A Local Algorithm for Finding Well-Connected Clusters

Zeyuan Allen Zhu ZEYUAN@CSAIL.MIT.EDU
MIT CSAIL, 32 Vassar St., Cambridge, MA 02139 USA

Silvio Lattanzi SILVIOL@GOOGLE.COM
Vahab Mirrokni MIRROKNI@GOOGLE.COM
Google Research, 111 8th Ave., 4th floor, New York, NY 10011 USA

Abstract

Motivated by applications of large-scale graph clustering, we study random-walk-based *local* algorithms whose running times depend only on the size of the output cluster, rather than the entire graph. In particular, we develop a method with better theoretical guarantee compared to all previous work, both in terms of the clustering accuracy and the conductance of the output set. We also prove that our analysis is tight, and perform empirical evaluation to support our theory on both synthetic and real data.

More specifically, our method outperforms prior work when the cluster is *well-connected*. In fact, the better it is well-connected inside, the more significant improvement we can obtain. Our results shed light on why in practice some random-walk-based algorithms perform better than its previous theory, and help guide future research about local clustering.

1. Introduction

As a central problem in machine learning, clustering methods have been applied to data mining, computer vision, social network analysis. Although a huge number of results are known in this area, there is still need to explore methods that are robust and efficient on large data sets, and have good theoretical guarantees. In particular, several algorithms restrict the number of clusters, or impose constraints that make these algorithms impractical for large data sets.

To solve those issues, recently, local random-walk clustering algorithms (Spielman & Teng, 2004; Andersen et al., 2006) have been introduced. The main idea behind those algorithms is to find a good cluster around a specific node. These techniques, thanks to their scalability, has had high impact in practical applications (Leskovec et al., 2009; Gargi et al., 2011; Gleich & Seshadhri, 2012; Andersen et al., 2012; Leskovec et al., 2010; Wu et al., 2012). Nevertheless, the theoretical understanding of these techniques is still very limited. In this paper, we make an important contribution in this direction. First, we relate for the first time the performance of these local algorithms with the *internal connectivity* of a cluster instead of analyzing only its external connectivity. This change of perspective is relevant for practical applications where we are not only interested to find clusters that are loosely connected with the rest of the world, but also clusters that are well-connected internally. In particular, we show theoretically and empirically that this internal connectivity is a fundamental parameter for those algorithms and, by leveraging it, it is possible to improve their performances.

Formally, we study the clustering problem where the data set is given by a similarity matrix as a graph: given an undirected[1] graph $G = (V, E)$, we want to find a set S that minimizes the relative number of edges going out of S with respect to the size of S (or the size of $\bar{S}$ if S is larger than $\bar{S}$). To capture this concept rigorously, we consider the *cut conductance* of a set S as:

$$\phi_{\mathsf{c}}(S) \stackrel{\text{def}}{=} \frac{|E(S,\bar{S})|}{\min\{\text{vol}(S), \text{vol}(\bar{S})\}} \ ,$$

where $\text{vol}(S) \stackrel{\text{def}}{=} \sum_{v \in S} \deg(v)$. Finding S with the smallest $\phi_{\mathsf{c}}(S)$ is called the conductance minimization. This measure is a well-studied measure in different disciplines (Shi & Malik, 2000; Spielman & Teng, 2004; Andersen et al., 2006; Gargi et al., 2011; Gleich & Seshadhri, 2012), and has been identified as one of the most important cut-based measures in the literature (Schaeffer, 2007). Many approximation algo-

Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28.

[1] All our results can be generalized to weighted graphs.

rithms have been developed for the problem, but most of them are global ones: their running time depends at least linearly on the size of the graph. A recent trend, initiated by Spielman and Teng (2004), and then followed by (Spielman & Teng, 2008; Andersen et al., 2006; Andersen & Peres, 2009; Gharan & Trevisan, 2012), attempts to solve this conductance minimization problem *locally*, with running time only dependent on the volume of the output set.

In particular, if there exists a set $A \subset V$ with $\phi_{\mathsf{c}}(A) \leq \Psi$, these local algorithms guarantee the existence of some set $A^g \subseteq A$ with at least half the volume, such that for any "good" starting vertex $v \in A^g$, they output a set S with conductance $\phi_{\mathsf{c}}(S) = \tilde{O}(\sqrt{\Psi})$.

Finding Well-Connectedness Clusters. All local clustering algorithms developed so far, both theoretical ones and empirical ones, only assume that $\phi_{\mathsf{c}}(A)$ is small, i.e., A is poorly connected to $\bar{A}$. Notice that such set A, no matter how small $\phi_{\mathsf{c}}(A)$ is, may be poorly connected or even disconnected inside. This cannot happen in reality if A is a "good" cluster, and in practice we are often interested in finding mostly good clusters. This motivates us to study an extra measure on A, that is the connectedness of A, denoted as $\mathtt{Conn}(A)$ and we will define it formally in Section 2. We assume that, in addition to prior work, the cluster A satisfies the *gap assumption*

$$\mathtt{Gap} = \mathtt{Gap}(A) \stackrel{\text{def}}{=} \frac{\mathtt{Conn}(A)}{\Psi} \geq \Omega(1) \quad ,$$

which says that A is better connected inside than it is connected to $\bar{A}$. This assumption is particularly relevant when the edges of the graph represent pairwise similarity scores extracted from a machine learning algorithm: we would expect similar nodes to be well connected within themselves while dissimilar nodes to be loosely connected. As a result, it is not surprising that the notion of connectedness is not new. For instance (Kannan et al., 2004) studied a bicriteria optimization for this objective. However, local algorithms based on the above gap assumption is not well studied.[2]

Our Results. Under the gap assumption $\mathtt{Gap} \geq \Omega(1)$, can we guarantee any better cut conductance than the previously shown $\tilde{O}(\sqrt{\Psi})$ ones? We prove that the answer is affirmative, along with some other desirable properties. In particular, we prove:

Theorem 1. *If there exists a non-empty set $A \subset V$ such that $\phi_{\mathsf{c}}(A) \leq \Psi$ and $\mathtt{Gap} \geq \Omega(1)$, then there exists some $A^g \subseteq A$ with $\mathrm{vol}(A^g) \geq \frac{1}{2}\mathrm{vol}(A)$ such that, when choosing a starting vertex $v \in A^g$, the* `PageRank-Nibble` *algorithm outputs a set S with*

1. $\mathrm{vol}(S \setminus A) \leq O(\frac{1}{\mathtt{Gap}})\mathrm{vol}(A)$,
2. $\mathrm{vol}(A \setminus S) \leq O(\frac{1}{\mathtt{Gap}})\mathrm{vol}(A)$,
3. $\phi_{\mathsf{c}}(S) \leq O(\sqrt{\Psi/\mathtt{Gap}})$, *and*

with running time $O(\frac{\mathrm{vol}(A)}{\Psi \cdot \mathtt{Gap}}) \leq O(\frac{\mathrm{vol}(A)}{\Psi})$.

We interpret the above theorem as follows. The first two properties imply that under $\mathtt{Gap} \geq \Omega(1)$, the volume for $\mathrm{vol}(S \setminus A)$ and $\mathrm{vol}(A \setminus S)$ are both small in comparison to $\mathrm{vol}(A)$, and the larger the gap is, the more accurate S approximates A.[3] For the third property on the cut conductance $\phi_{\mathsf{c}}(S)$, we notice that our guarantee $O(\sqrt{\Psi/\mathtt{Gap}}) \leq O(\sqrt{\Psi})$ outperforms all previous work on local clustering under this gap assumption. In addition, $\mathtt{Gap}$ might be very large in reality. For instance when A is a very-well-connected cluster it might satisfy $\mathtt{Conn}(A) = \mathsf{polylog}(n)$, and as a consequence $\mathtt{Gap}$ may be as large as $\tilde{\Omega}(1/\Psi)$. In this case our Theorem 1 guarantees a $\mathsf{polylog}(n)$ approximation to the cut conductance.

Our proof of Theorem 1 uses almost the same PageRank algorithm as (Andersen et al., 2006), but with a very different analysis specifically designed for our gap assumption. This algorithm is simple and clean, and can be described in four steps: 1) compute the (approximate) PageRank vector starting from a vertex $v \in A^g$ with carefully chosen parameters, 2) sort all the vertices according to their (normalized) probabilities in this vector, 3) study all *sweep cuts* that are those separating high-value vertices from low-value ones, and 4) output the sweep cut with the best cut conductance. See Algorithm 1 for details.

We also prove that our analysis is tight.

Theorem 2. *There exists a graph $G = (V, E)$ and a non-empty $A \subset V$ with Ψ and $\mathtt{Gap} = \Omega(1)$, such that for all starting vertices $v \in A$, none of the sweep-cut based algorithm on the PageRank vector can output a set S with cut conductance better than $O(\sqrt{\Psi/\mathtt{Gap}})$.*

We prove this tightness result by illustrating a hard instance, and proving upper and lower bounds on the probabilities of reaching specific vertices (up to a very high precision). Theorem 2 does not rule out existence of another local algorithm that can perform better than $O(\sqrt{\Psi/\mathtt{Gap}})$. However, we conjecture that all existing (random-walk-based) local clustering algorithms share the same hard instance and do not

[2] One relevant paper using this assumption is (Makarychev et al., 2012), who provided a *global* SDP-based algorithm to approximate the cut conductance.

[3] Very recently, (Wu et al., 2012) studied a variant of the PageRank random walk and their first experiment — although analyzed in a different perspective— essentially confirmed our first two properties in Theorem 1. However, they have not attempted to explain this in theory.

outperform $O(\sqrt{\Psi/\texttt{Gap}})$, similar to the classical case where they all provide only $\tilde{O}(\sqrt{\Psi})$ guarantee due to Cheeger's inequality. It is an interesting open question to design a flow-based local algorithm to overcome this barrier under our gap assumption.

Prior Work. Related work is discussed in depth in the full version of this paper.

Roadmap. We provide preliminaries in Section 2, and they are followed by the high level ideas of the proofs for Theorem 1 in Section 3 and Section 4. We then briefly describe how to prove our tightness result in Theorem 5, and end this extended abstract with empirical studies in Section 6.

2. Preliminaries

2.1. Problem Formulation

Consider an undirected graph $G(V, E)$ with $n = |V|$ vertices and $m = |E|$ edges. For any vertex $u \in V$ the degree of u is denoted by $\deg(u)$, and for any subset of the vertices $S \subseteq V$, *volume* of S is denoted by $\mathrm{vol}(S) \stackrel{\text{def}}{=} \sum_{u \in S} \deg(u)$. Given two subsets $A, B \subset V$, let $E(A, B)$ be the set of edges between A and B.

For a vertex set $S \subseteq V$, we denote by $G[S]$ the induced subgraph of G on S with outgoing edges removed, by $\deg_S(u)$ the degree of vertex $u \in S$ in $G[S]$, and by $\mathrm{vol}_S(T)$ the volume of $T \subseteq S$ in $G[S]$.

We respectively define the *cut conductance* and the *set conductance* of a non-empty set $S \subseteq V$ as follows:

$$\phi_{\mathtt{c}}(S) \stackrel{\text{def}}{=} \frac{|E(S, \bar{S})|}{\min\{\mathrm{vol}(S), \mathrm{vol}(\bar{S})\}} \ ,$$

$$\phi_{\mathtt{s}}(S) \stackrel{\text{def}}{=} \min_{\emptyset \subset T \subset S} \frac{|E(T, S \setminus T)|}{\min\{\mathrm{vol}_S(T), \mathrm{vol}_S(S \setminus T)\}} \ .$$

Here $\phi_{\mathtt{c}}(S)$ is classically known as the conductance of S, and $\phi_{\mathtt{s}}(S)$ is classically known as the conductance of S on the induced subgraph $G[S]$.

We formalize our goal in this paper as a *promise problem*. Specifically, we assume the existence of a non-empty cluster of the vertices $A \subset V$ satisfying $\mathrm{vol}(A) \leq \frac{1}{2}\mathrm{vol}(V)$ as well as $\phi_{\mathtt{s}}(A) \geq \Phi$ and $\phi_{\mathtt{c}}(A) \leq \Psi$. This set A is *not* known to the algorithm. The goal is to find some set S that "reasonably" approximates A, and at the same time be *local*: running in time proportional to $\mathrm{vol}(A)$ rather than n or m.

Our assumption. We assume that the following gap assumption:

$$\texttt{Gap} \stackrel{\text{def}}{=} \frac{\texttt{Conn}(A)}{\Psi} \stackrel{\text{def}}{=} \frac{\Phi^2/\log \mathrm{vol}(A)}{\Psi} \geq \Omega(1) \qquad \text{(Gap Assumption)}$$

holds throughout this paper. This assumption can be understood as the cluster A is more well-connected inside than it is connected to $\bar{A}$.

(This assumption can be weakened by replacing the definition of $\texttt{Conn}(A)$ *with* $\texttt{Conn}(A) \stackrel{\text{def}}{=} \frac{1}{\tau_{\mathrm{mix}}(A)}$*, where* $\tau_{\mathrm{mix}}(A)$ *is the* mixing time for the relative pointwise distance *in* $G[A]$*; or less weakly* $\texttt{Conn}(A) \stackrel{\text{def}}{=} \frac{\lambda(A)}{\log \mathrm{vol}(A)}$ *where* $\lambda(A)$ *is the* spectral gap*, i.e.,* 1 *minus the second largest eigenvalue of the random walk matrix on* $G[A]$*. We discuss them in the full version of this paper.)*

Input parameters. Similar to prior work on local clustering, we assume the algorithm takes as input:

- *Some "good" starting vertex* $v \in A$*, and an oracle to output the set of neighbors for any given vertex.* This requirement is essential because without such an oracle the algorithm may have to read all inputs and cannot be sublinear in time; and without a starting vertex the sublinear-time algorithm may be unable to even find an element in A.
  We also need v to be "good", as for instance the vertices on the boundary of A may not be helpful enough in finding good clusters. We call the set of good vertices $A^g \subseteq A$, and a local algorithm needs to ensure that A^g is large, i.e., $\mathrm{vol}(A^g) \geq \frac{1}{2}\mathrm{vol}(A)$.[4]
- *The value of* Φ*.*
  In practice Φ can be viewed as a parameter and can be tuned for specific data. This is in contrast to the value of Ψ that is the target cut conductance and does not need to be known by the algorithm.[5]
- *A value* vol_0 *satisfying* $\mathrm{vol}(A) \in [\mathrm{vol}_0, 2\mathrm{vol}_0]$.[6]

2.2. PageRank Random Walk

We use the convention of writing vectors as row vectors in this paper. Let A be the adjacency matrix of G, and let D be the diagonal matrix with $D_{ii} = \deg(i)$, then the *lazy random walk matrix* $W \stackrel{\text{def}}{=} \frac{1}{2}(I + D^{-1}A)$. Accordingly, the PageRank vector $pr_{s,\alpha}$, is defined to be the unique solution of the following linear equation (cf. (Andersen et al., 2006)):

$$pr_{s,\alpha} = \alpha s + (1 - \alpha) pr_{s,\alpha} W \ ,$$

where $\alpha \in (0, 1]$ is the *teleport probability* and s is a *starting vector*. Here s is usually a probability vector:

[4] This assumption is unavoidable in all local clustering work. One can replace this $\frac{1}{2}$ by any other constant at the expense of worsening the guarantees by a constant factor.

[5] In prior work when Ψ is the only quantity studied, Ψ plays both roles as a tuning parameter and as a target.

[6] This requirement is optional since otherwise the algorithm can try out different powers of 2 and pick the smallest one with a valid output. It blows up the running time only by a constant factor for local algorithms, since the running time of the last trial dominates.

its entries are in $[0,1]$ and sum up to 1. For technical reasons we may use an arbitrary (and possibly negative) vector s inside the proof. When it is clear from the context, we drop α in the subscript for cleanness.

Given a vertex $u \in V$, let $\chi_u \in \{0,1\}^V$ be the indicator vector that is 1 only at vertex u. Given non-empty subset $S \subseteq V$ we denote by π_S the degree-normalized uniform distribution on S, that is, $\pi_S(u) = \frac{\deg(u)}{\mathrm{vol}(S)}$ when $u \in S$ and 0 otherwise. Very often we study a PageRank vector when $s = \chi_v$ is an indicator vector, and if so we abbreviate pr_{χ_v} by pr_v.

One equivalent way to study pr_s is to imagine the following random procedure: first pick a non-negative integer $t \in \mathbb{Z}_{\geq 0}$ with probability $\alpha(1-\alpha)^t$, then perform a lazy random walk starting at vector s with exactly t steps, and at last define pr_s to be the vector describing the probability of reaching each vertex in this random procedure. In its mathematical formula we have (cf. (Haveliwala, 2002; Andersen et al., 2006)):

Proposition 2.1. $pr_s = \alpha s + \alpha \sum_{t=1}^{\infty} (1-\alpha)^t (sW^t)$.

This implies that pr_s is linear: $a \cdot pr_s + b \cdot pr_t = pr_{as+bt}$.

2.3. Approximate PageRank Vector

In the seminal work of (Andersen et al., 2006), they defined approximate PageRank vectors and designed an algorithm to compute them efficiently.

Definition 2.2. *An ε-approximate PageRank vector p for pr_s is a nonnegative PageRank vector $p = pr_{s-r}$ where the vector r is nonnegative and satisfies $r(u) \leq \varepsilon \deg(u)$ for all $u \in V$.*

Proposition 2.3. *For any starting vector s with $\|s\|_1 \leq 1$ and $\varepsilon \in (0,1]$, one can compute an ε-approximate PageRank vector $p = pr_{s-r}$ for some r in time $O\left(\frac{1}{\varepsilon\alpha}\right)$, with $\mathrm{vol}(\mathrm{supp}(p)) \leq \frac{2}{(1-\alpha)\varepsilon}$.*

For completeness we provide the algorithm and its proof in the full version. It can be verified that:

$$\forall u \in V, \quad pr_s(u) \geq p(u) \geq pr_s(u) - \varepsilon \deg(u) \ . \quad (2.1)$$

2.4. Sweep Cuts

Given any approximate PageRank vector p, the *sweep cut (or threshold cut)* technique is the one to sort all vertices according to their degree-normalized probabilities $\frac{p(u)}{\deg(u)}$, and then study only those cuts that separate high-value vertices from low-value vertices. More specifically, let $v_1, v_2, \ldots, v_n$ be the decreasing order over all vertices with respect to $\frac{p(u)}{\deg(u)}$. Then, define *sweep sets* $S_j^p \stackrel{\text{def}}{=} \{v_1, \ldots, v_j\}$ for each $j \in [n]$, and sweep cuts are the corresponding cuts $(S_j^p, \overline{S_j^p})$. Usually given a vector p, one looks for the best cut:

$$\min_{j \in [n-1]} \phi_{\mathsf{c}}(S_j^p) \ .$$

In almost all the cases, one only needs to enumerate j over $p(v_j) > 0$, so the above sweep cut procedure runs in time $O\big(\mathrm{vol}(\mathrm{supp}(p)) + |\mathrm{supp}(p)| \cdot \log |\mathrm{supp}(p)|\big)$. This running time is dominated by the time to compute p (see Proposition 2.3), so it is negligible.

2.5. Lovász-Simonovits Curve

Our proof requires the technique of *Lovász-Simonovits Curve* that has been more or less used in all local clustering algorithms so far. This technique was originally introduced by Lovász and Simonovits (1990; 1993) to study the mixing rate of Markov chains. In our language, from a probability vector p on vertices, one can introduce a function $p[x]$ on real number $x \in [0, 2m]$. This function $p[x]$ is piecewise linear, and is characterized by all of its end points as follows (letting $p(S) \stackrel{\text{def}}{=} \sum_{a \in S} p(a)$):

$$p[0] \stackrel{\text{def}}{=} 0, \quad p[\mathrm{vol}(S_j^p)] \stackrel{\text{def}}{=} p(S_j^p) \text{ for each } j \in [n] \ .$$

In other words, for any $x \in [\mathrm{vol}(S_j^p), \mathrm{vol}(S_{j+1}^p)]$,

$$p[x] \stackrel{\text{def}}{=} p(S_j^p) + \frac{x - \mathrm{vol}(S_j^p)}{\deg(v_{j+1})} p(v_{j+1}) \ .$$

Note that $p[x]$ is increasing and concave.

3. Guarantee Better Accuracy

In this section, we study PageRank random walks that start at a vertex $v \in A$ with teleport probability α. We claim the range of interesting α is $\left[\Omega(\Psi), O(\frac{\Phi^2}{\log n})\right]$. This is because, at a high level, when $\alpha \ll \Psi$ the random walk will leak too much to $\bar{A}$; while when $\alpha \gg \frac{\Phi^2}{\log n}$ the random walk will not mix well inside A. In prior work, α is chosen to be $\Theta(\Psi)$, and we will instead choose $\alpha = \Theta(\frac{\Phi^2}{\log n}) = \Theta(\Psi \cdot \mathtt{Gap})$. Intuitively, this choice of α ensures that under the condition the random walk mixes inside, it makes the walk leak as little as possible to $\bar{A}$. We prove the above intuition rigorously in this section. Specifically, we first show some properties on the exact PageRank vector in Section 3.1, and then move to the approximate vector in Section 3.2. This essentially proves the first two properties of Theorem 1.

3.1. Properties on the Exact Vector

We first introduce a new notation $\widetilde{pr}_s$, that is the PageRank vector (with teleport probability α) starting at vector s but walking on the subgraph $G[A]$.

Next, we choose the set of "good" starting vertices A^g to satisfy two properties: (1) the total probability of leakage is upper bounded by $\frac{2\Psi}{\alpha}$, and (2) pr_v is close to $\widetilde{pr}_v$ for vertices in A. Note that the latter implies that pr_v mixes well inside A as long as $\widetilde{pr}_v$ does so.

Lemma 3.1. *There exists a set $A^g \subseteq A$ with volume $\mathrm{vol}(A^g) \geq \frac{1}{2}\mathrm{vol}(A)$ such that, for any vertex $v \in A^g$, in a PageRank vector with teleport probability α starting at v, we have:*

$$\sum_{u \notin A} pr_v(u) \leq \frac{2\Psi}{\alpha} \ . \tag{3.1}$$

In addition, there exists a non-negative leakage vector *$l \in [0,1]^V$ with norm $\|l\|_1 \leq \frac{2\Psi}{\alpha}$ satisfying*

$$\forall u \in A, \quad pr_v(u) \geq \widetilde{pr}_v(u) - \widetilde{pr}_l(u) \ . \tag{3.2}$$

(Details of the proof are in the full version.)

Proof sketch. The proof for the first property (3.1) is classical and can be found in (Andersen et al., 2006). The idea is to study an auxiliary PageRank random walk with teleport probability α starting at the degree-normalized uniform distribution π_A, and by simple computation, this random walk leaks to $\bar{A}$ with probability no more than Ψ/α. Then, using Markov bound, there exists $A^g \subseteq A$ with $\mathrm{vol}(A^g) \geq \frac{1}{2}\mathrm{vol}(A)$ such that for each starting vertex $v \in A^g$, this leakage is no more than $\frac{2\Psi}{\alpha}$. This implies (3.1) immediately.

The interesting part is (3.2). Note that pr_v can be viewed as the probability vector from the following random procedure: start from vertex v, then at each step with probability α let the walk stop, and with probability $(1-\alpha)$ follow the matrix W to go to one of its neighbors (or itself) and continue. Now, we divide this procedure into two rounds. In the first round, we run the same PageRank random walk but whenever the walk wants to use an outgoing edge from A to leak, we let it stop and temporarily "hold" this probability mass. We define l to be the non-negative vector where $l(u)$ denotes the amount of probability that we have "held" at vertex u. In the second round, we continue our random walk only from vector l. It is worth noting that l is non-zero only at boundary vertices in A.

Similarly, we divide the PageRank random walk for $\widetilde{pr}_v$ into two rounds. In the first round we hold exactly the same amount of probability $l(u)$ at boundary vertices u, and in the second round we start from l but continue this random walk only within $G[A]$. To bound the difference between pr_v and $\widetilde{pr}_v$, we note that they share the same procedure in the first round; while for the second round, the random procedure for pr_v starts at l and walks towards $V \setminus A$ (so in the worst case it may never come back to A again), while that for $\widetilde{pr}_v$ starts at l and walks only inside $G[A]$ so induces a probability vector $\widetilde{pr}_l$ on A. This gives (3.2).

At last, to see $\|l\|_1 \leq \frac{2\Psi}{\alpha}$, one just needs to verify that $l(u)$ is essentially the probability that the original PageRank random walk leaks from vertex u. Then, $\|l\|_1 \leq \frac{2\Psi}{\alpha}$ follows from the fact that the total amount of leakage is upper bounded by $\frac{2\Psi}{\alpha}$. □

As mentioned earlier, we want to use (3.2) to lower bound $pr_v(u)$ for vertices $u \in A$. We achieve this by first lower bounding $\widetilde{pr}_v$ which is the PageRank random walk on $G[A]$. Given a teleport probability α that is small compared to $\frac{\Phi^2}{\log \mathrm{vol}(A)}$, this random walk should mix well. We formally state it as the following lemma, and provide its proof in the the full version.

Lemma 3.2. *When $\alpha \leq O(\Psi \cdot \mathtt{Gap})$ we have that*

$$\forall u \in A, \quad \widetilde{pr}_v(u) \geq \frac{4}{5}\frac{\deg_A(u)}{\mathrm{vol}(A)} \ .$$

Here $\deg_A(u)$ is the degree of u on $G[A]$, but $\mathrm{vol}(A)$ is with respect to the original graph.

3.2. Properties of the Approximate Vector

From this section on we always use $\alpha \leq O(\Psi \cdot \mathtt{Gap})$. We then fix a starting vertex $v \in A^g$ and study an ε-approximate Pagerank vector for pr_v. We choose

$$\varepsilon = \frac{1}{10 \cdot \mathrm{vol}_0} \in \left[\frac{1}{20\mathrm{vol}(A)}, \frac{1}{10\mathrm{vol}(A)}\right] \ . \tag{3.3}$$

For notational simplicity, we denote by p this ε-approximation and recall from Section 2.3 that $p = pr_{\chi_v - r}$ where r is a non-negative vector with $0 \leq r(u) \leq \varepsilon \deg(u)$ for every $u \in V$. Recall from (2.1) that $pr_v(u) \geq p(u) \geq pr_v(u) - \epsilon \cdot \deg(u)$ for all $u \in V$.

We now rewrite Lemma 3.1 in the language of approximate PageRank vectors using Lemma 3.2:

Corollary 3.3. *For any $v \in A^g$ and $\alpha \leq O(\Psi \cdot \mathtt{Gap})$, in an ε-approximate PageRank vector to pr_v denoted by $p = pr_{\chi_v - r}$, we have:*

$$\sum_{u \notin A} p(u) \leq \frac{2\Psi}{\alpha} \ \text{ and } \ \sum_{u \notin A} r(u) \leq \frac{2\Psi}{\alpha} \ .$$

In addition, there exists a non-negative leakage vector *$l \in [0,1]^V$ with norm $\|l\|_1 \leq \frac{2\Psi}{\alpha}$ satisfying*

$$\forall u \in A, \quad p(u) \geq \frac{4}{5}\frac{\deg_A(u)}{\mathrm{vol}(A)} - \frac{\deg(u)}{10\mathrm{vol}(A)} - \widetilde{pr}_l(u) \ .$$

Proof. The only inequality that requires a proof is $\sum_{u \notin A} r(u) \leq \frac{2\Psi}{\alpha}$. In fact, if one takes a closer look at the algorithm to compute an approximate Pagerank vector (see the full version), the total probability mass that will be sent to r on vertices outside A, is upper bounded by the probability of leakage. However, the latter is upper bounded by $\frac{2\Psi}{\alpha}$ when we choose A^g. □

We are now ready to state the main lemma of this section. We show that for all reasonable sweep sets S on this probability vector p, it satisfies that $\mathrm{vol}(S \setminus A)$ and $\mathrm{vol}(A \setminus S)$ are both at most $O\big(\frac{\Psi}{\alpha}\mathrm{vol}(A)\big)$.

Lemma 3.4. *In the same definition of α and p from Corollary 3.3, let sweep set $S_c \stackrel{\text{def}}{=} \{u \in V : p(u) \geq c\frac{\deg(u)}{\text{vol}(A)}\}$ for any constant $c < \frac{3}{5}$, then we have the following guarantees on the size of $S_c \setminus A$ and $A \setminus S_c$:*

1. $\text{vol}(S_c \setminus A) \leq \frac{2\Psi}{\alpha c}\text{vol}(A)$, *and*
2. $\text{vol}(A \setminus S_c) \leq \left(\frac{2\Psi}{\alpha(\frac{3}{5}-c)} + 8\Psi\right)\text{vol}(A)$.

Proof. First we notice that $p(S_c \setminus A) \leq p(V \setminus A) \leq \frac{2\Psi}{\alpha}$ owing to Corollary 3.3, and for each vertex $u \in S_c \setminus A$ it must satisfy $p(u) \geq c\frac{\deg(u)}{\text{vol}(A)}$. Those combined imply $\text{vol}(S_c \setminus A) \leq \frac{2\Psi}{\alpha c}\text{vol}(A)$ proving the first property.

We show the second property in two steps. First, let A_b be the set of vertices in A such that $\frac{4}{5}\frac{\deg_A(u)}{\text{vol}(A)} - \frac{\deg(u)}{10\text{vol}(A)} < \frac{3}{5}\frac{\deg(u)}{\text{vol}(A)}$. Any such vertex $u \in A_b$ must have $\deg_A(u) < \frac{7}{8}\deg(u)$. This implies that u has to be on the boundary of A and $\text{vol}(A_b) \leq 8\Psi\text{vol}(A)$.

Next, for a vertex $u \in A \setminus A_b$ we have (using Corollary 3.3 again) $p(u) \geq \frac{3}{5}\frac{\deg(u)}{\text{vol}(A)} - \widetilde{pr}_l(u)$. If we further have $u \notin S_c$ so $p(u) < c\frac{\deg(u)}{\text{vol}(A)}$, it implies that $\widetilde{pr}_l(u) \geq (\frac{3}{5} - c)\frac{\deg(u)}{\text{vol}(A)}$. As a consequence, the total volume for such vertices (i.e., $\text{vol}(A \setminus (A_b \cup S_c))$) cannot exceed $\frac{\|\widetilde{pr}_l\|_1}{3/5-c}\text{vol}(A)$. At last, we notice that $\widetilde{pr}_l$ is a non-negative probability vector coming from a random walk procedure, so $\|\widetilde{pr}_l\|_1 = \|l\|_1 \leq \frac{2\Psi}{\alpha}$. This in sum provides that

$$\begin{aligned}\text{vol}(A \setminus S_c) &\leq \text{vol}(A \setminus (A_b \cup S_c)) + \text{vol}(A_b)\\ &\leq \left(\frac{2\Psi}{\alpha(\frac{3}{5}-c)} + 8\Psi\right)\text{vol}(A)\ .\end{aligned}$$ □

Note that if one chooses $\alpha = \Theta(\Psi \cdot \texttt{Gap})$ in the above lemma, both those two volumes are at most $O(\text{vol}(A)/\texttt{Gap})$ satisfying the first two properties of Theorem 1.

4. Guarantee Better Cut Conductance

In the classical work of (Andersen et al., 2006), they have shown that when $\alpha = \Theta(\Psi)$, among all sweep cuts on vector p there exists one with cut conductance $O(\sqrt{\Psi \log n})$. In this section, we improve this result under our gap assumption $\texttt{Gap} \geq \Omega(1)$.

Lemma 4.1. *Letting $\alpha = \Theta(\Psi \cdot \texttt{Gap})$, among all sweep sets $S_c = \{u \in V : p(u) \geq c\frac{\deg(u)}{\text{vol}(A)}\}$ for $c \in [\frac{1}{8}, \frac{1}{4}]$, there exists one, denoted by S_{c^*}, with cut conductance $\phi_{\texttt{c}}(S_{c^*}) = O(\sqrt{\Psi/\texttt{Gap}})$.*

Proof sketch. To convey the idea of the proof, we only consider the case when $p = pr_v$ is the exact PageRank vector, and the proof for the approximate case is a bit more involved and deferred to the full version.

Suppose that all sweep sets S_c for $c \in [\frac{1}{8}, \frac{1}{4}]$ satisfy $|E(S_c, V \setminus S_c)| \geq E_0$ for some value E_0, then it suffices to prove $E_0 \leq O(\frac{\Psi}{\sqrt{\alpha}})\text{vol}(A)$. This is because, if so, then there exists some S_{c^*} with $|E(S_{c^*}, V \setminus S_{c^*})| \leq E_0$ and this combined with the result in Lemma 3.4 (i.e., $\text{vol}(S_{c^*}) = (1 \pm O(1/\texttt{Gap}))\text{vol}(A)$) gives

$$\phi_{\texttt{c}}(S_{c^*}) \leq O\Big(\frac{E_0}{\text{vol}(S_{c^*})}\Big) = O(\Psi/\sqrt{\alpha}) = O(\sqrt{\Psi/\texttt{Gap}})\ .$$

We introduce some classical notations before we proceed in the proof. For any vector q we denote by $q(S) \stackrel{\text{def}}{=} \sum_{u \in S} q(u)$. Also, given a directed edge[7], $e = (a, b) \in E$ we let $p(e) = p(a, b) \stackrel{\text{def}}{=} \frac{p(a)}{\deg(a)}$, and for a set of directed edges E' we let $p(E') \stackrel{\text{def}}{=} \sum_{e \in E'} p(e)$. We also let $E(A, B) \stackrel{\text{def}}{=} \{(a, b) \in E \mid a \in A \wedge b \in B\}$ be the set of directed edges from A to B.

Now for any set $S_{1/4} \subseteq S \subseteq S_{1/8}$, we compute that

$$\begin{aligned} p(S) = pr_v(S) &= \alpha\chi_v(S) + (1-\alpha)(pW)(S)\\ &\leq \alpha + (1-\alpha)(pW)(S)\\ \implies (1-\alpha)p(S) &\leq \alpha(1-p(S)) + (1-\alpha)(pW)(S)\\ \implies (1-\alpha)p(S) &\leq 2\Psi + (1-\alpha)(pW)(S)\\ \implies p(S) &< O(\Psi) + (pW)(S)\ . \end{aligned} \tag{4.1}$$

Here we have used the fact that when $p = pr_v$ is exact, it satisfies $1 - p(S) = p(V - S) \leq 2\Psi/\alpha$ according to Corollary 3.3. In the next step, we use the definition of the lazy random walk matrix W to compute that

$$\begin{aligned}&(pW)(S)\\ =&\Big(\sum_{(a,b)\in E(S,S)} p(a,b) + \sum_{(a,b)\in E(S,\bar{S})} \frac{p(a,b)+p(b,a)}{2}\Big)\\ =&\Big(\frac{1}{2}p\Big(E(S,S)\Big) + \frac{1}{2}p\Big(E(S,S)\cup E(S,\bar{S})\cup E(\bar{S},S)\Big)\Big)\\ \leq&\Big(\frac{1}{2}p\Big[\big|E(S,S)\big|\Big] + \frac{1}{2}p\Big[\big|E(S,S)\cup E(S,\bar{S})\cup E(\bar{S},S)\big|\Big]\Big)\\ =&\Big(\frac{1}{2}p\Big[\text{vol}(S) - \big|E(S,\bar{S})\big|\Big] + \frac{1}{2}p\Big[\text{vol}(S) + \big|E(S,\bar{S})\big|\Big]\Big)\\ \leq&\Big(\frac{1}{2}p\big[\text{vol}(S) - E_0\big] + \frac{1}{2}p\big[\text{vol}(S) + E_0\big]\Big)\ . \end{aligned} \tag{4.2}$$

Here the first inequality is due to the definition of the Lovász-Simonovits curve $p[x]$, and the second inequality is because $p[x]$ is concave. Next, suppose that in addition to $S_{1/4} \subseteq S \subseteq S_{1/8}$, we also know that S is a sweep set, i.e., $\forall a \in S, b \notin S$ we have $\frac{p(a)}{\deg(a)} \geq \frac{p(b)}{\deg(b)}$. This implies $p(S) = p[\text{vol}(S)]$ and combining (4.1) and (4.2) we obtain that

$$\begin{aligned}&\big(p[\text{vol}(S)] - p\big[\text{vol}(S) - E_0\big]\big)\\ \leq\ &O(\Psi) + \big(p\big[\text{vol}(S) + E_0\big] - p[\text{vol}(S)]\big)\ .\end{aligned}$$

[7] G is an undirected graph, but we study undirected edges with specific directions for analysis purpose only.

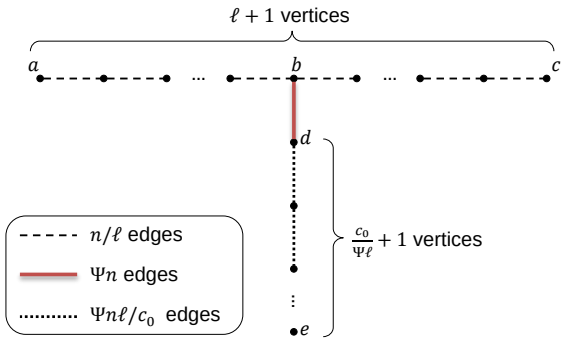

Figure 1. Our hard instance for proving tightness. One can pick for instance $\ell \approx n^{0.4}$ and $\Psi \approx \frac{1}{n^{0.9}}$, so that $n/\ell \approx n^{0.6}$, $\Psi n \approx n^{0.1}$ and $\Psi n\ell \approx n^{0.5}$.

Since we can choose S to be an arbitrary sweep set between $S_{1/4}$ and $S_{1/8}$, we have that the inequality $p[x] - p[x - E_0] \le O(\Psi) + p[x + E_0] - p[x]$ holds for all end points $x \in [\mathrm{vol}(S_{1/4}), \mathrm{vol}(S_{1/8})]$ on the piecewise linear curve $p[x]$. This implies that the same inequality holds for any real number $x \in [\mathrm{vol}(S_{1/4}), \mathrm{vol}(S_{1/8})]$ as well. We are now ready to draw our conclusion by repeatedly applying this inequality. Letting $x_1 := \mathrm{vol}(S_{1/4})$ and $x_2 := \mathrm{vol}(S_{1/8})$, we have

$$\begin{aligned}
\frac{E_0}{4\mathrm{vol}(A)} &\le p[x_1] - p[x_1 - E_0] \\
&\le O(\Psi) + (p[x_1 + E_0] - p[x_1]) \\
&\le 2 \cdot O(\Psi) + (p[x_1 + 2E_0] - p[x_1 + E_0]) \le \cdots \\
&\le \left\lfloor \frac{x_2 - x_1}{E_0} + 1 \right\rfloor O(\Psi) + (p[x_2 + E_0] - p[x_2]) \\
&\le \frac{\mathrm{vol}(S_{1/8} \setminus S_{1/4})}{E_0} O(\Psi) + \frac{E_0}{8\mathrm{vol}(A)} \\
&\le \frac{\mathrm{vol}(S_{1/8} \setminus A) + \mathrm{vol}(A \setminus S_{1/4})}{E_0} O(\Psi) + \frac{E_0}{8\mathrm{vol}(A)} \\
&\le \frac{O(\Psi/\alpha) \cdot \mathrm{vol}(A)}{E_0} O(\Psi) + \frac{E_0}{8\mathrm{vol}(A)} \ ,
\end{aligned}$$

where the first inequality uses the definition of $S_{1/4}$, the fifth inequality uses the definition of $S_{1/8}$, and last inequality uses Lemma 3.4 again. After re-arranging the above inequality we conclude that $E_0 \le O\big(\frac{\Psi}{\sqrt{\alpha}}\big)\mathrm{vol}(A)$ and finish the proof. □

The lemma above essentially shows the third property of Theorem 1 and finishes the proof of Theorem 1. For completeness of the paper, we still provide the formal proof for Theorem 1 in the full version, and summarize our final algorithm in Algorithm 1.

5. Tightness of Our Analysis

It is a natural question to ask under our newly introduced assumption $\texttt{Gap} \ge \Omega(1)$: is $O(\sqrt{\Psi/\texttt{Gap}})$ the best cut conductance we can obtain from a local algorithm? We show that this is true if one sticks to a sweep-cut algorithm using PageRank vectors.

Algorithm 1 `PageRank-Nibble`

Input: v, Φ and $\mathrm{vol}_0 \in [\frac{\mathrm{vol}(A)}{2}, \mathrm{vol}(A)]$.
Output: set S.

1: $\alpha \leftarrow \Theta(\frac{\Phi^2}{\log \mathrm{vol}(A))}) = \Theta(\Psi \cdot \texttt{Gap})$.
2: $p \leftarrow$ a $\frac{1}{10 \cdot \mathrm{vol}_0}$-approximate PageRank vector with starting vertex v and teleport probability α.
3: Sort all vertices in $\mathrm{supp}(p)$ according to $\frac{p(u)}{\deg(u)}$.
4: Consider all sweep sets $S'_c \stackrel{\text{def}}{=} \{u \in \mathrm{supp}(p) : p(u) \ge \frac{c \deg(u)}{\mathrm{vol}_0}\}$ for $c \in [\frac{1}{8}, \frac{1}{2}]$, and let S be the one among them with the best $\phi_{\mathsf{c}}(S)$.

More specifically, we show that our analysis in Section 4 is tight by constructing the following hard instance. Consider a (multi-)graph with two chains (see Figure 1) of vertices, and there are multi-edges connecting them.[8] In particular:

- the top chain (ended with vertex a and c and with midpoint b) consists of $\ell + 1$ vertices where ℓ is even with $\frac{n}{\ell}$ edges between each consecutive pair;
- the bottom chain (ended with vertex d and e) consists of $\frac{c_0}{\Psi\ell} + 1$ vertices with $\frac{\Psi n\ell}{c_0}$ edges between each consecutive pair, where the constant c_0 is to be determined later; and
- vertex b and d are connected with Ψn edges.

We let the top chain to be our promised cluster A. The total volume of A is $2n + \Psi n$, while the total volume of the entire graph is $4n + 2\Psi n$. The mixing time for A is $\tau_{\mathrm{mix}}(A) = \Theta(\ell^2)$, and the cut conductance $\phi_{\mathsf{c}}(A) = \frac{\Psi n}{\mathrm{vol}(A)} \approx \frac{\Psi}{2}$. Suppose that the gap assumption[9] $\texttt{Gap} \stackrel{\text{def}}{=} \frac{1}{\tau_{\mathrm{mix}}(A) \cdot \phi_{\mathsf{c}}(A)} \approx \frac{1}{\Psi\ell^2} \gg 1$ is satisfied, i.e., $\Psi\ell^2 = o(1)$. (For instance one can let $\ell \approx n^{0.4}$ and $\Psi \approx \frac{1}{n^{0.9}}$ to achieve this requirement.)

We then consider a PageRank random walk that starts at vertex $v = a$ and with teleport probability $\alpha = \frac{\gamma}{\ell^2}$ for some arbitrarily small constant $\gamma > 0$.[10] Let pr_a be this PageRank vector, and we prove in the full version the following lemma:

Lemma 5.1. *For any $\gamma \in (0, 4]$ and letting $\alpha = \gamma/\ell^2$, there exists some constant c_0 such that when studying the PageRank vector pr_a starting from vertex a in Figure 1, the following holds $\frac{pr_a(d)}{\deg(d)} > \frac{pr_a(c)}{\deg(c)}$.*

[8] One can transform this example into a graph without parallel edges by splitting vertices into expanders, but that goes out of the purpose of this section.

[9] We are using Theorem 1 in the language of gap assumption on τ_{mix}. See Section 2.1 for details.

[10] Although we promised in Theorem 2 to study all starting vertices $v \in A$, in this version of the paper we only concentrate on $v = a$ because other choices of v are only easier and can be analyzed similarly. In addition, this choice of $\alpha = \frac{\gamma}{\ell^2}$ is consistent with the one used Theorem 1.

Digit	0	1	2	3	4	5	6	7	8	9
$\Psi = \phi_{\mathsf{c}}(A)$	0.00294	0.00304	0.08518	0.03316	0.22536	0.08580	0.01153	0.03258	0.09761	0.05139
$\phi_{\mathsf{c}}(S)$	0.00272	0.00067	0.03617	0.02220	0.00443	0.01351	0.00276	0.00456	0.03849	0.00448
Precision	0.993	0.995	0.839	0.993	0.988	0.933	0.946	0.985	0.941	0.994
Recall	0.988	0.988	0.995	0.773	0.732	0.896	0.997	0.805	0.819	0.705

Table 1. Clustering results on the USPS zipcode data set. We report precision $|A \cap S|/|S|$ and recall $|A \cap S|/|A|$.

This lemma implies that, for any sweep-cut algorithm based on this vector pr_a, even if it computes pr_a exactly and looks for all possible sweep cuts, then none of them gives a better cut conductance than $O(\sqrt{\Psi/\texttt{Gap}})$. More specifically, for any sweep set S:

- if $c \notin S$, then $|E(S, V \setminus S)|$ is at least $\frac{n}{\ell}$ because it has to contain a (multi-)edge in the top chain. Therefore, the cut conductance $\phi_{\mathsf{c}}(S) \geq \Omega(\frac{n}{\ell \mathrm{vol}(S)}) \geq \Omega(\frac{1}{\ell}) \geq \Omega(\sqrt{\Psi/\texttt{Gap}})$; or
- if $c \in S$, then d must be also in S because it has a higher normalized probability than c using Lemma 5.1. In this case, $|E(S, V \setminus S)|$ is at least $\frac{\Psi n \ell}{c_0}$ because it has to contain a (multi-)edge in the bottom chain. Therefore, the cut conductance $\phi_{\mathsf{c}}(S) \geq \Omega(\frac{\Psi n \ell}{\mathrm{vol}(S)}) \geq \Omega(\Psi \ell) = \Omega(\sqrt{\Psi/\texttt{Gap}})$.

This ends the proof of Theorem 2. □

6. Empirical Evaluation

The PageRank local clustering method has been studied empirically in various previous work. For instance, Gleich and Seshadhri (2012) performed experiments on 15 datasets and confirmed that PageRank outperformed many others in terms of cut conductance, including the famous Metis algorithm. Moreover, (Leskovec et al., 2009) studied PageRank against Metis+MQI which is the Metis algorithm plus a flow-based post-processing. Their experiments confirmed that although Metis+MQI outperforms PageRank in terms of cut conductance, however, the PageRank algorithm's outputs are more "community-like", and they enjoy other desirable properties.

Since our `PageRank-Nibble` is essentially the same PageRank method as before with only theoretical changes in the parameters, it certainly embraces the same empirical behavior as those literatures above. Therefore, in this section we perform experiments *only for the sake of* demonstrating our theoretical discoveries in Theorem 1, without comparisons to other methods. We run our algorithm against both synthetic and real datasets, and due to the page limit, we defer the details of our experiment setups to the full version.

Recall that Theorem 1 has three properties. The first two properties are *accuracy guarantees* that ensure the output set S well approximates A in terms of volume; and the third property is a *cut-conductance guarantee* that ensures the output set S has small $\phi_{\mathsf{c}}(S)$. We now provide experimental results to support them.

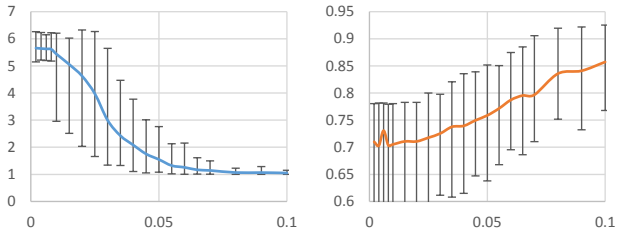

Figure 2. Experimental result on the synthetic data. The horizontal axis represents the value of β for constructing our graph, the blue curve (left) represents the ratio $\frac{\phi_{\mathsf{c}}(S)}{\Psi}$, and the red curve (right) represents the clustering accuracy. The vertical bars are 94% confidence intervals for 100 runs.

In the first experiment, we study a synthetic random graph of 870 vertices. Our desired cluster A is constructed from the Watts-Strogatz random model with a parameter $\beta \in [0, 1]$ to control the connectivity of $G[A]$: the larger β is the larger `Gap` is. We therefore present in Figure 2 our experimental results as two curves, both in terms of β: the cut conductance over Ψ ratio, i.e., $\frac{\phi_{\mathsf{c}}(S)}{\Psi}$, and the clustering accuracy, i.e., $1 - \frac{|A \Delta S|}{|V|}$. Our experiment confirms our result in Theorem 1: `PageRank-Nibble` performs better both in accuracy and cut conductance as `Gap` goes larger.

In the second experiment, we use the USPS zipcode dataset that was also used in the work from (Wu et al., 2012). This dataset has 9298 images of handwritten digits between 0 to 9, and we treat them as 10 separate binary-classification problems. We report our results in Table 1. For each of the 10 binary-classifications, we have a ground-truth cluster A that contains all data points associated with the given digit. We then compare the cut conductance of our output set $\phi_{\mathsf{c}}(S)$ against the desired cut conductance $\Psi = \phi_{\mathsf{c}}(A)$, and our algorithm consistently outperforms the desired one on all 10 clusters. (Notice that it is possible to see an output set S to have smaller conductance than A, because A is not necessarily the sparest cut in the graph.) In addition, one can also confirm from Table 1 that our algorithm enjoys high precision and recall.

Acknowledgments. We thank Lorenzo Orecchia, Jon Kelner, Aditya Bhaskara for helpful conversations.

References

Alamgir, Morteza and von Luxburg, Ulrike. Multi-agent random walks for local clustering on graphs. ICDM '10, pp. 18–27, 2010.

Alon, Noga. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

Alvisi, L., Clement, A., Epasto, A., Lattanzi, S., and Panconesi, A. The evolution of sybil defense via social networks. In *IEEE Symposium on Security and Privacy*, 2013.

Andersen, Reid and Lang, Kevin J. Communities from seed sets. WWW '06, pp. 223–232, 2006.

Andersen, Reid and Peres, Yuval. Finding sparse cuts locally using evolving sets. STOC, 2009.

Andersen, Reid, Chung, Fan, and Lang, Kevin. Using pagerank to locally partition a graph. 2006. An extended abstract appeared in FOCS '2006.

Andersen, Reid, Gleich, David F., and Mirrokni, Vahab. Overlapping clusters for distributed computation. WSDM '12, pp. 273–282, 2012.

Arora, Sanjeev and Kale, Satyen. A combinatorial, primal-dual approach to semidefinite programs. STOC '07, pp. 227–236, 2007.

Arora, Sanjeev, Rao, Satish, and Vazirani, Umesh V. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM*, 56(2), 2009.

Arora, Sanjeev, Hazan, Elad, and Kale, Satyen. O(sqrt(log(n)) approximation to sparsest cut in $\tilde{o}(n^2)$ time. *SIAM Journal on Computing*, 39(5): 1748–1771, 2010.

Chawla, Shuchi, Krauthgamer, Robert, Kumar, Ravi, Rabani, Yuval, and Sivakumar, D. On the hardness of approximating multicut and sparsest-cut. *Computational Complexity*, 15(2):94–114, June 2006.

Gargi, Ullas, Lu, Wenjun, Mirrokni, Vahab S., and Yoon, Sangho. Large-scale community detection on youtube for topic discovery and exploration. In *AAAI Conference on Weblogs and Social Media*, 2011.

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

Gleich, David F. and Seshadhri, C. Vertex neighborhoods, low conductance cuts, and good seeds for local community methods. In *KDD '2012*, 2012.

Haveliwala, Taher H. Topic-sensitive pagerank. In *WWW '02*, pp. 517–526, 2002.

Kannan, Ravi, Vempala, Santosh, and Vetta, Adrian. On clusterings: Good, bad and spectral. *Journal of the ACM*, 51(3):497–515, 2004.

Leighton, Frank Thomson and Rao, Satish. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM*, 46(6):787–832, 1999.

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

Leskovec, Jure, Lang, Kevin J., and Mahoney, Michael. Empirical comparison of algorithms for network community detection. WWW, 2010.

Lin, Frank and Cohen, William W. Power iteration clustering. In *ICML '10*, pp. 655–662, 2010.

Lovász, László and Simonovits, Miklós. The mixing rate of markov chains, an isoperimetric inequality, and computing the volume. FOCS, 1990.

Lovász, László and Simonovits, Miklós. Random walks in a convex body and an improved volume algorithm. *Random Struct. Algorithms*, 4(4):359–412, 1993.

Makarychev, Konstantin, Makarychev, Yury, and Vijayaraghavan, Aravindan. Approximation algorithms for semi-random partitioning problems. In *STOC '12*, pp. 367–384, 2012.

Merca, Mircea. A note on cosine power sums. *Journal of Integer Sequences*, 15:12.5.3, May 2012.

Morris, Ben and Peres, Yuval. Evolving sets and mixing. STOC '03, pp. 279–286. ACM, 2003.

Motwani, Rajeev and Raghavan, Prabhakar. *Randomized algorithms*. Cambridge University Press, 1995.

Schaeffer, S. E. Graph clustering. *Computer Science Review,*, 1(1):27–64, 2007.

Shalev-Shwartz, Shai and Srebro, Nathan. SVM optimization: inverse dependence on training set size. In *ICML*, 2008.

Sherman, Jonah. Breaking the multicommodity flow barrier for $o(\sqrt{\log n})$-approximations to sparsest cut. FOCS '09, pp. 363–372, 2009.

Shi, J. and Malik, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

Sinclair, Alistair and Jerrum, Mark. Approximate counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82 (1):93–133, 1989.

Spielman, Daniel and Teng, Shang-Hua. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. STOC, 2004.

Spielman, Daniel and Teng, Shang-Hua. A local clustering algorithm for massive graphs and its application to nearly-linear time graph partitioning. *CoRR*, abs/0809.3232, 2008.

Wu, Xiao-Ming, Li, Zhenguo, So, Anthony Man-Cho, Wright, John, and Chang, Shih-Fu. Learning with partially absorbing random walks. In *NIPS*, 2012.

Zhu, Zeyuan Allen, Chen, Weizhu, Zhu, Chenguang, Wang, Gang, Wang, Haixun, and Chen, Zheng. Inverse time dependency in convex regularized learning. ICDM, 2009.