f\| = 1$, and integer $t > 0$,*

$$\langle Q^t f, f \rangle \geq \langle Q^{t-1} f, f \rangle \langle Qf, f \rangle \geq \cdots \geq \langle Qf, f \rangle^t.$$

Proof. Since all of the inequalities in lemma's statement follow from the first inequality, we only prove the first inequality. Let $f_1, f_2, \ldots, f_n$ be the set of orthonormal eigenfunctions of Q with the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. For any integer $t \geq 1$, using the spectral theorem,

$$\langle Q^t f, f \rangle = \sum_{i=1}^{n} \langle f, f_i \rangle^2 \lambda_i^t. \tag{15}$$

On the other hand, since $\{f_1, \ldots, f_n\}$ is an orthonormal system, we have

$$\sum_{i=1}^{n} \langle f, f_i \rangle^2 = \|f\|^2 = 1.$$

Therefore, using Chebyshev's sum inequality, see Theorem 3.3 below, we can write

$$\sum_{i=1}^{n} \langle f, f_i \rangle^2 \left(\lambda_i^{t-1}\right) \cdot (\lambda_i) \geq \sum_{i=1}^{n} \langle f, f_i \rangle^2 \lambda_i^{t-1} \cdot \sum_{i=1}^{n} \langle f, f_i \rangle^2 \lambda_i.$$

In the above we are using that Q is PSD, that is, $\lambda_i \geq 0$ for all i, which implies that $\lambda_i^{t-1} < \lambda_j^{t-1}$ if and only if $\lambda_i < \lambda_j$. By Equation (15), the above inequality is equivalent to the lemma's conclusion. □

FACT 3.3 (CHEBYSHEV'S SUM INEQUALITY). *Let $a_1 \geq a_2 \geq \cdots \geq a_n$, and $b_1 \geq b_2 \geq \cdots \geq b_n$. Then, for any probability distribution $\mathcal{D}$ defined on $1, 2, \ldots, n$,*

$$\mathbb{E}_{i \sim \mathcal{D}} [a_i \cdot b_i] \geq \mathbb{E}_{i \sim \mathcal{D}} [a_i] \cdot \mathbb{E}_{i \sim \mathcal{D}} [b_i].$$

Variants of Theorem 3.2 have been previously studied in the literature. For example, the following lemma is proved in Mulholland and Smith [1959] and Blakley and Roy [1965] (see also London [1966] for various generalizations).

LEMMA 3.4 ([MULHOLLAND AND SMITH 1959; BLAKLEY AND ROY 1965]). *For any non-negative symmetric matrix $Q \in \mathbb{R}^{V \times V}$, and any non-negative unit norm $f : V \to \mathbb{R}$, and any integer $t > 0$,*

$$\langle Q^t f, f \rangle \geq \langle Qf, f \rangle^t.$$

We remark that for even values of t the above lemma can be proven by an argument similar to Theorem 3.2 using the Jensen's inequality instead of the Chebyshev's sum inequality.

LEMMA 3.5. *There is a universal constant $c_0 > 0$ satisfying the following. Let $Q \in \mathbb{R}^{n \times n}$ be a PSD matrix with eigenvalues at most 1. For any integer $t > 0$, and any $f, g \in \mathbb{R}^n$ such that $\langle f, g \rangle = 0$, $\|f + g\| = 1$, and $\|f\| \geq \|g\|$,*

$$\langle Q^t f, f \rangle \geq c_0 (3 \langle Q(f + g), (f + g) \rangle - 2)^t.$$

PROOF. Let $h := f + g$. Since f is orthogonal to g, we have $\|g\|^2 \leq 1/2 \leq \|f\|^2$. Let $f_1, f_2, \ldots, f_n$ be the set of orthonormal eigenfunctions of Q with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let $\alpha > 0$ be a constant that will be fixed later in the proof. Define $B := \{i : |\langle f, f_i \rangle| \geq \alpha |\langle g, f_i \rangle|\}$. First, observe that,

$$\langle Q^t f, f \rangle = \sum_{i=1}^{n} \langle f, f_i \rangle^2 \lambda_i^t \geq \sum_{i \in B} \langle f, f_i \rangle^2 \lambda_i^t \geq \frac{1}{(1 + 1/\alpha)^2} \sum_{i \in B} \langle h, f_i \rangle^2 \lambda_i^t, \tag{16}$$

where the equality follows from Equation (15), the first inequality uses the assumption that Q is PSD, and the last inequality follows from the definition of B, that is, for any $i \in B$, $\langle f, f_i \rangle^2 \geq (\langle h, f_i \rangle / (1 + 1/\alpha))^2$. Let $s := \sum_{i \in B} \langle h, f_i \rangle^2$. First, we lower bound s by a function of α, and then we use the Jensen's inequality to lower bound the right-hand side of (16),

$$s = \sum_{i \in B} \langle h, f_i \rangle^2 = 1 - \sum_{i \notin B} \langle h, f_i \rangle^2 \geq 1 - (1 + \alpha)^2 \sum_{i \notin B} \langle g, f_i \rangle^2,$$

where the second equality uses $\|h\| = 1$, and the last inequality follows from the fact that for any $i \notin B$,

$$\langle h, f_i\rangle^2 \le (|\langle f, f_i\rangle| + |\langle g, f_i\rangle|)^2 \le (\alpha|\langle g, f_i\rangle| + |\langle g, f_i\rangle|)^2 = (1+\alpha)^2\langle g, f_i\rangle^2.$$

Now, since $\|g\|^2 \le 1/2$, we lower bound s as follows:

$$s \ge 1 - (1+\alpha)^2\|g\|^2 \ge \frac{1-\alpha^2-2\alpha}{2}. \tag{17}$$

On the other hand, since Q is PSD, using Jensen's inequality,

$$\begin{aligned}\frac{1}{s}\sum_{i\in B}\langle h, f_i\rangle^2\lambda_i^t &\ge \left(\frac{1}{s}\sum_{i\in B}\langle h, f_i\rangle^2\lambda_i\right)^t \\ &\ge \left(\frac{\sum_{i=1}^n\langle h, f_i\rangle^2\lambda_i - (1-s)}{s}\right)^t \\ &\ge \left(1 - \frac{1-\langle Qh, h\rangle}{(1-\alpha^2-2\alpha)/2}\right)^t,\end{aligned} \tag{18}$$

where the second inequality follows from the assumptions that $\max_{1\le i\le n}\lambda_i \le 1$, and that $\|h\| = 1$, and the last inequality follows from Equation (17) and that $\langle Qh, h\rangle \le 1$. Putting Equations (16) and (18), and letting $\alpha = 0.154$, we get

$$\langle Q^t f, f\rangle \ge \frac{(1-\alpha^2-2\alpha)/2}{(1+1/\alpha)^2}\left(1 - \frac{1-\langle Qh, h\rangle}{(1-\alpha^2-2\alpha)/2}\right)^t \ge \frac{1}{200}(3\langle Qh, h\rangle - 2))^t.$$

Letting $c_0 := 1/200$ proves the lemma. □

3.1. Lower Bounds on Uniform Mixing Time of Random Walks

In this section we prove lower bounds on the mixing time of reversible Markov chains. Since any reversible finite state Markov chain can be realized as a random walk on a weighted undirected graph, for simplicity of notations, we model the Markov chain as a random walk on a weighted graph G.

The ϵ-*mixing time* of a random walk in total variation distance is defined as

$$\tau_V(\epsilon) := \min\left\{t : \sum_{v\in V}|\mathbb{P}_u[X_t = v] - \pi(v)| < \epsilon, \forall u \in V\right\}.$$

The mixing time of the chain is usually defined as $\tau_V(1/4)$. The ϵ-*uniform* mixing time of the chain is defined as

$$\tau(\epsilon) := \min\left\{t : \left|1 - \frac{\mathbb{P}_u[X_t = v]}{\pi(v)}\right| < \epsilon, \forall u, v \in V\right\}. \tag{19}$$

It is worth noting that the uniform mixing time can be considerably larger than the mixing time in total variation distance.

Let $\phi(G) := \min_{S:\mathrm{vol}(S)\le\mathrm{vol}(V)/2}\phi(S)$. Jerrum and Sinclair [1989] prove that the ϵ-uniform mixing time of any *lazy* random walk is bounded from above by

$$\tau(\epsilon) \le \frac{2}{\phi(G)^2}\left(\log\frac{1}{\min_v\pi(v)} + \log\frac{1}{\epsilon}\right).$$

On the other hand, one can use $\phi(G)$ as the *bottleneck ratio* to provide a lowerbound on the mixing time of the random walks. It follows from the Cheeger's inequality that

(see, e.g., Levin et al. [2006, Section 7.2]),

$$\tau_V(1/4) \geq \frac{1}{4\phi(G)}.$$

In the next proposition we prove stronger lower bounds on the uniform mixing time of any reversible Markov chain.

PROPOSITION 3.6. *For any (weighted) graph $G = (V, E)$, any $S \subseteq V$ with* $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$*, and* $0 < \epsilon < 1$*, if* $\phi(S) \leq 0.7$*, then*

$$\tau(\epsilon) \geq \frac{\log(\mathrm{vol}(V)/(1+\epsilon)\mathrm{vol}(S))}{2\phi(S)} - 1.$$

PROOF. Let $t \geq -\log((1+\epsilon)\pi(S))/2\phi(S) - 1$ be an integer. Since the random walk is not necessarily a lazy walk, P is not necessarily a PSD operator, so we cannot directly apply Theorem 3.1. Instead we can use Theorem 3.4 that does not need a PSD assumption. So, for $Q = D^{1/2} I_S P I_S D^{-1/2}$,

$$\mathbb{E}_{v\sim\pi_S}\left[\mathrm{rem}(v, t, S)\right] = \langle Q^t\sqrt{\pi_S}, \sqrt{\pi_S}\rangle \geq \langle Q\sqrt{\pi_S}, \sqrt{\pi_S}\rangle^t = (1-\phi(S))^t,$$

where the first equality follows from Equations (6) and (9), and the last equality follows from Equation (10). Thus, there exists a vertex $u \in S$ such that

$$\mathrm{rem}(u, t, S) \geq (1-\phi(S))^t \geq (1+\epsilon)\pi(S),$$

where in the last inequality we used the assumption that $\phi(S) \leq 0.7$. Since $\mathbb{P}_{X_0=u}[X_t \in S] \geq \mathrm{rem}(u, t, S)$, there is a vertex $v \in S$ such that,

$$\frac{P^t(u, v)}{\pi(v)} \geq \frac{\mathbb{P}_{X_0\sim u}[X_t \in S]}{\pi(S)} \geq \frac{(1+\epsilon)\pi(S)}{\pi(S)} = 1+\epsilon,$$

where the first inequality uses $\mathbb{P}_{X_0\sim u}[X_t \in S] = \sum_{v\in S} P^t(u, v)$. Therefore, $\frac{|P^t(u,v)-\pi(v)|}{\pi(v)} \geq \epsilon$, and using Equation (19), $\tau(\epsilon) \geq t$. □

We remark that the above bound only holds for the uniform mixing time, and it can provide much stronger lower bound than the bottleneck ratio, if $\mathrm{vol}(S) \ll \mathrm{vol}(V)$.

4. ALMOST OPTIMAL LOCAL GRAPH CLUSTERING

In this section, we show how to find sets with small conductance by generating sample paths from the volume-biased ESP. The following theorem is the main result of this section. We show that for any target set A and $\epsilon > 0$, if we simulate the volume-biased ESP starting from certain vertices of A for $T = \Theta(\epsilon \log \mathrm{vol}(A)/\phi(A))$ steps, then with probability of at least $\Omega(\mathrm{vol}(A)^{-\epsilon})$ one of the observed states has conductance $O(\sqrt{\phi(A)/\epsilon})$ and volume $O(\mathrm{vol}(A)^{1+\epsilon})$.

THEOREM 4.1. *Let $A \subset V$ be a set of vertices of volume* $\mathrm{vol}(A) \leq k$ *and conductance $\phi(A) \leq \phi$. For any $\epsilon \in (0, 1)$, fix $T = \epsilon \log k/3\phi$. There is a universal constant $c_0 > 0$, and a subset $A' \subseteq A$ of volume* $\mathrm{vol}(A') \geq \mathrm{vol}(A)/2$ *for which the following holds. For any $v \in A'$, with probability of at least $c_0 k^{-\epsilon}/8$, a sample path $(S_1, S_2, \ldots, S_T)$ of the volume biased ESP started from $S_0 = \{v\}$ satisfies the following:*

(1) For some $t \in [0, T]$, $\phi(S_t) \leq \Phi_\epsilon(\phi)$, where $\Phi_\epsilon(\phi) := \sqrt{100(1-\log c_0)\phi/\epsilon}$,
(2) For all $i \in [0, T]$, $\mathrm{vol}(S_i \cap A) \geq c_0 k^{-\epsilon}\mathrm{vol}(S_i)/2$*, and henceforth,*
(3) For all $i \in [0, T]$, $\mathrm{vol}(S_i) \leq \mathcal{K}_\epsilon(k)$*, where $\mathcal{K}_\epsilon(k) := 2k^{1+\epsilon}/c_0$.*

The proof of Theorem 4.1 is at the end of this section, after we present two necessary lemmas. Consider a sample path from the volume-biased ESP. The following lemma

shows it is unlikely for the sample path to contain many sets with large conductance. In particular, we show that, for any stopping time τ, with probability of at least $1-1/\alpha$, there is a set of conductance of at most $O(\sqrt{\frac{1}{\tau}\log(\alpha \cdot \mathrm{vol}(S_\tau))})$. Intuitively, this is true because at each step the quantity $\mathrm{vol}(S_t)$ tends to increase at a rate that depends on $\phi(S_t)$. Eventually, the sample path will absorb in the state V, whose conductance is $\phi(V)=0$.

LEMMA 4.2. *For any starting set $S_0 \subseteq V$ and any stopping time τ for the volume-biased ESP,*

$$\widehat{\mathbf{E}}_{S_0}\left[\sum_{i=0}^{\tau}\phi(S_i)^2\right] \le 4\widehat{\mathbf{E}}\left[\log\frac{\mathrm{vol}(S_\tau)}{\mathrm{vol}(S_0)}\right] \le 4\log \mathrm{vol}(V).$$

Moreover, for any $\alpha > 1$,

$$\widehat{\mathbf{P}}_{S_0}\left[\sum_{i=1}^{\tau}\phi^2(S_i) \le 8\left(\log\alpha + \log\frac{\mathrm{vol}(S_\tau)}{\mathrm{vol}(S_0)}\right)\right] \ge 1-\frac{1}{\alpha}.$$

PROOF. The idea is to define a martingale M_t using the rate of change at each step $1-\psi(S_t)$ and then use the optional stopping theorem to lower bound the growth in the size of the set at the stopping time τ. We define

$$M_t := F_t\frac{\sqrt{\mathrm{vol}(S_0)}}{\sqrt{\mathrm{vol}(S_t)}}, \quad \text{where } F_t := \prod_{j=0}^{t-1}(1-\psi(S_j))^{-1}, \text{ and } F_0 := 1, \tag{20}$$

First, we verify that (M_t) is a martingale in the volume-biased ESP:

$$\begin{aligned}
\widehat{\mathbf{E}}[M_t|S_0,\ldots,S_{t-1}] &= \sqrt{\mathrm{vol}(S_0)}F_t\widehat{\mathbf{E}}\left[\frac{1}{\sqrt{\mathrm{vol}(S_t)}}\,\middle|\,S_{t-1}\right]\\
&= F_t\frac{\sqrt{\mathrm{vol}(S_0)}}{\sqrt{\mathrm{vol}(S_{t-1})}}\widehat{\mathbf{E}}_{S_{t-1}}\left[\frac{\sqrt{\mathrm{vol}(S_{t-1})}}{\sqrt{\mathrm{vol}(S_t)}}\right]\\
&= F_t\frac{\sqrt{\mathrm{vol}(S_0)}}{\sqrt{\mathrm{vol}(S_{t-1})}}\mathbf{E}_{S_{t-1}}\left[\frac{\sqrt{\mathrm{vol}(S_t)}}{\sqrt{\mathrm{vol}(S_{t-1})}}\right]\\
&= F_t\frac{\sqrt{\mathrm{vol}(S_0)}}{\sqrt{\mathrm{vol}(S_{t-1})}}(1-\psi(S_{t-1}))\\
&= F_{t-1}\frac{\sqrt{\mathrm{vol}(S_0)}}{\sqrt{\mathrm{vol}(S_{t-1})}} = M_{t-1},
\end{aligned}$$

where the third equality follows from Theorem 2.3; also see Theorem 2.2 for the definition of growth gauge, $\psi(.)$. Let τ be a stopping time for the volume-biased ESP. By use of the optional stopping theorem for non-negative martingales (see Williams [1991]), we have

$$\widehat{\mathbf{E}}[M_\tau] = M_0 = 1.$$

Then, by applying Jensen's inequality, and Markov's Inequality to the above equation, we have

$$\widehat{\mathbf{E}}[\log M_\tau] \le \log(\widehat{\mathbf{E}}[M_\tau]) = 0, \tag{21}$$

$$\widehat{\mathbf{P}}[\log M_\tau \le \log \alpha] \ge 1 - \frac{1}{\alpha}. \tag{22}$$

By the definition of M_τ,

$$\begin{aligned}
\log M_\tau &= \log F_\tau + \frac{1}{2}\log\frac{\mathrm{vol}(S_0)}{\mathrm{vol}(S_\tau)} \\
&= \log \prod_{i=0}^{\tau-1} \frac{1}{1-\psi(S_i)} - \frac{1}{2}\log\frac{\mathrm{vol}(S_\tau)}{\mathrm{vol}(S_0)} \\
&\ge \sum_{i=0}^{\tau-1} \psi(S_i) - \frac{1}{2}\log\frac{\mathrm{vol}(S_\tau)}{\mathrm{vol}(S_0)} \\
&\ge \frac{1}{8}\sum_{i=0}^{\tau-1} \phi^2(S_i) - \frac{1}{2}\log\frac{\mathrm{vol}(S_\tau)}{\mathrm{vol}(S_0)},
\end{aligned}$$

where the first inequality follows from the fact that $1/(1-\psi(S_i)) \ge e^{\psi(S_i)}$, and the last inequality follows from Theorem 2.2. The first statement of the lemma follows from taking expectation of both sides of the above inequality and using (21), and the second statement follows from putting (22) and the above equation together. □

The previous lemma shows that for any $k, \phi > 0$, if we can run the process for $T \approx \epsilon \log k/\phi$ steps without observing a set larger than $k^{O(1)}$, then, with probability $1 - 1/k$, one of the sets in the sample path must have a conductance of $O(\sqrt{\phi/\epsilon})$, which is what we are looking for. Using the following lemma, together with Theorem 3.1, we can show the above event occurs with some non-zero probability. That is, for any $\epsilon < 1$, with probability at least $\approx k^{-\epsilon}$, the volume of all sets in the sample path of the process are at most $O(k^{1+\epsilon})$.

Lemma 4.3. *For any set $A \subseteq V$, vertex $v \in A$, and integer $T \ge 0$, the following holds for all $\beta > 0$:*

$$\widehat{\mathbf{P}}_v\left[\max_{t \le T} \frac{\mathrm{vol}(S_t \backslash A)}{\mathrm{vol}(S_t)} > \beta\, \mathrm{esc}(v, T, A)\right] < \frac{1}{\beta}.$$

Proof. We use the Diaconis-Fill coupling between the volume-biased ESP Markov chain (S_t) and the random-walk Markov chain (X_t). Recall that for any $t \ge 0$,

$$\mathbf{P}^*[X_t = u \mid S_0, \ldots, S_t] = \frac{d(u)}{\mathrm{vol}(S_t)}\mathbb{I}[u \in S_t].$$

Fix a value $\lambda \in [0, 1]$ and let τ be the first time t when $\mathrm{vol}(S_t \backslash A) > \lambda \mathrm{vol}(S_t)$, or let $\tau = \infty$ if this does not occur. Consider the probability that $X_\tau \notin A$, conditioned on S_τ:

$$\mathbf{P}^*[X_\tau \notin A \mid S_\tau = S] = \sum_{u \in S \backslash A} \frac{d(u)}{\mathrm{vol}(S)} = \frac{\mathrm{vol}(S \backslash A)}{\mathrm{vol}(S)}.$$

By the definition of τ, we have $\mathbf{P}^*[X_\tau \notin A \mid \tau \le T] > \lambda$, so

$$\begin{aligned}
\mathrm{esc}(v, T, A) &= \mathbf{P}^*\left[\cup_{i=0}^{T} (X_i \notin A)\right] \\
&\ge \mathbf{P}^*[X_\tau \notin A \wedge \tau \le T] \\
&= \mathbf{P}^*[X_\tau \notin A \mid \tau \le T]\mathbf{P}^*[\tau \le T] \\
&\ge \lambda \mathbf{P}^*[\tau \le T].
\end{aligned}$$

Therefore

$$\widehat{\mathbf{P}}_v\left[\max_{t\le T}\frac{\mathrm{vol}(S_t\backslash A)}{\mathrm{vol}(S_t)}>\lambda\right]=\mathbf{P}^*[\tau\le T]<\frac{\mathrm{esc}(v,T,A)}{\lambda}.$$

The lemma follows by taking $\lambda=\beta\,\mathrm{esc}(v,T,A)$. □

We now combine the results from this section to prove Theorem 4.1. The proof simply follows from applying the union bound to argue that with some non-zero probability there is a small non-expanding set in the sample path of the volume-biased ESP.

Proof of Theorem 4.1. If $\phi\ge 1/2$, then we simply return v. Otherwise, assume $\phi<1/2$. We let c_0 be the constant defined in Theorem 3.1. Let A' be the set of vertices $v\in A$ such that

$$\mathrm{rem}(v,T,A)\ge c_0\left(1-\frac{3\phi(A)}{2}\right)^T.$$

By use of Theorem 3.1, $\mathrm{vol}(A')\ge\mathrm{vol}(A)/2$. In the rest of the proof, let v be a vertex in A'. We have

$$\mathrm{esc}(v,T,A)\le 1-c_0\left(1-\frac{3\phi(A)}{2}\right)^T\le 1-c_0\left(1-\frac{3\phi}{2}\right)^{\frac{\epsilon\log k}{3\phi}}\le 1-c_0k^{-\epsilon},$$

where we used $\phi\le 1/2$. Now, let $\beta:=1+c_0k^{-\epsilon}/2$. By use of Theorem 4.3, we have

$$\widehat{\mathbf{P}}_v\left[\max_{t\le T}\frac{\mathrm{vol}(S_t\backslash A)}{\mathrm{vol}(S_t)}\le\beta\,\mathrm{esc}(v,T,A)\le 1-\frac{c_0k^{-\epsilon}}{2}\right]\ge 1-\frac{1}{\beta}\ge\frac{c_0k^{-\epsilon}}{4}.$$

Since for any $S\subset V$, $\mathrm{vol}(S\backslash A)+\mathrm{vol}(S\cap A)=\mathrm{vol}(S)$, we have

$$\widehat{\mathbf{P}}_v\left[\min_{t\le T}\frac{\mathrm{vol}(S_t\cap A)}{\mathrm{vol}(S_t)}\ge\frac{c_0k^{-\epsilon}}{2}\right]\ge\frac{c_0k^{-\epsilon}}{4}.$$

On the other hand, let $\alpha:=k$. By use of Theorem 4.2, with probability $1-1/k$, for some $t\in[0,T]$,

$$\phi^2(S_t)\le\frac{1}{T}\sum_{i=0}^{T}\phi^2(S_i)\le\frac{8(\log k+\log\mathrm{vol}(S_T))}{T}.$$

Therefore, since $\epsilon<1$, using the union bound we have

$$\widehat{\mathbf{P}}_v\left[\min_{t\le T}\frac{\mathrm{vol}(S_t\cap A)}{\mathrm{vol}(S_t)}\ge\frac{c_0k^{-\epsilon}}{2}\ \bigwedge\ \exists\, t:\ \phi(S_t)\le\sqrt{\frac{8(\log k+\log\mathrm{vol}(S_T))}{T}}\right]\ge\frac{c_0k^{-\epsilon}}{8}$$

Finally, since for any set $S\subseteq V$, $\mathrm{vol}(S\cap A)\le\mathrm{vol}(A)\le k$, in the above event, $\mathrm{vol}(S_T)\le\frac{2k^{1+\epsilon}}{c_0}$. Therefore,

$$\phi(S_t)\le\sqrt{\frac{8(\log k+\log(2k^{1+\epsilon}/c_0))}{T}}\le\sqrt{\frac{100(1-\log c_0)\phi}{\epsilon}},$$

which completes the proof. □

5. SIMULATING THE VOLUME-BIASED EVOLVING SET PROCESS

In this section, we describe a subroutine `GenerateSample` that simulates the volume-biased ESP until a certain stopping time τ is reached, generating a sample path $(S_0,\ldots,S_\tau)$ and producing as output the set S_τ. We choose τ to be the first time that

S_t has sufficiently small conductance, while the volume of S_t is sufficiently small. We assign a cost to each sample path that depends on the boundaries of the sets in the path and the difference in volume between successive sets. We then show that the work performed by `GenerateSample` is $O(\text{polylog}(n))$ times the cost of the sample path it generates.

At the end of this section we prove the main Theorem 1.1. We design an algorithm, `EvoPar`, that runs multiple copies of the volume-biased ESP and returns the first small non-expanding set S_τ that `GenerateSample` finds. To bound the work/volume ratio of `EvoPar`, we directly bound the expected ratio between the cost of $(S_0, \ldots, S_\tau)$ and the volume of S_τ.

Definition 5.1. The *cost* of a sample path $(S_0, \ldots, S_t)$ is

$$\text{cost}(S_0, \ldots, S_t) := \text{vol}(S_0) + \sum_{i=1}^{t} \big(\text{vol}(S_i \Delta S_{i-1}) + \partial(S_{i-1})\big), \tag{23}$$

where Δ denotes the symmetric difference between two sets.

Definition 5.2. Given integers T, B, $\mathcal{K}$, and $\Phi \in [0, 1]$ let $\tau(T, B, \Phi, \mathcal{K})$ be the first time one of the following occurs:

(1) $\phi(S_t) \leq \Phi$, and $\text{vol}(S_t) \leq \mathcal{K}$.
(2) $t = T$ or $\text{cost}(S_0, \ldots, S_t) > B$.

The following proposition shows that `GenerateSample` generates a sample path from the volume-biased ESP with stopping rule $\tau(T, B, \Phi, \mathcal{K})$ and that its complexity is at most $O(\log n)$ times the cost of the sample path it generates. The complexity is also bounded by $O(B \log n)$.

Proposition 5.3. *The algorithm* `GenerateSample`$(v, T, B, \Phi, \mathcal{K})$ *takes as input a vertex* v *and integers* $T, B, \mathcal{K} \geq 0$ *and* $\Phi \in [0, 1]$. *Let* $S_0 = \{x\}$ *and let* $\tau = \tau(T, B, \Phi, \mathcal{K})$. *The algorithm generates a sample path* $(S_0, \ldots, S_\tau)$ *and outputs the last set* S_τ. *The following hold.*

(1) *The probability that* `GenerateSample` *generates the sample path* $(S_0, \ldots, S_\tau)$ *is* $\widehat{\mathbf{P}}_v[S_0, \ldots, S_\tau]$.
(2) *If* `GenerateSample` *generates* $(S_0, \ldots, S_\tau)$, *then its output is* S_τ *and its complexity is*

$$O(\log n) \min(B, \text{cost}(S_0, \ldots, S_\tau)).$$

The description of `GenerateSample` and the proof of Theorem 5.3 are given in Section 5.1.

In the following theorem, we bound the expected ratio between the cost of the sample path $(S_0, \ldots, S_\tau)$ and the volume of S_τ, which is later used to bound the work/volume ratio for `EvoPar`.

Theorem 5.4. *For any starting set* S_0 *and any stopping time* τ *that is bounded by* T, *we have*

$$\widehat{\mathbf{E}}_{S_0}\left[\frac{\text{cost}(S_0, \ldots, S_\tau)}{\text{vol}(S_\tau)}\right] \leq 1 + 4\sqrt{T \log \text{vol}(V)}.$$

The proof uses a martingale argument and the transform between the ESP and volume-biased ESP. It bounds the work/volume ratio in a more direct way than previous local partitioning algorithms, which required the user to guess the approximate volume of the output set [Spielman and Teng 2004, 2013; Andersen et al. 2006].

PROOF. Let c_i be the cost of the step in which S_i is selected,

$$c_i := \mathrm{vol}(S_i \Delta S_{i-1}) + \partial(S_{i-1}).$$

We define $c_0 := \mathrm{vol}(S_0)$ and recall that $\mathrm{cost}(S_0, \ldots, S_t) = c_0 + \cdots + c_t$.

Consider the conditional expectation of c_i in the ESP. We have

$$\mathbf{E}[c_i \mid S_{i-1}] = \mathbf{E}[\mathrm{vol}(S_i \Delta S_{i-1}) \mid S_{i-1}] + \partial(S_{i-1}).$$

We now compute the expected volume of the symmetric difference. Let Z be the uniform random threshold used to select S_i from S_{i-1} in the ESP, and recall that $S_i \subseteq S_{i-1}$ when $Z \geq 1/2$ and $S_{i-1} \subseteq S_i$ when $Z < 1/2$. By Theorem 2.1,

$$\begin{aligned}
\mathbf{E}[\mathrm{vol}(S_i \Delta S_{i-1}) \mid S_{i-1}] &= \mathbf{E}[|\mathrm{vol}(S_i) - \mathrm{vol}(S_{i-1})| \mid S_{i-1}] \\
&= \frac{1}{2}\mathbf{E}\left[\mathrm{vol}(S_i) - \mathrm{vol}(S_{i-1}) \,\middle|\, S_{i-1}, Z < \frac{1}{2}\right] \\
&\quad + \frac{1}{2}\mathbf{E}\left[\mathrm{vol}(S_{i-1}) - \mathrm{vol}(S_i) \,\middle|\, S_{i-1}, Z \geq \frac{1}{2}\right] \\
&= \frac{1}{2}\partial(S_{i-1}) + \frac{1}{2}\partial(S_{i-1}) = \partial(S_{i-1}).
\end{aligned}$$

Therefore, $\mathbf{E}[c_i \mid S_{i-1}] = 2\partial(S_{i-1})$.

Let $R_t := \frac{\mathrm{cost}(S_0, \ldots, S_t)}{\mathrm{vol}(S_t)}$, and consider the conditional expectation of R_t in the volume-biased ESP. Using Theorem 2.3,

$$\begin{aligned}
\widehat{\mathbf{E}}[R_t \mid S_0, \ldots, S_{t-1}] &= \mathbf{E}\left[\frac{\mathrm{cost}(S_0, \ldots, S_t)}{\mathrm{vol}(S_t)} \frac{\mathrm{vol}(S_t)}{\mathrm{vol}(S_{t-1})} \,\middle|\, S_0, \ldots, S_{t-1}\right] \\
&= \frac{1}{\mathrm{vol}(S_{t-1})}(\mathrm{cost}(S_0, \ldots, S_{t-1}) + \mathbf{E}[c_t \mid S_0, \ldots, S_{t-1}]) \\
&= \frac{1}{\mathrm{vol}(S_{t-1})}(\mathrm{cost}(S_0, \ldots, S_{t-1}) + 2\partial(S_{t-1})) = R_{t-1} + 2\phi(S_{t-1}).
\end{aligned}$$

We define

$$M_t := R_t - Q_t, \qquad \text{where} \qquad Q_t := 1 + 2\sum_{i=1}^{t} \phi(S_{i-1}).$$

By construction, (M_t) is a martingale in the volume-biased ESP. Notice that $M_0 = R_0 - 1 = 0$. Now let τ be an arbitrary stopping time that is bounded by T. By the optional stopping theorem for martingales (see Williams [1991]), we have $\widehat{\mathbf{E}}[M_\tau] = M_0 = 0$, so $\widehat{\mathbf{E}}[R_\tau] = \widehat{\mathbf{E}}[Q_\tau]$. By Cauchy-Schwarz,

$$\sum_{i=0}^{T-1} \phi(S_i) \leq \sqrt{T} \sqrt{\sum_{i=0}^{T-1} \phi^2(S_i)}.$$

By Jensen's inequality,

$$\begin{aligned}\widehat{\mathbf{E}}[R_\tau] = \widehat{\mathbf{E}}[Q_\tau] \le \widehat{\mathbf{E}}[Q_T] &= 1 + 2\sum_{i=0}^{T-1}\phi(S_i)\\ &\le 1 + 2\sqrt{T}\,\widehat{\mathbf{E}}\left[\sqrt{\sum_{i=0}^{T-1}\phi^2(S_i)}\right]\\ &\le 1 + 2\sqrt{T}\sqrt{\widehat{\mathbf{E}}\left[\sum_{i=0}^{T-1}\phi^2(S_i)\right]} \le 1 + 4\sqrt{T\log \mathrm{vol}(V)}.\end{aligned}$$

In the last step, we used Theorem 4.2. □

5.1. Implementing Volume-Biased ESP

In this section we describe the subroutine `GenerateSample` and we prove Theorem 5.3. At a high level, `GenerateSample` simulates the volume-biased ESP by updating the boundary of the current set at each step. We define $\delta(S)$ to be the *two-sided vertex boundary* of S,

$$\delta(S) := \{v : v \in S \wedge e(v, \overline{S}) > 0\} \cup \{v : v \in \overline{S} \wedge e(v, S) > 0\}.$$

The algorithm maintains a dynamic data structure that stores the current state S, its two-sided boundary $\delta(S)$, and the values of $p(v, S)$ for vertices in the two-sided boundary. This allows the algorithm to ignore the vertices in the interior of the set when selecting the next state. The complexity of `GenerateSample` is dominated by the work required to iterate over the boundary of the current set, select the next state using the stored values of $p(v, S)$, and update the set-with-boundary data structure.

The following proposition describes the set-with-boundary data structure that is used by `GenerateSample`.

PROPOSITION 5.5. *There is a set-with-boundary data structure S that supports these operations:*

—**add** *or* **remove** *a vertex v from S in time $O(d(v)\log \mathrm{vol}(S))$.*
—**get** *the value of $e(v, S)$, $\mathbb{I}[v \in S]$, or $p(v, S)$ in time $O(\log \mathrm{vol}(S))$.*
—**iterate** *over the vertices in the boundary of $\delta(S)$ in time $O(|\delta(S)|)$.*

PROOF. The set-with-boundary data structure can be implemented using two standard dictionary data structures. We maintain a *membership dictionary* $\mathcal{M}$ that contains the vertices in the set S and a *boundary dictionary* $\mathcal{B}$ that contains the vertices in $\delta(S)$ and stores the associated value $\mathcal{B}(v) = e(v, S)$ for each $v \in \delta(S)$. These dictionaries must support the following operations: inserting and deleting a key and its value, checking whether a given key is in the dictionary, and looking up the value associated with a key. A red/black tree supports these operations in the $O(\log N)$ worst-case time per operation, where N is the number of keys currently in the tree (see Cormen et al. [2009]).

The value of $\mathbb{I}[v \in S]$ can be computed by checking whether $v \in \mathcal{M}$. For any vertex $v \in V$, the value of $e(v, S)$ can be computed using two lookups, one into the membership dictionary and one into the boundary dictionary:

$$e(v, S) = \begin{cases} \mathcal{B}(v) & \text{if } v \in \mathcal{B},\\ d(v) & \text{if } v \notin \mathcal{B} \text{ and } v \in \mathcal{M},\\ 0 & \text{if } v \notin \mathcal{B} \text{ and } v \notin \mathcal{M}.\end{cases} \tag{24}$$

It is straightforward to compute $p(v, S)$ from $e(v, S)$ and $\mathbb{I}[v \in S]$.

Each time a node v is added or removed from S, we update the membership dictionary. For each neighbor $u \sim v$, we increment or decrement the value of $e(u, S)$ in the boundary dictionary. For each node u that was updated (including v and its neighbors), we determine whether the node is contained in $\delta(S)$ by examining the values of $e(u, S)$ and $\mathbb{I}[u \in S]$ and then add or remove u from the boundary dictionary when necessary. In total, adding or removing v takes $O(d(v))$ dictionary operations. The size of either dictionary is $O(\mathrm{vol}(S))$, so each dictionary operation takes time $O(\log \mathrm{vol}(S))$. □

We now describe `GenerateSample`. The input is a starting vertex v, a time limit $T \geq 0$, a budget $B \geq 0$, a target conductance Φ, and a target volume $\mathcal{K}$. Note that each of the Φ or $\mathcal{K}$ may set to infinity. The output is a set S_τ sampled from the volume-biased ESP with the stopping rule $\tau = \tau(T, B, \Phi, \mathcal{K})$. The algorithm simulates the volume-biased ESP using the coupling described in Theorem 2.4. It uses an instance S of the set-with-boundary data structure to maintain the current state S_t and also stores a vertex X that represents the current walk position X_t. Initially, $S = S_0 = \{v\}$ and $X = X_0 = v$. The algorithm proceeds in steps. At the beginning of step t, we have $S = S_{t-1}$ and $X = X_{t-1}$. The algorithm continues until the stopping time τ is reached, then outputs S_τ.

Each step has two stages. In the first stage we select S_t and compute a list of the vertices that need to be added or removed from S_{t-1} to form S_t. This stage requires $O(1) + O(\partial(S_{t-1}))$ operations. We stop after the first stage if $\mathrm{cost}(S_0, \ldots, S_t) > B$. Otherwise, we proceed to the second stage in which we update S to S_t, which requires $O(1) + O(\mathrm{vol}(S_t \Delta S_{t-1}))$ operations. We stop after the second stage if the stopping time $\tau(T, B, \Phi, \mathcal{K})$ is occurred. Each operation is either a constant time operation or a dictionary operation requiring time $O(\log n)$.

In stage 1, we begin with $X = X_{t-1}$ and then update the walk particle. Given that $X_{t-1} = v_{t-1}$, we choose $X_t = v_t$ with probability $p(v_{t-1}, v_t)$ and update $X = X_t$. We assume that a random neighbor of v_{t-1} can be selected in time $O(\log n)$. We compute $p(v_t, S)$ by a lookup into the set-with-boundary data structure and select a random threshold Z uniformly from the interval $[0, p(v_t, S)]$. At this point we define $S_t = \{u \mid p(u, S_{t-1}) \geq Z\}$, but we do not yet update S to reflect S_t. Instead, we create a list D of the vertices in the set difference $S_t \Delta S_{t-1}$. We populate the list by iterating over each node $u \in \delta(S)$, looking up the value of $p(u, S_{t-1})$ and comparing this value with the threshold Z. While doing this, we update the values of $\mathrm{vol}(S_t)$ and $\mathrm{cost}(S_0, \ldots, S_t)$. We the check whether either of the stopping conditions $t = T$ or $\mathrm{cost}(S_0, \ldots, S_t) > B$ is satisfied. If so, then we stop and output $S_t = S_{t-1} \Delta D$. Otherwise, we proceed to the next stage.

In stage 2, we update S to S_t by adding or removing the vertices from D. While making these updates to S, we also update $\partial(S_{t-1})$ to $\partial(S_t)$. We compute $\phi(S_t) = \partial(S_t)/\mathrm{vol}(S_t)$ and check whether $\phi(S_t) \leq \Phi$ and $\mathrm{vol}(S_t) \leq \mathcal{K}$. If so, then we halt and output the set S_t. Otherwise, we proceed to the next step.

Proof of Theorem 5.3. By construction, `GenerateSample` simulates the coupling from Section 2.5. By Theorem 2.4, the sequence $(S_0, \ldots, S_\tau)$ it generates is a sample path from the volume-biased ESP.

Let $c_t := \partial(S_{t-1}) + \mathrm{vol}(S_{t-1} \Delta S_t)$. We will show that the number of operations performed in step t is $O(c_t)$. Each operation is either a constant time operation or a dictionary operation requiring time $O(\log n)$. The number of operations performed in stage 1 is dominated by step 7 in 1, in which lookup operations are performed for each vertex in $O(\delta(S_{t-1}))$. This requires $O(1) + O(\partial(S_{t-1}))$ operations. The number of operations performed in stage 2 is dominated by step 13, in which the vertices from $S_{t-1} \Delta S_t$ are added or removed from the set-with-boundary data structure S. Using Theorem 5.5, this requires $O(1) + O(\mathrm{vol}(S_t \Delta S_{t-1}))$ operations. In total, the number of operations required in step t is

$$O(\partial(S_{t-1}) + \mathrm{vol}(S_{t-1} \Delta S_t)) + O(1) = O(c_t).$$

ALGORITHM 1: `GenerateSample`$(v, T, B, \Phi, \mathcal{K})$:

Input: A starting vertex v, and three integers $T, B, \mathcal{K} \geq 0$, and a target conductance $0 \leq \Phi \leq 1$.

Output: A set S_τ sampled from the volume-biased ESP with stopping rule $\tau = \tau(T, B, \Phi, \mathcal{K})$.

Internal State:

$S =$ an instance of the set-with-boundary data structure.

$X =$ the current location of the random-walk particle.

We also maintain the current values of $\partial(S)$, $\text{vol}(S)$, and $\text{cost}(S_0, \ldots, S)$.

Initialization: $S = S_0 = \{v\}$ and $X = v_0 = v$.

1: **for** $t = 1 \to T$ **do** ▷ Loop assumptions: $S = S_{t-1}$ and $X = X_{t-1}$.
2: **Stage 1** (Select the vertices to add or remove from S):
3: Given $X_{t-1} = v_{t-1}$, select v_t with probability $p(v_{t-1}, v_t)$ and update $X \leftarrow X_t = v_t$.
4: Lookup $p(v_t, S_{t-1})$ and pick Z uniformly at random from the interval $[0, p(v_t, S_{t-1})]$.
5: Define $S_t = \{u \mid p(u, S_{t-1}) \geq Z\}$, $D = \emptyset$.
6: **for all** $u \in \delta(S_{t-1})$ **do**
7: Lookup $p(u, S_{t-1})$. If $u \in S_{t-1} \Delta S_t$, then add u to D.
8: **end for**
9: Update $\text{vol}(S_t)$ and $\text{cost}(S_0, \ldots, S_t)$.
10: If $t = T$ or $\text{cost}(S_0, \ldots, S_t) > B$, then return $S_t = S_{t-1} \Delta D$.
11: **EndStage 1**
12: **Stage 2** (Update S):
13: Update S to $S_t = S_{t-1} \Delta D$ by adding or removing the vertices in D from S.
14: Update $\partial(S_t)$, $\phi(S_t) = \partial(S_t)/\text{vol}(S_t)$.
15: **EndStage 2**
16: If $\phi(S_t) \leq \Phi$, and $\text{vol}(S_t) \leq \mathcal{K}$, then return S_t.
17: **end for**

The $O(1)$ term above can be ignored safely because $\partial(S_{t-1}) > 0$ for $t \leq \tau$. We define $c_0 = d(v_0) = \text{vol}(S_0)$ to account for the work required to create S_0. Then, the total number of operations performed by `GenerateSample` is $O(c_0 +, \ldots, +c_\tau) = O(\text{cost}(S_0, \ldots, S_\tau))$.

If $\text{cost}(S_0, \ldots, S_\tau) > B$, then `GenerateSample` halts after stage 1 during step τ, and the number of operations performed in step τ is $O(\partial(S_{\tau-1}))$. The total number of operations performed is therefore

$$O(\text{cost}(S_0, \ldots, S_{\tau-1})) + \partial(S_{\tau-1}) = O(2B),$$

because $\partial(S_{\tau-1}) \leq \text{vol}(S_{\tau-1}) \leq \text{cost}(S_0, \ldots, S_{\tau-1}) \leq B$. The proposition follows. □

5.2. Proof of the Main Theorem

In this subsection we finally prove Theorem 1.1. To prove Theorem 1.1, we can simply run k^ϵ copies of the volume-biased ESP in parallel. By Theorem 4.1, with a constant probability at least one of the copies finds a non-expanding set. We bound the time complexity of the algorithm using Theorem 5.3 and Theorem 5.4. The details of the algorithm is described in 2.

Now we are ready to prove Theorem 1.1.

THEOREM 1.1 (MAIN). `EvoPar`(v, k, ϕ, ϵ) *takes as input a starting vertex* $v \in V$*, a target conductance* $\phi \in (0, 1)$*, a target volume* k*, and* $0 < \epsilon < 1$ *and outputs a set of vertices. For*

ALGORITHM 2: `EvoPar`(v, k, ϕ, ϵ)

$T \leftarrow \epsilon \log k/6\phi$.
Run $k^{\epsilon/2}$ independent copies of `GenerateSample`$(v, T, \infty, \Phi_{\epsilon/2}(\phi), \mathcal{K}_{\epsilon/2}(k))$, in parallel. As soon as any of the copies finds a set S, of volume $\text{vol}(S) \leq \mathcal{K}_{\epsilon/2}(k)$, and conductance $\phi(S) \leq \Phi_{\epsilon/2}(\phi)$, stop the algorithm and return S.

a given run of the algorithm let S be the output and let W be the computational complexity of the algorithm. Then, S and W depend on the randomness of the algorithm, and

(1) The work per volume ratio $W/\text{vol}(S)$ satisfies,

$$\mathbb{E}[W/\text{vol}(S)] = O(k^{\epsilon}\phi^{-1/2}\log^2 n).$$

(2) If $A \subseteq V$ is a set of vertices that satisfy $\phi(A) \leq \phi$, and $\text{vol}(A) \leq k$, then there is a subset $A' \subseteq A$ with volume at least $\text{vol}(A)/2$, such that if $v \in A'$, then with a constant probability S satisfies,
(a) $\phi(S) = O(\sqrt{\phi/\epsilon})$,
(b) $\text{vol}(S) = O(k^{1+\epsilon})$.

Proof. Let A' be as defined in Theorem 4.1. First, for any $v \in A'$, using Theorem 4.1, each copy finds a set of volume $\mathcal{K}_{\epsilon/2}(k)$ and conductance $\Phi_{\epsilon/2}(\phi)$ with probability $\Omega(k^{-\epsilon/2})$; in this case the algorithm will output a set satisfying theorem's statement. But, since $k^{\epsilon/2}$ copies are executed independently, at least one of them will succeed with a constant probability. This proves the correctness of the algorithm.

It remains to compute the time complexity. Let $\ell := k^{\epsilon/2}$ be the number of copies and $W_1, \ldots, W_\ell$ be random variables indicating the work done by each of the copies in a single run of `EvoPar`, thus $\sum_i W_i$ is the time complexity of the algorithm. Note that it is possible that $W_i < W_j$ for some $i \neq j$, since the ith copy may stop without finding any small non-expanding sets. Let S_{out} be the output of the algorithm. If the algorithm returns the output of the i-th copy, then we define $I_i = 1/\text{vol}(S_{out})$ and we let $I_i = 0$ otherwise. Also, let $I := \sum I_i$; note that if the algorithm returns the empty set, then $I = 0$. We write $\widehat{\mathbf{P}}_v^{\ell}[.]$ to denote the probability measure of the l-independent volume-biased ESP all started from $S_0 = \{v\}$, and $\widehat{\mathbf{E}}_v^{\ell}[.]$ for the expectation. To prove the theorem it is sufficient to show

$$\widehat{\mathbf{E}}_v^{\ell}\left[I\sum_{i=1}^{\ell} W_i\right] = O(k^{\epsilon}\phi^{-1/2}\log^2 n).$$

By linearity of expectation, it is sufficient to show that for all $1 \leq i \leq k$,

$$\widehat{\mathbf{E}}_v^{\ell}\left[I_i\sum_{j=1}^{\ell} W_j\right] = O(k^{\epsilon/2}\phi^{-1/2}\log^2 n),$$

By symmetry of the copies, it is sufficient to show the above equation only for $i = 1$. Furthermore, since conditioned on $I_1 \neq 0$, $W_1 = \max_i W_i$, we just need to show

$$\widehat{\mathbf{E}}_v^{\ell}[I_1 W_1] = O(\phi^{-1/2}\log^2 n),$$

Let τ be a $\tau(T, \infty, \Phi_{\epsilon/2}(\phi), \mathcal{K}_{\epsilon/2}(k))$ stopping time for the first copy, S_τ be the output, and $W_1(\tau)$ be the amount of work done by that time. Note that we always have $W_1 \leq W_1(\tau)$ because the first copy may be stopped since one of the other copies succeeded. Since $I_1 \leq 1/\text{vol}(S_\tau)$, for any element of the joint probability space we have $I_1 W_1 \leq W_1(\tau)/\text{vol}(S_\tau)$.

Therefore,

$$\widehat{\mathbf{E}}_v^{\ell}[I_1 W_1] \leq \widehat{\mathbf{E}}_v^{\ell}\left[\frac{W_1(\tau)}{\mathrm{vol}(S_\tau)}\right] = \widehat{\mathbf{E}}_v\left[\frac{W(\tau)}{\mathrm{vol}(S_\tau)}\right] = O(T^{1/2}\log^{3/2} n) = O(\phi^{-1/2}\log^2 n),$$

where the second-to-last equation follows from Theorem 5.4 and Theorem 5.3. □

6. FINDING BALANCED CUTS

Spielman and Teng [2013] studied local graph clustering algorithm for their goal of designing almost linear time approximation algorithms for the balanced separator problem. Although the spectral clustering algorithm finds a square-root approximation for the sparsest cut problem, the output set may be very small, so to obtain a balanced cut, one needs iterated applications of the algorithm that can result in quadratic running time.

Using a local graph clustering algorithm, we can guarantee that the size of the output is proportional to the running time of the algorithm. Therefore, by iterated applications of local graph clustering algorithms in a time proportional to the size of the graph we can a find an almost balanced cut. Say we are looking to find a nonexpanding set with volume $\Omega(m/\beta)$ where $\beta = \Theta(m^{\epsilon})$ encodes the balancedness of the cut. First, we design an algorithm, `EvoNibble` (analogous to the `Nibble` algorithm of Spielman and Teng [2013]), with a running time $m^{\epsilon}/\sqrt{\phi}$ up to polylogarithmic factors, that returns a set S where S is either an empty set or a set with small conductance and it contains $\Omega(\frac{1}{\beta^2 m})$ fraction of the target set in expectation. The `EvoNibble` simply runs the `GenerateSample` procedure from a uniformly random vertex of G with a random budget that is at most polylog(n) in expectation.

Later, in Section 6.2, we describe `EvoPartition` that is an adaptation of the `EvoNibble` algorithm of Spielman and Teng [2013]. The `EvoPartition` iteratively calls and extracts the output of `EvoNibble` from G. Since the output of `EvoNibble` contains $\Omega(\frac{1}{\beta^2 m})$ fraction of the target set, after $m\,\mathrm{poly}(\beta)$ iterations, the union of the output sets is a β-balanced cut (in expectation). When this event occurs, the `EvoPartition` stops and returns this cut. Therefore, `EvoPartition` runs in time $m\,\mathrm{poly}(\beta)/\sqrt{\phi}$ up to polylogarithmic factors and returns a β-balanced cut with a constant probability. Since for any two disjoint sets S_1, S_2, $\phi(S_1 \cup S_2) \leq \max\{\phi(S_1), \phi(S_2)\}$, it follows that the conductance of the output β-balanced cut is at most square root of the target cut, up to an error of $O(1/\sqrt{\epsilon})$.

6.1. `EvoNibble`

In this section we describe the subroutine `EvoNibble`. We use the parameter

$$\beta := \frac{\mathcal{K}_{\epsilon}(\mathrm{vol}(V))}{\mathrm{vol}(V)} = 2\mathrm{vol}(V)^{\epsilon}/c_0.$$

ALGORITHN 3: `EvoNibble`(ϕ, ϵ)

Let $T = \epsilon \log \mathrm{vol}(V)/3\phi$.
Choose a random vertex $X \in V$ with probability $\mathbb{P}[X = v] = d(v)/\mathrm{vol}(V)$.
Choose a random budget as follows. Let $J_{\max} = \log_2 \mathrm{vol}(V)$, and let J be an integer from $[0, J_{\max}]$ chosen with probability $\mathbb{P}[J = j] \propto 2^{-j}$. Let $B_J = 16\alpha\beta 2^J$, where $\alpha = 1 + 4\sqrt{T \log \mathrm{vol}(V)}$.
Compute $S =$ `GenerateSample`$(X, T, B_J, \Phi_{\epsilon}(\phi), \infty)$.
If $\phi(S) \leq \Phi_{\epsilon}(\phi)$ and $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$, then output S. Otherwise output $\emptyset$.

PROPOSITION 6.1. *The randomized algorithm* `EvoNibble`(ϕ, ϵ) *takes as input* $\phi, \epsilon \in (0, 1)$ *and outputs a set* $S \subseteq V$. *The following hold:*

(1) The expected complexity is $O(\mathrm{vol}(V)^{\epsilon}\phi^{-1/2}\log^3 \mathrm{vol}(V))$.
(2) Either $S = \emptyset$ *or* S *satisfies* $\phi(S) = \Phi_{\epsilon}(\phi)$ *and* $\mathrm{vol}(S) \le \mathrm{vol}(V)/2$.
(3) For any set $A \subseteq V$ *that satisfies* $\beta\,\mathrm{vol}(A) \le \mathrm{vol}(V)/2$ *and* $\phi(A) \le \phi$,

$$\mathbb{E}\left[\frac{\mathrm{vol}(S \cap A)}{\mathrm{vol}(A)}\right] \ge \frac{1}{32\beta^2 \mathrm{vol}(V)}.$$

PROOF. First we prove conclusion (1). Let W be the complexity of the algorithm. By Theorem 5.3 and Theorem 5.4, we have

$$\mathbb{E}[W \mid J] = O(B_J \log \mathrm{vol}(V)) = O(\alpha\beta 2^J \log \mathrm{vol}(V)),$$

The expected complexity is

$$\begin{aligned}
\mathbb{E}[W] &= \sum_{j \in [0, J_{\max}]} \mathbb{E}[W | J = j]\mathbb{P}[J = j] \\
&= \sum_{j \in [0, J_{\max}]} O(\alpha\beta 2^j \log \mathrm{vol}(V)) O(2^{-j}) \\
&= O(\alpha\beta J_{\max} \log \mathrm{vol}(V)) \\
&= O(\mathrm{vol}(V)^{\epsilon} \log^3(\mathrm{vol}(V))\phi^{-1/2}),
\end{aligned}$$

where in the last equation we used $\alpha = O(\log \mathrm{vol}(V)\sqrt{\epsilon/\phi})$.

Conclusion (2) is immediate from the definition of the algorithm. We now prove conclusion (3). Let S_{out} be the output of `EvoNibble`. Let X be the starting vertex. Let A be a set that satisfies the requirements of conclusion (3), and let $A' \subseteq A$ be the subset defined in Theorem 4.1. We will prove the following:

$$\text{if } v \in A', \text{ then } \mathbb{E}[\mathrm{vol}(S_{out} \cap A) \mid X = v] \ge 1/16\beta^2. \tag{25}$$

After that, conclusion (3) follows by taking the expectation over the choice of the starting vertex:

$$\begin{aligned}
\mathbb{E}[\mathrm{vol}(S_{out} \cap A)] &= \sum_{v \in V} \mathbb{E}[\mathrm{vol}(S_{out} \cap A) \mid X = v]\mathbb{P}[X = v] \\
&\ge \frac{1}{16\beta^2}\mathbb{P}[X \in A'] \ge \frac{\mathrm{vol}(A)}{32\beta^2 \mathrm{vol}(V)}.
\end{aligned}$$

We now prove (25). Consider a sample path from the volume-biased ESP started from $\{X\}$, and let $\tau = \tau(T, \infty, \Phi_{\epsilon}(\phi), \infty)$. Let $\mathcal{E}$ be the event that all of the following hold:

(i) $\mathrm{cost}(S_0, \ldots, S_{\tau}) \le 8\alpha\beta \cdot \mathrm{vol}(S_{\tau})$,
(ii) $\phi(S_{\tau}) \le \Phi_{\epsilon}(\phi)$,
(iii) $\mathrm{vol}(S_{\tau} \cap A) \ge \mathrm{vol}(S_{\tau})/\beta$.
(iv) $\mathrm{vol}(S_{\tau}) \le \mathrm{vol}(V)/2$,

Note that item (iv) is just a consequence of (iii) and the assumption that $\beta\,\mathrm{vol}(A) \le \mathrm{vol}(V)/2$. Using Theorem 4.1, if $v \in A'$, then (ii) and (iii) hold with probability $1/4\beta$; on the other hand, using Theorem 5.4 and Markov inequality, (i) holds with probability $1 - 1/8\beta$. Therefore, using the union bound, if $v \in A'$, then $\mathbb{P}[\mathcal{E} \mid X = v] \ge 1/8\beta$.

The subroutine `GenerateSample`$(X, T, B, \Phi_{\epsilon}(\phi), \infty)$ returns the set $S_{\tau(T,B,\Phi_{\epsilon}(\phi),\infty)}$ rather than S_{τ}. To deal with this, we define the events $\mathcal{E}_j = \mathcal{E} \wedge (\mathrm{vol}(S_{\tau}) \in [2^j, 2^{j+1}))$.

Note that

$$(\text{cost}(S_0, \ldots, S_\tau) \mid \mathcal{E}_j) \le 16\alpha\beta 2^j = B_j.$$

This implies that if the event $(\mathcal{E}_j \wedge (J \ge j))$ holds, then $S_{out} = S_{\tau(T,B,\Phi_\epsilon(\phi),\infty)} = S_\tau$, and, furthermore, $\text{vol}(S_{out} \cap A) \ge 2^j/\beta$. For any $v \in A'$, we have

$$\begin{aligned}
&\mathbb{E}[\text{vol}(S_{out} \cap A) \mid X = v] \\
&\quad \ge \sum_{j \in [0, J_{\max}]} \mathbb{E}[\text{vol}(S_{out} \cap A) \mid X = v, J = j, \mathcal{E}_j]\mathbb{P}[\mathcal{E}_j \mid X = v]\mathbb{P}[J = j] \\
&\quad \ge \sum_{j \in [0, J_{\max}]} \frac{2^j}{\beta}(2^{-j-1})\mathbb{P}[\mathcal{E}_j \mid X = v] \\
&\quad \ge \frac{1}{2\beta} \sum_{j \in [0, J_{\max}]} \mathbb{P}[\mathcal{E}_j \mid X = v] \\
&\quad \ge \frac{1}{2\beta}\mathbb{P}[\mathcal{E} \mid X = v] \ge \frac{1}{16\beta^2}.
\end{aligned}$$

This establishes (25) and completes the proof. □

6.2. `EvoPartition`

The algorithm `EvoPartition` described in the following theorem is constructed by substituting the subroutine `EvoNibble` for `Nibble` in Spielman and Teng's algorithm `Partition`. The proof is just an adaptation of Spielman and Teng [2013, Theorem 3.2].

ALGORITHM 4: `EvoPartition`(ϕ, ϵ)

1: $V_0 = V$, $S = \emptyset$, $\ell = 96\beta^2 m = \Theta(m^{1+2\epsilon})$.
2: **for** $j = 1 \to \ell$ **do**
3: Compute $P_j =$ `EvoNibble`$(2\phi, \epsilon)$ on $G[V_{j-1}]$.
4: $S = S \cup P_j$.
5: $V_j = V_{j-1} \backslash P_j$.
6: If $\text{vol}(V_j) < \text{vol}(V)/2$, return S.
7: **end for**
return S.

The following is the main theorem of this section.

THEOREM 6.2. *The randomized algorithm* `EvoPartition`(ϕ) *takes inputs* $\phi, \epsilon \in (0, 1)$*, and it outputs a set* $S \subseteq V$*. The expected complexity is* $O(m^{1+3\epsilon}\phi^{-1/2}\,\text{polylog}(n))$*, where* $m = \text{vol}(V)$*. S satisfies the following:*

(1) $\text{vol}(S) \le 3\text{vol}(V)/4$.
(2) If $S \ne \emptyset$*, then* $\phi(S) = O(\sqrt{\phi/\epsilon})$.
(3) If $A \subseteq V$ *is any set satisfying* $\phi(A) \le \phi$ *and* $\beta\,\text{vol}(A) \le \text{vol}(V)/4$*, where* $\beta = \Theta(m^\epsilon)$*, then with probability at least* $1/2$*, at least one of the following holds:*
(3.a) $\text{vol}(S) \ge \text{vol}(V)/2\beta$,
(3.b) $\text{vol}(S \cap A) \ge \text{vol}(A)/2$.

We remark that in the specific application of the balanced separator problem, we can only guarantee $\phi(S) = O(\sqrt{\phi \log n})$ in the worst case. In particular, if every set of volume at most $\text{vol}(V)/4$ is highly expanding, then we can only invoke the above theorem for $\epsilon = \Theta(1/\log n)$; this implies a $\sqrt{\log n}$ dependency on n in the conductance of the output set.

PROOF. It is easy to see that the expected complexity of the algorithm is $O(m^{1+3\epsilon}\phi^{-1/2}$ polylog$(n))$. This is because we execute the loop for $O(m^{1+2\epsilon})$ iterations, and, using Theorem 6.1, the complexity of each iteration is $O(m^{\epsilon}\phi^{-1/2}\,\text{polylog}(n))$.

Let S_{out} be the output of the algorithm. Conclusion (1) is easy to verify. For any $1 \le j \le m^{1+2\epsilon}$, $\text{vol}(P_j) \le \text{vol}(V_j)/2 \le \text{vol}(V)/4$. Therefore, $\text{vol}(S_{out}) \le 3\text{vol}(V)/4$. Conclusion (2) simply follows from the fact that the conductance of the union of two disjoint sets is less than the maximum of their conductances. That is,

$$\phi(S_{out}) \le \max_{1\le j\le \ell} \phi(P_j) \le \Phi_\epsilon(2\phi) = O\big(\sqrt{\phi/\epsilon}\big).$$

It remains to prove (3). For $1 \le i \le \ell$, define the random variable

$$X_i = \frac{\text{vol}(P_i \cap A)}{\text{vol}(A)}.$$

If the algorithm stops before the i-th iteration of the loop, then we let $X_i = 0$. Let $\mathcal{E}_j$ be the event that $\sum_{i=1}^{j} X_i \ge 1/2$. Since $P_1, P_2, \ldots$ are disjoint, if $\mathcal{E}_j$ occurs for some $j \le \ell$, then conclusion (3.b) is satisfied and we are done. Let $\mathcal{F}_j$ be the event that $\text{vol}(V_i) < \text{vol}(V)/2$ for some $i \le j$. On the other hand, if some $\mathcal{F}_j$ occurs, then conclusion (3.a) is satisfied. We show that, with a constant probability, either of $\mathcal{E}_j$ or $\mathcal{F}_j$ occurs for some $j \le \ell$. □

CLAIM 6.3. *If for some $j < \ell$ neither $\mathcal{E}_j$ nor $\mathcal{F}_j$ hold, then $\mathbb{E}[X_{j+1}] \ge \frac{1}{64\beta^2 m}$.*

PROOF. We show that the conditions of part (3) of Theorem 6.1 are satisfied for the graph $G[V_j]$ and the set $A \cap V_j$; the claim follows from the conclusion of part (3). First, we upper bound $\phi(A \cap V_j)$, and then we upper bound $\beta\ \text{vol}(A \cap V_j)$. Since $\mathcal{E}_j$ does not hold,

$$\text{vol}(V_j \cap A) = \text{vol}(A) - \sum_{i=1}^{j} \text{vol}(P_i \cap A) = \text{vol}(A)\left(1 - \sum_{i=1}^{j} X_i\right) \ge \text{vol}(A)/2.$$

For a set $S \subset V_j$, let $\phi_{G[V_j]}(S)$ be the conductance of S in the induced graph $G[V_j]$. Therefore,

$$\phi_{G[V_j]}(A \cap V_j) = \frac{\partial_{G[V_j]}(A \cap V_j)}{\text{vol}_{G[V_j]}(A \cap V_j)} \le \frac{|E(A \cap V_j), V_j \backslash A)|}{\text{vol}(A)/2} \le \frac{\partial(A)}{\text{vol}(A)/2} \le 2\phi(A) \le 2\phi.$$

On the other hand, since $\mathcal{F}_j$ does not hold, $\text{vol}(V_j) \ge \text{vol}(V)/2$. Therefore,

$$\beta\ \text{vol}(A \cap V_j) \le \beta\ \text{vol}(A) \le \text{vol}(V)/4 \le \text{vol}(V_j)/2.$$

Therefore, using part (3) of Theorem 6.1,

$$\mathbb{E}[X_{j+1}] \ge \frac{1}{32\beta^2\ \text{vol}(V_j)} \ge \frac{1}{64\beta^2 m}. \quad \square$$

Now, let

$$Y_j = \begin{cases} \frac{1}{64\beta^2 m} & \text{if } \mathcal{E}_j \text{or } \mathcal{F}_j \\ X_j & \text{otherwise.} \end{cases}$$

Then, using the above claim,

$$\mathbb{E}\left[\sum_{i=1}^{\ell} Y_i\right] \ge \frac{\ell}{64\beta^2 m} = 3/2.$$

On the other hand,

$$\sum_{i=1}^{\ell} Y_i \leq \sum_{i=1}^{\ell} X_i + \frac{\ell}{64\beta^2 m} \leq 5/2.$$

Therefore, with probability at least 1/2, $\sum_{j=1}^{\ell} Y_j \geq 1/2$. Therefore, at least one of $\mathcal{E}_\ell$ or $\mathcal{F}_\ell$ occur.

7. CONCLUDING REMARKS

It remains a fascinating open problem to prove or refute the SSE conjecture of Raghavendra and Steurer [2010]. As we mentioned in the introduction, our Theorem 1.5 indicates that the hard instances of the small set expansion problem are those where $\phi(m^{1-\Omega(1)}) \approx 1$, while $\phi(\delta m) \leq \phi$.

One promising approach to refute this conjecture is to use the powerful Theorem 4.2. We conjectured that if G has a set A of conductance ϕ, then there is a vertex $v \in A$ such that, with probability $1/\mathrm{vol}(A)^{O(1)}$, all of the sets in a sample path of the volume-biased ESP, started from $\{v\}$ and ran for $T = \Omega(\log(\mathrm{vol}(A))/\phi)$ steps, have size at most $O(\mathrm{vol}(A))$. This conjecture was recently refuted by Chan et al. [2015] for an extension of the "noisy hypercube graph." Nonetheless, the conjecture may still hold for graphs with a moderate threshold rank, that is, G, where $\mathrm{rank}_{1-\eta}(D^{-1}A)$ is a polylogarithmic function of n. It is easy to see that an affirmative answer of the conjecture combined with Theorem 4.2 provides a quasi polynomial time algorithm for the small set expansion problem.

ACKNOWLEDGMENTS

We thank Or Meir and Amin Saberi for stimulating discussions. We also thank Noga Alon for pointing us to the reference London [1966].

REFERENCES

N Alon. 1986. Eigen values and expanders. *Combinatorica* 6 (January 1986), 83–96. Issue 2. http://portal.acm.org/citation.cfm?id=18497.18498.

N. Alon and V. Milman. 1985. Isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theor. Ser. B* 38, 1 (Feb 1985), 73–88.

Reid Andersen, Fan R. K. Chung, and Kevin J. Lang. 2006. Local graph partitioning using pagerank vectors. In *FOCS*. 475–486.

Reid Andersen and Kevin J. Lang. 2006. Communities from seed sets. In *WWW*. ACM, New York, NY, 223–232.

Reid Andersen and Yuval Peres. 2009. Finding sparse cuts locally using evolving sets. In *STOC*. 235–244.

Sanjeev Arora, Boaz Barak, and David Steurer. 2010a. Subexponential algorithms for unique games and related problems. In *FOCS*. IEEE Computer Society, Washington, DC, 563–572.

Sanjeev Arora, Elad Hazan, and Satyen Kale. 2010b. $O(\sqrt{\log n})$ Approximation to sparsest cut in $\tilde{O}(n^2)$ time. *SIAM J. Comput.* 39, 5 (2010), 1748–1771.

Nikhil Bansal, Uriel Feige, Robert Krauthgamer, Konstantin Makarychev, Viswanath Nagarajan, Joseph Naor, and Roy Schwartz. 2011. Min-max graph partitioning and small set expansion. In *FOCS*. IEEE, Washington, DC, 17–26.

G. R. Blakley and Prabir Roy. 1965. A hölder type inequality for symmetric matrices with nonnegative entries. In *Proc. Am. Math. Soc.* Vol. 16. 1244–1245.

Siu On Chan, Tsz Chiu Kwok, and Lap Chi Lau. 2015. Random Walks and Evolving Sets: Faster Convergences and Limitations. (2015). http://arxiv.org/abs/1507.02069.

T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. 2009. *Introduction to Algorithms* (3rd ed.). MIT Press, Cambridge, MA.

Persi Diaconis and James A. Fill. 1990. Strong stationary times via a new form of duality. *Ann. Probab.* 18, 4 (1990), 1483–1522.

Gary William Flake, Steve Lawrence, and C. Lee Giles. 2000. Efficient identification of Web communities. In *SIGKDD*. ACM, New York, NY, 150–160.

M. Jerrum and Alistair Sinclair. 1989. Approximating the permanent. *SIAM J. Comput.* 18, 6 (1989), 1149–1178.

Ravi Kannan, Santosh Vempala, and Adrian Vetta. 2004. On clusterings: Good, bad and spectral. *J. ACM* 51 (2004), 497–515.

Jon M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *J. ACM* 46 (1999), 668–677.

Tsz Chiu Kwok and Lap Chi Lau. 2012. Finding small sparse cuts by random walk. In *APPROX-RANDOM*. 615–626.

Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. 2013. Improved cheeger's inequality: Analysis of spectral partitioning algorithms through higher order spectral gap. In *STOC*. 11–20.

Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. 2008. Statistical properties of community structure in large social and information networks. In *WWW*. ACM, New York, NY, 695–704.

Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. 2009. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Math.* 6, 1 (2009), 29–123.

Jure Leskovec, Kevin J. Lang, and Michael Mahoney. 2010. Empirical comparison of algorithms for network community detection. In *WWW*. New York, NY, 631–640.

David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. 2006. *Markov chains and mixing times*. American Mathematical Society.

Angsheng Li, Yicheng Pan, and Pan Peng. 2011. Testing conductance in general graphs. *Electron. Colloq. Comput. Complex. (ECCC)* 18 (2011), 101.

Angsheng Li and Pan Peng. 2011. Community structures in classical network models. *Internet Math.* 7, 2 (2011), 81–106.

David London. 1966. Inequalities in quadratic forms. *Duke Math. J.* 33, 3 (1966), 511–522.

László Lovász and Ravi Kannan. 1999. Faster mixing via average conductance. In *STOC*. ACM, 282–287.

Ravi Montenegro. 2007. Sharp edge, vertex, and mixed cheeger inequalities for finite markov kernels. *Electronic Communications in Probability [electronic only]* 12 (2007), 377–389.

Ravi Montenegro. 2009. The simple random walk and max-degree walk on a directed graph. *Rand. Struct. Algorithm.* 34, 3 (May 2009), 395–407.

Ben Morris and Yuval Peres. 2003. Evolving sets and mixin. In *STOC*. 279–286.

H. P. Mulholland and C. A. B. Smith. 1959. An inequality arising in genetical theory. *Am. Math. Month.* 66 (1959), 673–683.

Ryan O'Donnell and David Witmer. 2012. Improved small-set expansion from higher eigenvalues. *CoRR* abs/1204.4688 (2012).

Lorenzo Orecchia, Sushant Sachdeva, and Nisheeth K. Vishnoi. 2012. Approximating the exponential, the lanczos method and an $\tilde{O}(m)$-time spectral algorithm for balanced separator. In *STOC*.

Lorenzo Orecchia and Nisheeth K. Vishnoi. 2011. Towards an sdp-based approach to spectral methods: A nearly-linear-time algorithm for graph partitioning and decomposition. In *SODA*. 532–545.

Shayan Oveis Gharan and Luca Trevisan. 2012. Approximating the expansion profile and almost optimal local graph clustering. In *FOCS*. 187–196.

Prasad Raghavendra and David Steurer. 2010. Graph expansion and the unique games conjecture. In *STOC*. ACM, New York, NY, 755–764.

Prasad Raghavendra, David Steurer, and Prasad Tetali. 2010. Approximations for the isoperimetric and spectral profile of graphs and related parameters. In *STOC*. ACM, New York, NY, 631–640.

Jonah Sherman. 2009. Breaking the multicommodity flow barrier for $O(\sqrt{\log n})$-approximations to sparsest cut. In *FOCS*. 363–372.

Jianbo Shi and Jitendra Malik. 2000. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 8 (2000), 888–905.

Daniel A. Spielman and Shang-Hua Teng. 2004. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *STOC*. 81–90.

Daniel A. Spielman and Shang-Hua Teng. 2013. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM J. Comput.* 42, 1 (2013), 1–26.

David A. Tolliver and Gary L. Miller. 2006. Graph partitioning by spectral rounding: Applications in image segmentation and clustering. In *CVPR*. IEEE Computer Society, Washington, DC, 1053–1060.

David Williams. 1991. *Probability with Martingales (Cambridge Mathematical Textbooks)*. Cambridge University Press, Cambridge.

Received February 2014; revised September 2015; accepted December 2015