Approximating the Expansion Profile and Almost Optimal Local Graph Clustering

Shayan Oveis Gharan[+]
Management Science and Engineering Dept.
Stanford University
Stanford, CA, USA
shayan@stanford.edu

Luca Trevisan[*]
Department of Computer Science
Stanford University
Stanford, CA, USA
trevisan@stanford.edu

Abstract—Spectral partitioning is a simple, nearly-linear time, algorithm to find sparse cuts, and the Cheeger inequalities provide a worst-case guarantee of the quality of the approximation found by the algorithm. Local graph partitioning algorithms [1], [2], [3] run in time that is nearly linear in the size of the output set, and their approximation guarantee is worse than the guarantee provided by the Cheeger inequalities by a poly-logarithmic $\log^{\Omega(1)} n$ factor. It has been an open problem to design a local graph clustering algorithm with an approximation guarantee close to the guarantee of the Cheeger inequalities and with a running time nearly linear in the size of the output.

In this paper we solve this problem; we design an algorithm with the same guarantee (up to a constant factor) as the Cheeger inequality, that runs in time slightly super linear 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. As a byproduct of our results, we prove a bicriteria approximation algorithm for the expansion profile of any graph. Let $\mu(S) = \sum_{v \in S} d(v)$ be the volume, and $\phi(S) := |E(S, \overline{S})|/\mu(S)$, be the conductance of a set S of vertices. If there is a set of volume at most γ and conductance ϕ, we can find a set of volume at most $\gamma^{1+\epsilon}$ and conductance at most $O(\sqrt{\phi/\epsilon})$, for any $\epsilon > 0$. Our proof techniques also provide a simpler proof of the structural result of Arora, Barak, Steurer [4], that can be applied to irregular graphs.

Our main technical tool is a lemma stating 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 $(1 - \phi(S)/2)^t$. The lemma also implies a new lower bound to the uniform mixing time of any finite states reversible markov chain.

Keywords-Local Graph Clustering; Cheeger Inequality; Small Set Expansion Problem;

I. INTRODUCTION

Let $G = (V, E)$ be an undirected graph, with $n := |V|$ vertices, 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 degrees of vertices in S,

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

The *conductance* of a set S is defined as

$$\phi(S) := \partial(S)/\mu(S)$$

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

$$\phi(G) := \min_{S: \mu(S) \le \mu(V)/2} \phi(S)$$

be the conductance of G. The conductance is (approximately) the same as uniform sparsest cut only in regular graphs, otherwise the denominator is different. The Cheeger inequalities [5], [6] 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 [7], [8], [9].

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 iteration might be used to find a *balanced* sparse cut if one exists (c.f. [10]), 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 [11]. 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 [12], and subsequently [2], [3] studied local graph partitioning algorithms that find a set S of approximately minimal conductance in time nearly linear in the size of the output set S. Note that the running time can be sub linear in the size of the input graph if the algorithm finds a small output set S. When iterated, such an algorithm

187

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.

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 [13] and the expansion profile of a graph (that is, what is the cut of 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 contain more interesting information.

A. 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 paper is an almost optimal local graph clustering algorithm.

The problem first studied in the remarkable work of Spielman and Teng [12]. Spielman and Teng design an algorithm Nibble such that for any set $U \subseteq V$, if the initial vertex, v, is sampled randomly according to the degree of vertices in U, with a constant probability, Nibble finds a set of conductance $O(\phi^{1/2}(U) \log^{3/2} n)$, with a work/volume ratio of $O(\phi^{-2}(U) \operatorname{polylog}(n))$, 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, Chung and Lang [2], used the approximate PageRank vector rather than approximate random walk distribution, and they managed to improve the conductance of the output set to $O(\sqrt{\phi(U) \log n})$, and the work/volume ratio to $O(\phi^{-1}(U) \operatorname{polylog} n)$. More recently, Andersen and Peres [3], use the *evolving set process* developed in the work of Diaconis and Fill [14], and they improved the work/volume ratio to $O(\phi^{-1/2}(U) \operatorname{polylog} n)$, while achieving the same guarantee as [2] on the conductance of the

output set.

It has been an open problem to design a local variant of the Cheeger's inequalities: that is to provide a local graph clustering algorithm with an approximation guarantee that does not depend on the size of G. In this work we come close to resolving this open problem, and we prove the following theorem:

Theorem 1. ParESP$(v, \gamma, \phi, \epsilon)$ *takes as input a starting vetex $v \in V$, a target conductance $\phi \in (0, 1)$, a target size γ, and $0 < \epsilon < 1$. For a given run of the algorithm it outputs a set S of vertices with the expected work per volume ratio of $O(\gamma^\epsilon \phi^{-1/2} \log^2 n)$. If $U \subseteq V$ is a set of vertices that satisfy $\phi(U) \leq \phi$, and $\mu(U) \leq \gamma$, then there is a subset $U' \subseteq U$ with volume at least $\mu(U)/2$, such that if $v \in U'$, with a constant probability S satisfies,*

1) *$\phi(S) = O(\sqrt{\phi/\epsilon})$,*
2) *$\mu(S) \leq O(\gamma^{1+\epsilon})$.*

We remark that unlike the previous local graph clustering algorithms, the running time of the algorithm is slightly super linear in the size of the optimum. Nonetheless, by choosing $\epsilon = \Theta(1/\log \gamma)$, one can reproduce the previous local clustering theorems.

B. Approximating the Expansion Profile

As a byproduct of the above result we give an approximation algorithm for the expansion profile of G. Lovasz and Kannan [15] defined the *expansion profile* of a graph G as follows:

$$\phi(\gamma) := \min_{S:\mu(S)\leq\gamma} \phi(S).$$

Lovasz 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 [13].

The small set expansion problem is defined as follows:

Problem 2 (GAP Small Set Expansion $SSE(\delta, \phi)$ [13])**.** *Given a graph G, and constants $\delta, \phi > 0$, distinguish whether,*

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

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 $\operatorname{rank}_{1-\eta}(D^{-1}A)$, be the number (with multiplicities) of eigenvalues λ of $D^{-1}A$, satisfying $\lambda > 1 - \eta$. Arora, Barak and Steurer [4] solved the small set expansion problem on regular graphs G, assuming that $\operatorname{rank}_{1-\eta}(D^{-1}A) \geq n^{\operatorname{poly}(\eta/\phi)}$, for some constant $\eta > 0$. Unfortunately, their algorithm does not provide any approximation of $\phi(\delta n)$ for general graphs, since they may have much fewer large eigenvalues.

Raghavendra, Steurer, Tetali [16], and Bansal et al. [17] used the semidefinite programming relaxation of the problem, and designed algorithms that approximate $\phi(\gamma)$ within $O(\sqrt{\phi(\gamma)^{-1}\log\frac{\mu(V)}{\gamma}})$, and $O(\sqrt{\log n \log\frac{\mu(V)}{\gamma}})$ of the optimum, respectively. However, in the interesting regime of $\gamma = o(\mu(V))$, which is of interests to the small set expansion problem, the quality of both approximation algorithms is not independent of γ.

Here, we prove γ independent approximation of $\phi(\gamma)$ as a function of $\phi(\gamma^{1-\epsilon})$, without any dependency in the size of the graph; specifically, we prove the following theorem:

Theorem 3. *There is a polynomial time algorithm that takes as input a target conductance ϕ, and $0 < \epsilon < 1/4$, and outputs a set S, s.t. if $\phi(U) \leq \phi$, for some $U \subseteq V$, then $\mu(S) \leq 2\mu(U)^{1+\epsilon}$, and $\phi(S) \leq \sqrt{2\phi/\epsilon}$.*

Our theorem indicates that the hard instance of the small set expansion problem are those where $\phi(n^{1-\Omega(1)}) \approx 1$, while $\phi(\delta n) \leq \phi$. In other words, if in an instance of SSE problem $\phi(n^{1-\Omega(1)})$ is bounded away from 1, then using our algorithm we can prove $\phi(\delta n) \leq \phi$ for any constant $\delta > 0$.

We remark that one can also use `ParESP` to approximate $\phi(\gamma)$ with slightly worse guarantees (up to constant factors), in sub linear time. Here, for the sake of clarity and simplicity of the arguments, we prove the theorem using random walks. Independent of our work, Kwok and Lau [18] have obtained a somewhat different proof of Theorem 3.

Our analysis techniques also provide a simpler proof of the structural result of Arora, Barak, Steurer [4], that can be applied to non-regular graphs.

Theorem 4. *For any graph G, and $0 < \epsilon \leq 1$, if $\mathrm{rank}_{1-\eta}(D^{-1}A) \geq n^{(1+\epsilon)\eta/\phi}$, then there exists a set $S \subseteq V$ of volume $\mu(S) \leq 4\mu(V)n^{-\eta/\phi}$ and $\phi(S) \leq \sqrt{2\phi/\epsilon}$. Such a set can be found by finding the smallest volume threshold set of conductance at most $\sqrt{2\phi/\epsilon}$, among the rows of $(D^{-1}A)^t$, for $t = O(\log n/\phi)$.*

We remark that Arora et al. [4] prove a variant of the above theorem for regular graphs with the stronger assumption that $\mathrm{rank}_{1-\eta}(D^{-1}A) \geq n^{100\eta/\phi}$. Our theorem resolves their question of whether the factor 100 can be improved to $1 + \epsilon$. Independent of our work, O'Donnell and Witmer [19] obtained a different proof of the above theorem.

C. Lower Bounds on uniform mixing time of Random Walks

Using our techniques, we also 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, we model the markov chain as a random walk on a weighted graph G.

For the vertices $u, v \in V$, let $\mathcal{P}_u[X_t = v]$ be the probability that a random walk started at u, goes to v in t steps, and let $\pi(.)$ be the stationary distribution of the walk. The

ϵ-*uniform* mixing time of the chain is defined as

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

We prove the following lower-bound on $\tau(\epsilon)$.

Proposition 5. *For any (weighted) graph $G = (V, E)$, $1 \leq \gamma \leq \mu(V)/2$, and $0 < \epsilon < 1$, if $\phi(\gamma) \leq 0.7$, then*

$$\tau(\epsilon) \geq \frac{\ln(\mu(V)/2\gamma)}{2\phi(\gamma)} - 2.$$

Due to the space limits, we defer the proof of Theorem 4, and Proposition 5 to the full version of this paper.

D. Techniques

Our main technical result is that if S is a set of vertices, and we consider a t-step *lazy* random walk started at a random element of S, then the probability that the walk is entirely contained in S is at least $(1 - \phi(S)/2)^t$. Previously, only the lower bound $1 - t\phi(S)/2$ was known, and the analysis of other local clustering algorithms implicitly or explicitly depended on such a bound.

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 $t = (\epsilon \ln n)/\phi$ 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 expansion bound, while the $1/n^\epsilon$ factor that we lose in the probability corresponds to the factor that we lose in the size of the non-expanding set. We also use this bound to prove stronger lower bounds on the uniform mixing time of reversible markov chains.

Our polynomial time algorithm to approximate the expansion profile of a graph is the same as the algorithm used by Arora, Barak and Steurer [4] to find small non-expanding sets in graphs of a given threshold rank, but our analysis is different. (Our analysis can be also be used to give a different proof of their result.) Arora, Barak and Steurer use the threshold rank to argue that a random walk started from a random vertex of G will be at the initial vertex after t steps with probability at least $\mathrm{rank}_{1-\eta}(G)(1-\eta)^t/n$. Then, they argue that if all sets of a certain size have large conductance, this probability must be small, which is a contradiction. To make this quantitative they use the second norm of the probability distribution vector ($\|\mathbf{p}_t\|$) as a potential function, where $\mathbf{p}_t$ is the distribution of the walk after t steps, and they choose t to get a sufficiently small potential function and argue that the probability of being at the initial vertex after t steps must be small. In our analysis, we use the fact that for any set S, there is a vertex $v \in S$ such that the probability that the walk started at v remains in S is at least $(1 - \phi(S)/2)^t$. Then, we use the

potential function $I(\mathbf{p}_t, \gamma)$ introduced in the work of Lovasz and Simonovits [20]. Roughly speaking, $I(\mathbf{p}_t, \gamma)$ is defined as follows: consider the distribution $\mathbf{p}_t$ of the vertex reached in a t-step random walk started from a random element of S, take the k vertices of highest probability under $\mathbf{p}_t$, where k is chosen so that their total volume is about γ, then $I(\mathbf{p}_t, \gamma)$ is the total probability under $\mathbf{p}_t$ of those k vertices.

Using the machinery of Lovasz and Simonovits, we can upper-bound $I(\mathbf{p}_t, \gamma)$ by $\frac{\gamma}{\Gamma} + \sqrt{\gamma}(1 - \phi^2/2)^t$ conditioned on all of the threshold sets of volume at most Γ of the probability distribution vectors up to time t having conductance more than ϕ. Letting $t = \Omega(\epsilon \log \mu(S)/\phi(S))$, $\Gamma = O(\mu(S)^{1+\epsilon})$, and $\phi = O(\sqrt{\phi(S)/\epsilon})$ since the walk remains in S with probability at least $\mu(S)^{-\epsilon} < I(\mathbf{p}_t, \gamma)$, at least one of the threshold sets of volume at most $O(\mu(S)^{1+\epsilon})$, must have conductance $O(\sqrt{\phi(S)/\epsilon})$.

Our local algorithm uses the evolving set process. The evolving set process starts with a vertex v of the graph, and then produces a sequence of sets $S_1, S_2, \ldots, S_\tau$, with the property that at least one set S_t is such that $\partial S_t/\mu(S_t) \leq O(\sqrt{\log \mu(S_\tau)/\tau})$. If one can show that up to some time T the process constructs sets all of volume at most γ, then we get a set of volume at most γ and conductance at most $O(\sqrt{\log \gamma/T})$. Andersen and Peres were able to show that if the graph has a set S of conductance ϕ, then the process is likely to construct sets all of volume at most $2\mu(S)$ for at least $T = \Omega(1/\phi)$ steps, if started from a random element of the S, leading to their $O(\sqrt{\phi \log n})$ guarantee. We show that for any chosen $\epsilon < 1/2$, the process will construct sets of volume at most $O(\mu(S)^{1+\epsilon})$ for $T = \Omega(\epsilon \log \mu(S)/\phi)$ steps, with probability at least $1/\mu(S)^\epsilon$. This is enough to guarantee that, at least with probability $1/\mu(S)^\epsilon$, the process constructs at least one set of conductance $O(\sqrt{\phi/\epsilon})$. To obtain this conclusion, we also need to strengthen the first part of the analysis of Andersen and Peres: they show that the process has at least a constant probability of constructing a set of low conductance in the first t steps, while we need to show that this happens with probability at least $1 - 1/\mu(S)^{\Omega(1)}$, because we need to take a union bound with the event that t is large, for which probability we only have a $\mu(S)^{-\Omega(1)}$ lower bound. Finally, to achieve a constant probability of success, we run $\mu(S)^\epsilon$ copies of the evolving set process simultaneously, and stop as soon as one of the copies finds a small non-expanding set.

II. PRELIMINARIES

A. Notations

Let $G = (V, E)$ be an undirected graph, with $n := |V|$ vertices and $m := |E|$ edges. Let A be the adjacency matrix of G, D be the diagonal matrix of vertex degrees, and $d(v)$ 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,

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

Let $E(S, V \setminus S) := \{\{u, v\} : u \in S, v \notin S\}$ be the set of the edges connecting S to $V \setminus S$, 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

$$\phi(S) := \partial(S)/\mu(S).$$

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

We define the following probability distribution vector on a set $S \subseteq V$ of vertices:

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

In particular, we use $\pi(v) \equiv \pi_V(v)$ as the stationary distribution of a random walk in G. We use $v \sim \boldsymbol{\pi}_S$ to denote the distribution where v is sampled with probability $\pi_S(v)$.

Throughout the paper, let I be the identity matrix, and for any subset $S \subseteq V$, let I_S be the diagonal matrix such that $I_S(v, v) = 1$, if $v \in S$ and 0 otherwise. Also, let $\mathbf{1}$ be all one vector, and $\mathbf{1}_S$ be the indicator vector of the set S. We may abuse the notation and use $\mathbf{1}_v$ instead of $\mathbf{1}_{\{v\}}$ for a vertex $v \in V$.

We use lower bold letters to denote the vectors, and capital letters for the matrices/sets. For a vector $\mathbf{x} : V \to R$, and a set $S \subseteq V$, we use $x(S) := \sum_{v \in S} x(v)$. Unless otherwise specified, $\mathbf{x}$ is considered to be a column vector, and $\mathbf{x}'$ is its transpose. We use $\sqrt{\mathbf{x}}$ to denote the vector where each entry is square root of the corresponding entry of $\mathbf{x}$.

For a square matrix A, we use $\lambda_{\min}(A)$ to denote the minimum eigenvalue of A, and $\lambda_{\max}(A)$ to denote the maximum eigenvalue of A.

B. 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. We use $\mathcal{G} := (D^{-1}A + I)/2$ as the transition probability matrix of this random walk, and $\pi(.)$ is the unique stationary distribution, that is $\boldsymbol{\pi}'\mathcal{G} = \boldsymbol{\pi}'$. We write $\mathcal{P}_v[.]$ to denote the probability measure of the lazy random walk started from a vertex $v \in V$.

Let X_t be the random variable indicating the tth step of the random walk started at v. Observe that the distribution of X_t is exactly, $\mathbf{1}_v'\mathcal{G}^t$. For a subset $S \subseteq V$, and $v \in V$, and

integer $t > 0$, we write $\mathrm{esc}(v,t,S) := \mathcal{P}_v\left[\cup_{i=0}^t X_i \notin S\right]$ 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. It follows that,

$$\mathrm{rem}(v,t,S) := \mathbf{1}'_v(I_S\mathcal{G}I_S)^t\mathbf{1}_S. \tag{1}$$

C. Spectral Properties of the Transition Probability Matrix

Although the transition probability matrix, $\mathcal{G}$, is not a symmetric matrix, it features many properties of the symmetric matrices. First of all, $\mathcal{G}$ can be transformed to a symmetric matrix simply by considering $D^{1/2}\mathcal{G}D^{-1/2}$. It follows that any eigenvector of $D^{1/2}\mathcal{G}D^{-1/2}$ can be transformed into a left (right) eigenvector of $\mathcal{G}$, once it is multiplied by $D^{1/2}$ ($D^{-1/2}$), respectively. Henceforth, the left and right eigenvalues of $\mathcal{G}$ are the same, and they are real.

Furthermore, since $\|D^{-1}A\|_\infty \leq 1$, and $\mathcal{G}$ is the average of $D^{-1}A$, and the identity matrix, we must have $\lambda_{\min}(\mathcal{G}) \geq 0$ and $\lambda_{\max}(\mathcal{G}) \leq 1$. Thus $D^{1/2}\mathcal{G}D^{-1/2}$ is a positive semidefinite matrix, symmetric matrix, whose largest eigenvalue is at most 1.

D. The Evolving Set Process

The *evolving set process* is a markov chain on the subsets of the vertex set V. The process together with the closely related *volume biased evolving set process* is introduced in the work of Diaconis and Fill [14] as the *strong stationary dual* of a random walk. Morris and Peres [21] use it to upper-bound the mixing time of random walks in terms of isoperimetric properties.

Given a subset $S_0 \subseteq V$, the next subset S_1 is chosen as follows: first we choose a threshold $R \in [0,1]$ uniformly at random. Then, we let

$$S_1 := \{u : \mathcal{P}_u\left[X_1 \in S_0\right] \geq R\}.$$

The transition kernel of the evolving set process is defined as, $\mathrm{K}(S,S') = \mathbb{P}\left[S_1 = S'|S_0 = S\right]$. It follows that $\emptyset$, and V are the absorbing states of this markov chain, and the rest of the states are transient. Morris and Peres [21] defined the *growth gauge* $\psi(S)$ of a set S as follows:

$$\psi(S) := 1 - \mathbb{E}\left[\sqrt{\frac{\mu(S_1)}{\mu(S)}} \;\middle|\; S_0 = S\right].$$

They showed that

Proposition 6 (Morris,Peres [21])**.** *For any set $S \subseteq V$, $\psi(S) \geq \phi(S)^2/8$.*

The *volume-biased evolving set process* is a special case of the evolving set process where the markov chain is conditioned to be absorbed in V. In particular, the transition kernel is defined of a volume biased ESP as follows:

$$\mathbf{K}(S,S') = \frac{\mu(S')}{\mu(S)}\mathrm{K}(S,S').$$

Given a state S_0, we write $\mathbf{P}_{S_0}[.] := \mathbf{P}[.|S_0]$ to denote the probability measure of the volume biased ESP started at state S_0, and we use $\mathbf{E}_{S_0}[.] := \mathbf{E}[.|S_0]$ for the expectation.

Andersen and Peres used the volume biased ESP as a local graph clustering algorithm [3]. They show that for any non-expanding set U, if we run the volume biased ESP from a randomly chosen vertex of U, with a constant probability, there is a set in the sample path of conductance $O(\sqrt{\phi(U)\log n})$, and volume at most $2\mu(U)$. As a part of their proof, they designed an efficient simulation of the volume biased ESP, called `GenerateSample`. They prove the following theorem,

Theorem 7 (Andersen, Peres [3, Theorems 3,4])**.** *There is an algorithm,* `GenerateSample`*, that simulates the volume biased ESP such that for any vertex $v \in V$, any sample path $(S_0 = \{v\},\ldots,S_\tau)$, is generated with probability $\mathbf{P}_v[S_0,\ldots,S_\tau]$. Furthermore, for a stopping time τ that is bounded above by T, let $W(\tau)$ be the time complexity of* `GenerateSample` *if it is run up to time τ. Then, the expected work per volume ratio of the algorithm is*

$$\mathbf{E}_v\left[\frac{W(\tau)}{\mu(S_\tau)}\right] = O(T^{1/2}\log^{3/2}\mu(V)).$$

III. UPPER BOUNDS ON THE ESCAPING PROBABILITY OF RANDOM WALKS

In this section we establish strong results on the escaping probability of the random walks. Spielman and Teng [1] show that for any set $S \subseteq V$, $t > 0$, the random walk started at a randomly (proportional to degree) chosen vertex of S, remain in S for t steps with probability at least $1 - t\phi(S)/2$. We strengthen this result, by improving the lower bound to $(1 - \phi(S)/2)^t$,

Proposition 8. *For any set $S \subseteq V$, and integer $t > 0$,*

$$\mathbb{E}_{v\sim\boldsymbol{\pi}_S}\mathrm{rem}(v,t,S) \geq \left(1 - \frac{\phi(S)}{2}\right)\mathbb{E}_{v\sim\boldsymbol{\pi}_S}\mathrm{rem}(v,t-1,S)$$

$$\geq \ldots \geq \left(1 - \frac{\phi(S)}{2}\right)^t. \tag{2}$$

Furthermore, there is a subset $S^t \subseteq S$, such that $\mu(S^t) \geq \mu(S)/2$, and for all $v \in S^t$

$$\mathrm{rem}(v,t,S) \geq \frac{1}{200}\left(1 - \frac{3\phi(S)}{2}\right)^t. \tag{3}$$

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 [1]. Whence, here both of the results incorporate non-trivial spectral arguments.

As a corollary, we prove strong lower bounds on the uniform mixing time of random walks in Proposition 5. In the rest of this section we prove Proposition 8. We start by proving (2).

Using equation (1), and a simple induction on t, (2) is equivalent to the following equation:

$$\boldsymbol{\pi}_S'(I_S\mathcal{G}I_S)^t\mathbf{1}_S \geq (1 - \phi(S)/2)\boldsymbol{\pi}'(I_S\mathcal{G}I_S)^{t-1}\mathbf{1}_S. \quad (4)$$

Let $P := D^{1/2}I_S\mathcal{G}I_SD^{-1/2}$. First we show that (4) is equivalent to the following equation:

$$\sqrt{\boldsymbol{\pi}_S}'P^t\sqrt{\boldsymbol{\pi}_S} \geq \left(\sqrt{\boldsymbol{\pi}_S}'P\sqrt{\boldsymbol{\pi}_S}\right)\left(\sqrt{\boldsymbol{\pi}_S}'P^{t-1}\sqrt{\boldsymbol{\pi}_S}\right). \quad (5)$$

Then, we use Lemma 9 that shows the above equation holds for any symmetric positive semidefinite matrix P, and any norm one vector $\mathbf{x} = \sqrt{\boldsymbol{\pi}_S}$. First observe that by the definition of P, for any $t > 0$,

$$\boldsymbol{\pi}_S'(I_S\mathcal{G}I_S)^t\mathbf{1}_S = \boldsymbol{\pi}_S'D^{-1/2}P^tD^{1/2}\mathbf{1}_S = \sqrt{\boldsymbol{\pi}_S}'P^t\sqrt{\boldsymbol{\pi}_S} \quad (6)$$

On the other hand,

$$\begin{aligned}
\boldsymbol{\pi}_S'(I_S\mathcal{G}I_S)\mathbf{1}_S &= \frac{1}{2}\boldsymbol{\pi}_S'(D^{-1}A + I)\mathbf{1}_S \\
&= \frac{1}{2\mu(S)}(\mathbf{1}_S'A\mathbf{1}_S) + \frac{1}{2}\boldsymbol{\pi}_S'\mathbf{1}_S \\
&= \frac{1}{2\mu(S)}2|E(S)| + \frac{1}{2} \\
&= 1 - \phi(S)/2. \quad (7)
\end{aligned}$$

Equation (5) is derived simply from equation (4), by putting (6), (7) together. Next we prove equation (5) using Lemma 9. First observe that $\sqrt{\boldsymbol{\pi}_S}$ is a norm one vector. On the other hand, by definition $P = \frac{1}{2}(D^{-1/2}I_SAI_SD^{-1/2} + D^{-1/2}I_SD^{-1/2})$ is a symmetric matrix.

It remains to show that P is positive semidefinite. This follows from the same reason that $D^{1/2}\mathcal{G}D^{-1/2}$ is positive semidefinite. In particular, since eigenvectors of P can be transformed to the eigenvectors of $I_S\mathcal{G}I_S$, the eigenvalues of $I_S\mathcal{G}I_S$ are the same as the eigenvalues of P. Finally, since $\|I_SD^{-1}AI_S\|_\infty \leq 1$, and $I_S\mathcal{G}I_S$ is the average of $I_SD^{-1}AI_S$ and I_S, we must have $\lambda_{\min}(I_S\mathcal{G}I_S) \geq 0$, and $\lambda_{max}(I_S\mathcal{G}I_S) \leq 1$. Thus P is positive semidefinite. Now, (5) simply follows from Lemma 9. This completes the proof of (2)

It remains to prove (3). We prove it by showing that for any set $X \subseteq S$, of volume $\mu(X) \geq \mu(S)/2$, the random walk started at a randomly (proportional to degree) chosen vertex of X, remains in X (and S), with probability at least $\frac{1}{200}(1 - 3\phi(S)/2)^t$,

$$\begin{aligned}
\mathbb{E}_{v \sim \boldsymbol{\pi}_X}\left[\text{rem}(v, t, X)\right] &= \boldsymbol{\pi}_X'(I_S\mathcal{G}I_S)^t\mathbf{1}_X \\
&\geq \frac{1}{200}\left(1 - \frac{3\phi(S)}{2}\right)^t. \quad (8)
\end{aligned}$$

Therefore, in any such set X, there is a vertex that satisfy (3), hence the volume of the set of vertices that satisfy (3) is at least half of $\mu(S)$.

Using equations (6) and (7), (8) is equivalent to the following equation,

$$\sqrt{\boldsymbol{\pi}_X}'P^t\sqrt{\boldsymbol{\pi}_X} \geq \frac{1}{200}\left(3\sqrt{\boldsymbol{\pi}_S}'P\sqrt{\boldsymbol{\pi}_S} - 2\right)^t. \quad (9)$$

We prove the above equation using Lemma 10. Let $Y = S \setminus X$, and define

$$\begin{aligned}
\mathbf{x} &:= I_X\sqrt{\boldsymbol{\pi}_S} = \sqrt{\mu(X)\boldsymbol{\pi}_X/\mu(S)} \quad (10) \\
\mathbf{y} &:= I_Y\sqrt{\boldsymbol{\pi}_S} = \sqrt{\mu(Y)\boldsymbol{\pi}_Y/\mu(S)}
\end{aligned}$$

Since $X \cap Y = \emptyset$, $\langle\mathbf{x}, \mathbf{y}\rangle = 0$, and $\|\mathbf{x} + \mathbf{y}\| = \|\sqrt{\boldsymbol{\pi}_S}\| = 1$. Furthermore, since $\mu(X) \geq \mu(S)/2 \geq \mu(Y)$, $\|\mathbf{x}\| \geq \|\mathbf{y}\|$. Therefore, $P, \mathbf{x}, \mathbf{y}$ satisfy the requirements of Lemma 10. Finally, since $\sqrt{\boldsymbol{\pi}_X}'P^t\sqrt{\boldsymbol{\pi}_X} \geq \mathbf{x}'P^t\mathbf{x}$, (8) follows from Lemma 10. This completes the proof of Proposition 8.

Lemma 9. *Let $P \in \mathbb{R}^{n \times n}$ be a symmetric positive semidefinite matrix. Then, for any $\mathbf{x} \in \mathbb{R}^n$ of norm $\|x\| = 1$, and integer $t > 0$,*

$$\mathbf{x}'P^t\mathbf{x} \geq \left(\mathbf{x}'P^{t-1}\mathbf{x}\right)\left(\mathbf{x}'P\mathbf{x}\right) \geq \ldots \geq \left(\mathbf{x}'P\mathbf{x}\right)^t.$$

Proof: Since all of the inequalities in lemma's statement follows from the first inequality, we only prove the first inequality. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be the set of orthonormal eigenvectors of P with the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. For any $k \geq 1$, we have

$$\begin{aligned}
\mathbf{x}'P^k\mathbf{x} &= \left(\sum_{i=1}^n\langle\mathbf{x}, \mathbf{v}_i\rangle\mathbf{v}_i\right)P^k\left(\sum_{i=1}^n\langle\mathbf{x}, \mathbf{v}_i\rangle\mathbf{v}_i\right) \\
&= \left(\sum_{i=1}^n\langle\mathbf{x}, \mathbf{v}_i\rangle\lambda_i^k\mathbf{v}_i\right) \cdot \left(\sum_{i=1}^n\langle\mathbf{x}, \mathbf{v}_i\rangle\mathbf{v}_i\right) \\
&= \sum_{i=1}^n\langle\mathbf{x}, \mathbf{v}_i\rangle^2\lambda_i^k. \quad (11)
\end{aligned}$$

On the other hand, since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthornormal system, we have

$$\sum_{i=1}^n\langle\mathbf{x}, \mathbf{v}_i\rangle^2 = \|\mathbf{x}\|^2 = 1.$$

For any $k > 0$, Let $f_k(\lambda) = \lambda^k$; it follows that,

$$\sum_i\langle\mathbf{x}, \mathbf{v}_i\rangle^2\lambda_i^k = \mathbb{E}_{\Lambda \sim \mathcal{D}}\left[f_k(\Lambda)\right],$$

where $\mathbb{P}_{\Lambda \sim \mathcal{D}}\left[\Lambda = \lambda_i\right] = \langle\mathbf{x}, \mathbf{v}_i\rangle^2$. Using equation (11) we may rewrite the lemma's statement as follows:

$$\mathbb{E}_{\mathcal{D}}\left[f_{t-1}(\Lambda)f_1(\Lambda)\right] \geq \mathbb{E}_{\mathcal{D}}\left[f_{t-1}(\Lambda)\right]\mathbb{E}_{\mathcal{D}}\left[f_1(\Lambda)\right]$$

Since P is positive semidefinite, $\lambda_{\min}(P) \geq 0$. Thus, for all $t > 0$, the function $f_t(.)$ is increasing in the support of $\mathcal{D}$. The above inequality follows from the Chebyshev's sum inequality. ∎

Lemma 10. *Let $P \in \mathbb{R}^{n \times n}$ be a symmetric positive semidefinite matrix such that $\lambda_{\max}(P) \leq 1$, and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*

such that $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, $\|\mathbf{x} + \mathbf{y}\| = 1$, and $\|\mathbf{x}\| \geq \|\mathbf{y}\|$. Then, for any integer $t > 0$,

$$\mathbf{x}'P^t\mathbf{x} \geq \frac{1}{200}\left(3(\mathbf{x} + \mathbf{y})'P(\mathbf{x} + \mathbf{y}) - 2\right)^t$$

Due to the space limits, we defer the proof of this lemma to the full version of our paper.

IV. APPROXIMATING THE EXPANSION PROFILE

In this section we use the machinery developed in the works of Lovasz and Simonovits [20], [22] to prove Theorem 3, Theorem 4. We start by introducing some notations.

Let $\mathbf{p}$ be a probability distribution vector on the vertices of V, and let $\sigma(.)$ be the permutation of the vertices that is decreasing with respect to $p(v)/d(v)$, and breaking ties lexicographically. That is, suppose

$$\frac{p(\sigma(1))}{d(\sigma(1))} \geq \ldots \geq \frac{p(\sigma(i))}{d(\sigma(i))} \geq \ldots \geq \frac{p(\sigma(n))}{d(\sigma(n))}.$$

We use $T_i(\mathbf{p}) := \{\sigma(1), \ldots, \sigma(i)\}$ to denote the *threshold set* of the first i vertices. Following Spielman, Teng [1] (c.f. Lovasz, Simonovits [20]), we use the following potential function,

$$I(\mathbf{p}, x) := \max_{\substack{w \in [0,1]^n \\ \sum w(v)d(v) = x}} \sum_{v \in V} w(v)p(v). \qquad (12)$$

Observe that for $I(\mathbf{p}, x)$ is a non-decreasing piecewise linear concave function of x, that is $I(\mathbf{p}, x) = p(T_j(\mathbf{p}))$, for $x = \mu(T_j(\mathbf{p}))$, and is linear in other values of x. We use $I(\mathbf{p}, x)$ as a potential function to measure the distance of the distribution $\mathbf{p}$ from the stationary distribution $\boldsymbol{\pi}$.

We find the small non-expanding set in Theorem 3, by running $\mathtt{Threshold}(\sqrt{2\phi/\epsilon}, \epsilon \ln \mu(V)/\phi)$. The algorithm simply returns the smallest non-expanding set among the threshold sets of the rows of $\mathcal{G}^t$, for $t = O(\epsilon \log \mu(V)/\phi)$. The details are described in Algorithm 1.

Algorithm 1 $\mathtt{Threshold}(\phi, \epsilon)$

Let $\mathcal{T}$ be the family of all threshold sets $T_i(\mathbf{1}'_v\mathcal{G}^t)$, for any vertex $v \in V$, and $1 \leq t \leq \epsilon \log \mu(V)/\phi$, with conductance at most $\sqrt{2\phi/\epsilon}$.
return the set with minimum volume in $\mathcal{T}$.

If none of the sets $T_i(\mathbf{1}'_v\mathcal{G}^t)$ is a non-expanding set, then Lovasz and Siminovits [20], [22] prove that the curve $I(\mathbf{1}'_v\mathcal{G}^t, x)$ lies far below $I(\mathbf{1}'_v\mathcal{G}^{t-1}, x)$. This is quantified in the following lemma:

Lemma 11 (Lovasz, Simonovits [22, Lemma 1.3]). *Let $\mathcal{G}$ be a transition probability matrix of a lazy random walk on a graph. For any probability distribution vector $\mathbf{p}$ on V, if*

$\phi(T_i(\mathbf{p}'\mathcal{G})) \geq \Phi$, *then for $x = \mu(T_i(\mathbf{p}'\mathcal{G}))$,*

$$\begin{aligned} I(\mathbf{p}'\mathcal{G}, x) &\leq \frac{1}{2}I(\mathbf{p}, x - 2\Phi\min(x, 2m - x)) \\ &+ \frac{1}{2}I(\mathbf{p}, x + 2\Phi\min(x, 2m - x)). \end{aligned}$$

By repeated application of the above lemma, Lovasz and Simonovits [20] argue that, if all of the sets $T_i(\mathbf{1}'_v\mathcal{G}^t)$ are expanding, then $I(\mathbf{1}'_v\mathcal{G}^t, .)$ approaches the straight line. In the next lemma we show that, if all of the small threshold sets (i.e., $\mu(T_i(\mathbf{1}'_v\mathcal{G}^t)) \leq \Gamma$), are expanding, then $I(\mathbf{1}'_v\mathcal{G}^t, .)$ approaches the curve x/Γ.

Lemma 12. *For any vertex $v \in V$, $t \geq 0$, and $0 \leq \Gamma \leq m$, $0 \leq \Phi \leq 1/2$, if for all $t \leq T$, all of threshold sets $T_i(\mathbf{1}'_v\mathcal{G}^t)$ of volume at most Γ, has conductance at least Φ, then for any $0 \leq t \leq T$,*

$$I(\mathbf{1}'_v\mathcal{G}^t, x) \leq \frac{x}{\Gamma} + \sqrt{x/\mu(v)}\left(1 - \frac{\Phi^2}{2}\right)^t.$$

Proof: We prove by induction. The lemma trivially holds for $t = 0$. This is because the LHS is $x/\mu(v)$ for $0 \leq x \leq \mu(v)$ and 1 for larger values of x, while the RHS is $\sqrt{x/\mu(v)}$ for all $x \geq 0$. Next, we prove the lemma's statement holds for t, assuming that it holds for $t - 1$. Let $\mathbf{p} := \mathbf{1}'_v\mathcal{G}^{t-1}$. First of all, since $I(\mathbf{p}'\mathcal{G}, .)$ is a piecewise-linear concave function of x, it is sufficient to prove the statement for values of $x = \mu(T_i(\mathbf{p}'\mathcal{G}))$. For $x \geq \Gamma$, the statement holds trivially, because the RHS is at least 1, while the LHS is less than or equal to 1. Now, suppose $x < \Gamma$, and $x = \mu(T_i(\mathbf{p}'\mathcal{G}))$. Using Lemma 11, we have

$$\begin{aligned} I(\mathbf{p}'\mathcal{G}, x) &\leq \frac{1}{2}\{I(\mathbf{p}, x - 2\Phi x) + I(\mathbf{p}, x + 2\Phi x)\} \\ &\leq \frac{1}{2}\left\{\frac{2x}{\Gamma} + \sqrt{x/\mu(v)}\left(1 - \frac{\Phi^2}{2}\right)^{t-1} \right. \\ &\qquad \left. \left(\sqrt{1 - 2\Phi} + \sqrt{1 + 2\Phi}\right)\right\} \\ &\leq \frac{x}{\Gamma} + \sqrt{x/\mu(v)}\left(1 - \frac{\Phi^2}{2}\right)^t, \end{aligned}$$

where the first inequality uses the assumption that $x < \Gamma \leq m$, the second inequality uses the induction hypothesis, and the last inequality uses the inequality

$$\frac{1}{2}\left(\sqrt{1 - 2\Phi} + \sqrt{1 + 2\Phi}\right) \leq 1 - \frac{\Phi^2}{2},$$

holds for any $\Phi \leq 1/2$. ∎

Now we are ready to prove Theorem 3: using Proposition 8 we show that $I(\mathbf{1}'_v\mathcal{G}^t)$ does not converge to x/Γ, for $t \approx \log \gamma/\phi$. Therefore, by the previous lemma, at least one of the small threshold sets is non-expanding.
Proof of Theorem 3. Let, $\gamma = \mu(U)$, $T = \epsilon \ln \gamma/\phi$, $\Gamma = 2\gamma^{1+\epsilon}$, and $\Phi = \sqrt{2\phi/\epsilon}$. We show that $\mathtt{Threshold}(\phi, \epsilon)$ returns a set of volume at most Γ, and conductance at most Φ. Wlog we may assume that $\Gamma < m$, otherwise

the statement is trivial. We prove by contradiction; assume that the output of the algorithm has volume larger than Γ, we show that Lemma 12 and Proposition 8 can not hold simultaneously. First of all, by Proposition 8, there exists a vertex $u \in U$, such that

$$\text{rem}(u,t,U) \geq \left(1 - \frac{\phi(U)}{2}\right)^t \geq \left(1 - \frac{\phi}{2}\right)^{\frac{\epsilon \ln \gamma}{\phi}} \geq e^{-\epsilon \ln \gamma}$$
$$\geq \gamma^{-\epsilon}.$$

Let $\mathbf{p} = \mathbf{1}'_u \mathcal{G}^T$. Let $w(v) = 1$, for all $v \in U$, and $w(v) = 0$ for the rest of the vertices. By equation (12), we have

$$I(\mathbf{p}, \gamma) \geq \sum_{v \in V} w(v) p(u) = \sum_{v \in U} p(v) \geq \text{rem}(u,t,U) \geq \gamma^{-\epsilon}.$$

On the other hand, by Lemma 12, we have:

$$I(\mathbf{p}, \gamma) \leq \frac{\gamma}{\Gamma} + \sqrt{\gamma}\left(1 - \frac{\Phi^2}{2}\right)^t \leq \frac{1}{2\gamma^\epsilon} + \sqrt{\gamma}(1 - \frac{\phi}{\epsilon})^{\frac{\epsilon \ln \gamma}{\phi}}$$
$$< \gamma^{-\epsilon},$$

which is a contradiction. The last inequality uses the fact that $\epsilon < 1/2$. $\qquad\square$

V. ALMOST OPTIMAL LOCAL GRAPH CLUSTERING

In this section we use the volume biased ESP to design a local graph clustering algorithm with a worst case guarantee on the conductance of output set that is independent of the size of G.

Let $(S_0, S_1, \ldots, S_\tau)$ be a sample path of the volume biased ESP, for a stopping time τ. Andersen and Peres show that with a constant probability the conductance of at least one of the sets in the sample path is at most $O(\sqrt{\frac{1}{\tau} \log \mu(S_\tau)})$,

Lemma 13 ([3, Lemma1,Corollary 1]). *For any starting set S_0, and any stopping time τ, and $\alpha > 0$,*

$$\mathbf{P}_{S_0}\left[\sum_{i=1}^\tau \phi^2(S_i) \leq 4\alpha \ln \frac{\mu(S_\tau)}{\mu(S_0)}\right] \geq 1 - \frac{1}{\alpha}.$$

Here, we strengthen the above result, and we show the event occurs with much higher probability. In particular, we show the with probability at least $1 - 1/\alpha$, the conductance of at least one of the sets in the sample path is at most $O(\sqrt{\frac{1}{\tau} \log(\alpha \cdot \mu(S_\tau))})$.

Lemma 14. *For any starting set $S_0 \subseteq V$, and any stopping time τ, and $\alpha \geq 0$,*

$$\mathbf{P}_{S_0}\left[\sum_{i=1}^\tau \phi^2(S_i) \leq 8\left(\ln \alpha + \ln \frac{\mu(S_\tau)}{\mu(S_0)}\right)\right] \geq 1 - \frac{1}{\alpha}$$

Proof: Let

$$M_t := \frac{\sqrt{\mu(S_0)}}{\sqrt{\mu(S_t)}} \prod_{i=0}^{t-1} \frac{1}{1 - \psi(S_i)}$$

Andersen, Peres [3, Lemma 1] show that M_t is a martingale in the volume biased ESP. It follows from the optional sampling theorem [23] that $\mathbf{E}[M_\tau] = M_0 = 1$. Thus, by the Markov inequality, for any $\alpha > 0$, we have

$$\mathbf{P}[M_\tau \leq \alpha] \geq 1 - \frac{1}{\alpha}$$

By taking logarithm, from both sides of the event in the above equation we obtain,

$$\mathbf{P}[\ln M_\tau \leq \ln \alpha] \geq 1 - \frac{1}{\alpha} \qquad (13)$$

On the other hand, by the definition of M_τ,

$$\ln M_\tau = \frac{1}{2} \ln \frac{\mu(S_0)}{\mu(S_\tau)} + \ln \prod_{i=0}^{\tau-1} \frac{1}{1 - \psi(S_i)}$$
$$\geq \frac{1}{2} \ln \frac{\mu(S_0)}{\mu(S_\tau)} + \sum_{i=0}^{\tau-1} \psi(S_i)$$
$$\geq \frac{1}{2} \ln \frac{\mu(S_0)}{\mu(S_\tau)} + \frac{1}{8} \sum_{i=0}^{\tau-1} \phi^2(S_i), \qquad (14)$$

where the first inequality follows by the fact that $1/(1 - \psi(S_i)) \geq e^{\psi(S_i)}$, and the last inequality follows by Proposition 6. Putting (13) and (14) together proves the lemma. $\qquad\blacksquare$

The previous lemma shows that for any $\gamma, \phi > 0$, if we can run the process for $T \approx \epsilon \log \gamma / \phi$ steps without observing a set larger than $\gamma^{O(1)}$, then, with probability $1 - 1/\gamma$, one of the sets in the sample path must have a conductance of $O(\sqrt{\phi/\epsilon})$, which is what we are looking for. Next we use the following lemma by Andersen and Peres, together with Proposition 8, to show that event occurs with some non-zero probability. That is, for any $\epsilon < 1$, with probability at least $\approx \gamma^{-\epsilon}$, the volume of all sets in the sample path of the process are at most $O(\gamma^{1+\epsilon})$. Then, by the union bound we can argue both event occur with probability at least $\Omega(\gamma^{-\epsilon})$.

Lemma 15 (Andersen, Peres [3, Lemma 2]). *For any set $U \subset V$, $v \in U$, and integer $T > 0$, the following holds,*

$$\mathbf{P}_v\left[\max_{t \leq T} \frac{\mu(S_t \setminus U)}{\mu(S_t)} > \beta \, \text{esc}(v, T, U)\right] < \frac{1}{\beta}, \quad \forall \beta > 0$$

Lemma 16. *Given $\gamma > 0$, $0 < \phi < 1/4$, $0 < \epsilon < 1$, such that G has a set $U \subset V$ of volume $\mu(U) \leq \gamma$, and conductance $\phi(U) \leq \phi$. Let $T = \epsilon \ln \gamma / 3\phi$. There is a constant $c > 0$, and a subset $U^T \subseteq U$ of volume $\mu(U^T) \geq \mu(U)/2$ such that for any $v \in U^T$, with probability at least $c\gamma^{-\epsilon}/8$, a sample path $(S_1, S_2, \ldots, S_T)$ of the volume biased ESP started from $S_0 = \{v\}$ satisfies the following,*

  i) For some $t \in [0, T]$, $\phi(S_t) \leq O(\sqrt{100(1 - \ln c)\phi/\epsilon})$,

  ii) For all $t \in [0, T]$, $\mu(S_t \cap U) \geq c\gamma^{-\epsilon}\mu(S_t)/2$, and henceforth, $\mu(S_t) \leq 2\gamma^{1+\epsilon}/c$.

Proof: First of all, we let U^T be the set of vertices $v \in U$ such that

$$\text{rem}(v, T, U) \geq c \left(1 - \frac{3\phi(U)}{2}\right)^T.$$

By Proposition 8, there exists a constant $c > 0$ such that $\mu(U^T) \geq \mu(U)/2$. In the rest of the proof let v be a vertex in U^T. We have,

$$\text{esc}(v, T, U) \leq 1 - c\left(1 - \frac{3\phi(U)}{2}\right)^T \leq 1 - c\left(1 - \frac{3\phi}{2}\right)^{\frac{\epsilon \ln \gamma}{3\phi}}$$
$$\leq 1 - c\gamma^{-\epsilon}$$

Now, let $\beta := 1 + c\gamma^{-\epsilon}/2$. By Lemma 15, we have

$$\mathbf{P}_v \left[\max_{t \leq T} \frac{\mu(S_t \setminus U)}{\mu(S_t)} \leq \beta \, \text{esc}(v, T, U) \leq 1 - \frac{c\gamma^{-\epsilon}}{2}\right] \geq 1 - \frac{1}{\beta}$$
$$\geq \frac{c\gamma^{-\epsilon}}{4}$$

Since for any $S \subset V$, $\mu(S \setminus U) + \mu(S \cap U) = \mu(S)$, we have

$$\mathbf{P}_v \left[\min_{t \leq T} \frac{\mu(S_t \cap U)}{\mu(S_t)} \geq \frac{c\gamma^{-\epsilon}}{2}\right] \geq \frac{c\gamma^{-\epsilon}}{4}$$

On the other hand, let $\alpha := \gamma$. By Lemma 14, with probability $1 - 1/\gamma$, for some $t \in [0, T]$,

$$\phi^2(S_t) \leq \frac{1}{T} \sum_{t=0}^{T} \phi^2(S_t) \leq \frac{8(\ln \gamma + \ln \mu(S_T))}{T}.$$

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

$$\mathbf{P}_v \left[\min_{t \leq T} \frac{\mu(S_t \cap U)}{\mu(S_t)} \geq \frac{c\gamma^{-\epsilon}}{2} \bigwedge \right.$$
$$\left. \exists t : \phi(S_t) \leq \sqrt{\frac{8(\ln \gamma + \ln \mu(S_T))}{T}} \right] \geq \frac{c\gamma^{-\epsilon}}{8}$$

Finally, since for any set $S \subseteq V$, $\mu(S \cap U) \leq \mu(U) \leq \gamma$, in the above event, $\mu(S_T) \leq \frac{2\gamma^{1+\epsilon}}{c}$. Therefore,

$$\phi(S_t) \leq \sqrt{\frac{8(\ln \gamma + \ln(2\gamma^{1+\epsilon}/c))}{T}} \leq \sqrt{\frac{100(1 - \ln c)\phi}{\epsilon}},$$

which completes the proof. ∎

To prove Theorem 1, we can simply run γ^ϵ copies of the volume biased ESP in parallel. Using the previous lemma with a constant probability at least one of the copies finds a non-expanding set. Moreover, we may bound the time complexity of the algorithm using Lemma 7. The details of the algorithm is described in Algorithm 2.

Now we are ready to prove Theorem 1.

Theorem 1. `ParESP`$(v, \gamma, \phi, \epsilon)$ *takes as input a starting vetex $v \in V$, a target conductance $\phi \in (0, 1)$, a target size γ, and $0 < \epsilon < 1$. For a given run of the algorithm it outputs a set S of vertices with the expected work per volume ratio*

Algorithm 2 `ParESP`$(v, \gamma, \phi, \epsilon)$

1: Let $S_0 \leftarrow \{v\}$, and $T \leftarrow \epsilon \ln \gamma / 6\phi$, and c as defined in Lemma 16.
2: Run $\gamma^{\epsilon/2}$ independent copies of the volume biased ESP, using the simulator `GenerateSample`, starting from S_0, in parallel. Stop each copy as soon as the length of its sample path reaches T.
3: If any of the copies finds a set S, of volume $\mu(S) \leq 2\gamma^{1+\epsilon/2}/c$, and conductance $\phi(S) \leq \sqrt{200(1 - \ln c)\phi/\epsilon}$, stop the algorithm and return S.

---

of $O(\gamma^\epsilon \phi^{-1/2} \log^2 n)$. If $U \subseteq V$ is a set of vertices that satisfy $\phi(U) \leq \phi$, and $\mu(U) \leq \gamma$, then there is a subset $U' \subseteq U$ with volume at least $\mu(U)/2$, such that if $v \in U'$, with a constant probability S satisfies,

1) $\phi(S) = O(\sqrt{\phi/\epsilon})$,
2) $\mu(S) \leq O(\gamma^{1+\epsilon})$.

Proof: Let $U' = U^T$ as defined in Lemma 16. First of all, for any $v \in U'$, by Lemma 16, each copy of volume biased ESP, with probability $\Omega(\gamma^{-\epsilon/2})$, finds a set S such that $\mu(S) \leq 2\gamma^{1+\epsilon/2}/c$, and $\phi(S) \leq \sqrt{200(1 - \ln c)\phi/\epsilon}$; but, since $\gamma^{\epsilon/2}$ copies are executed independently, at least one of them will succeed with a constant probability. Therefore, with a constant probability the output set will satisfy properties (1), (2) in theorem's statement. This proves the correctness of the algorithm.

It remains to compute the time complexity. Let, $k := \gamma^{\epsilon/2}$ be the number of copies, and $W_1, \ldots, W_k$ be random variables indicating the work done by each of the copies in a single run of `ParESP`, thus $\sum_i W_i$ is the time complexity of the algorithm. Let M be the random variable indicating the volume of the output set of the algorithm; we let $M = 0$ if the algorithm does not return any set. Also, for $1 \leq i \leq k$, let X_i be $1/M$ if the output set is chosen from the ith copy, and 0 otherwise, and let $X := \sum X_i$. We write $\mathbf{P}_v^k[.]$ to denote the probability measure of the k independent volume biased ESP all started from $S_0 = \{v\}$, and $\mathbf{E}_v^k[.]$ for the expectation. To prove the theorem it is sufficient to show

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

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

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

By symmetry of the copies, it is sufficient to show that only for $i = 1$. Furthermore, since conditioned on $X_1 \neq 0$, $W_1 = \max_i W_i$, we just need to show,

$$\mathbf{E}_v^k [X_1 W_1] = O(\phi^{-1/2} \log^2 n),$$

Let τ be a stopping time, bounded from above by T, indicating the first time where a set S_τ of volume $\mu(S_\tau) \leq 2\gamma^{1+\epsilon}/c$, and conductance $\phi(S_\tau) \leq \sqrt{200(1 - \log c)\phi/\epsilon}$ is observed in the first copy if it is executed up to time τ, and $W_1(\tau)$ as the amount of work done by that time. Observe that for any element of the joint probability space, $X_1 W_1 \leq W_1(\tau)/\mu(S_\tau)$. This is because, we always have $W_1 \leq W_1(\tau)$, and $X_1 \leq \mu(S_\tau)$. Therefore,

$$
\begin{aligned}
\mathbf{E}_v^k[X_1 W_1] &\leq \mathbf{E}_v^k\left[\frac{W_1(\tau)}{\mu(S_\tau)}\right] = \mathbf{E}_v\left[\frac{W(\tau)}{\mu(S_\tau)}\right] \\
&= O(T^{1/2} \log^{3/2} n) = O(\phi^{-1/2} \log^2 n),
\end{aligned}
$$

where the second to last equation follows from Lemma 7. ∎

VI. Discussion

We proved that by running multiple copies of the Evolving Set Process algorithm of Andersen and Peres [3], we get a local graph clustering algorithm with almost the same guarantee as of the Cheeger's inequalities. Unfortunately, although the running time of this algorithm is independent of the size of G, it is not linear in the size of the output. It remains an open problem if we can make the running time truly linear in the size of the output, while providing the same approximating guarantee. Such a result would give a truly optimal local graph clustering algorithm.

Raghavendra and Steurer [13] conjectured that Problem 2 is NP-hard. Although our approximation algorithm to the expansion profile does not refute this conjecture, we are not aware of any bad examples. It is a fascinating open problem if one can provide hard examples of the SSE problem that can not be solved by running $\text{poly}(n)$ copies of the evolving set processes.

Acknowledgements.

References

[1] D. A. Spielman and S.-H. Teng, "A local clustering algorithm for massive graphs and its application to nearly-linear time graph partitioning," *CoRR*, vol. abs/0809.3232, 2008.

[2] R. Andersen, F. R. K. Chung, and K. J. Lang, "Local graph partitioning using pagerank vectors," in *FOCS*, 2006, pp. 475–486.

[3] R. Andersen and Y. Peres, "Finding sparse cuts locally using evolving sets," in *STOC*, 2009, pp. 235–244.

[4] S. Arora, B. Barak, and D. Steurer, "Subexponential algorithms for unique games and related problems," in *FOCS*, 2010, pp. 563–572.

[5] N. Alon and V. Milman, "Isoperimetric inequalities for graphs, and superconcentrators," *Journal of Combinatorial Theory, Series B*, vol. 38, no. 1, pp. 73–88, feb 1985.

[6] N. Alon, "Eigen values and expanders," *Combinatorica*, vol. 6, pp. 83–96, January 1986.

[7] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, pp. 668–677, 1999.

[8] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[9] D. A. Tolliver and G. L. Miller, "Graph partitioning by spectral rounding: Applications in image segmentation and clustering," in *CVPR*, 2006, pp. 1053–1060.

[10] L. Orecchia, S. Sachdeva, and N. K. Vishnoi, "Approximating the exponential, the lanczos method and an $\tilde{O}(m)$-time spectral algorithm for balanced separator," in *STOC*, 2012.

[11] R. Kannan, S. Vempala, and A. Vetta, "On clusterings: Good, bad and spectral," *Journal of The ACM*, vol. 51, pp. 497–515, 2004.

[12] D. A. Spielman and S.-H. Teng, "Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems," in *STOC*, 2004, pp. 81–90.

[13] P. Raghavendra and D. Steurer, "Graph expansion and the unique games conjecture," in *STOC*, 2010, pp. 755–764.

[14] P. Diaconis and J. A. Fill, "Strong stationary times via a new form of duality," *Ann. Probab.*, vol. 18, no. 4, pp. 1483–1522, 1990.

[15] L. Lovász and R. Kannan, "Faster mixing via average conductance," in *STOC*, 1999, pp. 282–287.

[16] P. Raghavendra, D. Steurer, and P. Tetali, "Approximations for the isoperimetric and spectral profile of graphs and related parameters," in *STOC*, 2010, pp. 631–640.

[17] N. Bansal, U. Feige, R. Krauthgamer, K. Makarychev, V. Nagarajan, J. Naor, and R. Schwartz, "Min-max graph partitioning and small set expansion." in *FOCS*, 2011, pp. 17–26.

[18] T. C. Kwok and L. C. Lau, "Finding small sparse cuts by random walk," in *APPROX-RANDOM*, 2012, pp. 615–626.

[19] R. O'Donnell and D. Witmer, "Improved small-set expansion from higher eigenvalues," *CoRR*, vol. abs/1204.4688, 2012.

[20] L. Lovász and M. Simonovits, "The mixing rate of markov chains, an isoperimetric inequality, and computing the volume," in *FOCS*, 1990, pp. 346–354.

[21] B. Morris and Y. Peres, "Evolving sets and mixing," in *STOC*, 2003, pp. 279–286.

[22] L. Lovász and M. Simonovits, "Random walks in a convex body and an improved volume algorithm," *Random Struct. Algorithms*, vol. 4, no. 4, pp. 359–412, 1993.

[23] D. Williams, *Probability with Martingales (Cambridge Mathematical Textbooks)*. Cambridge University Press, Feb. 1991.