LOCAL CHEEGER INEQUALITIES AND SPARSE CUTS

CAELAN GARRETT
18.434 FINAL PROJECT
CAELAN@MIT.EDU

Abstract. When computing on massive graphs, it is not tractable to iterate over the full graph. Local graph methods aim to analyze properties in a neighborhood of a seed vertex. Graph partitioning is an example of a problem that has a local analog. It attempts to find a sparse cut of a specific volume near a seed vertex. Previous work attempts to approximate the optimal partition through standard, PageRank, and Heat Kernel PageRank random walks. We review these methods and focus on a recent improved approximation result from the Heat Kernel PageRank. Specifically, we show that the result was obtained via a false Local Cheeger Inequality. We present a counterexample, explain the flaw in the proof, obtain weaker, correct results, and cite implications on Hardness of Approximation if the algorithm is able to achieve its approximation ratio.

1. Introduction

The big data revolution and the emergence of massive information networks in everyday life motivate the analysis of properties and algorithms on large graphs. These networks are represented as graphs by treating each network connection as an edge between two vertices and studied with combinatorial and spectral tools. Early research looked at ranking vertices in the graph based on their influence for use in ranking Webpages for Web search engines. This Hub-and-Authority algorithm is an early way of producing a ranking by using an eigenvector to weigh importance. However, computing an eigenvector can be prohibitively expensive for large graphs because it requires iteration over the full graph structure. This brought about the need to study local graph algorithms that perform computations on a small subset of the graph around a seed vertex. Random walks prove useful for these computations because their running times are proportional to the volume they operate on. We will focus on local graph partitioning, the problem of finding local sparse cuts, and fast spectral algorithms that produce approximately optimal cuts through random walks.

Graph partitioning is classical problem that involves finding a partition or cut of the vertices of a graph into two sets that have large volumes without removing many edges. Formally, this is the process of minimizing the conductance of a selected cut. The optimal graph partitioning algorithm is NP-complete, but approximation algorithms have proven effective at efficiently finding an approximate solution. A spectral approximation algorithm finds partitions by ordering vertices by a ranking, usually the second eigenvector, and cut along the sorted vector to produce a partition composed of the high scoring vertices. Because it scans the vector of vertices only once, it is called a one-sweep algorithm. This is advantageous because it considers $O(n)$ potential cuts instead of an exponential number of

Date: December 1, 2014.

potential cuts. And the resulting cuts have good performance guarantees giving approximation ratios that are $O(\sqrt{\mathrm{OPT}})$ by using analysis from Cheeger Inequalities. Although the best approximation algorithm achieves an approximation ratio of $O(\sqrt{n})$ via semi-definite relaxation [9], one-sweep algorithms are still popular for their simplicity. But even an $O(n)$ runtime can be intractable for the large graphs we would like to process.

Local graph partitioning is variant of this problem tractable for our model. The goal of the local graph partitioning is to find a sparse cut of a specified size given a starting, seed vertex. Because these algorithms are limited to a specific volume, they only need to process a small subset of the full graph to find their partition. Spectral approximation algorithms operate similarly to their behavior in the classical graph partitioning problem by finding sweeps of ranking vectors except they compute these ranking vectors through random walks starting at the seed vertex. Although standard random walks are effective ways to do this, the PageRank random walk currently gives the best cut that is $O(\sqrt{\phi \log s})$ where ϕ is the target conductance and s is the desired cut size [6]. The performance guarantees of these algorithms are given by Local Cheeger Inequalities.

Recently, an algorithm that uses a PageRank variant called Heat Kernel PageRank to perform the random walk showed the potential to produce cuts that are $O(\sqrt{\log s})$ better than previously known, removing the approximation dependence on the target cut volume and giving a constant factor approximation [2]. If this algorithm is correct, it has ramifications in Hardness of Approximation because it implies that the conjectured NP-hard small-set expansion problem can be solved efficiently and gives strong evidence against the Unique Games Conjecture, one of the major open problems in complexity theory.

However, the proof of its approximation ratio uses an incorrect Local Cheeger Inequality. We explore where the proof goes astray by first constructing a counterexample that exemplifies why the bound cannot be true both explicitly and intuitively. Next, we review the proof and point out where it fails. Finally, we explore recoverable weaker results from the proof that may still be useful to analyze the performance of Heat Kernel PageRank.

2. Background

2.1. **Definitions.** A graph $G = (V, E)$ is composed of a set of vertices V and edges E. An adjacency matrix A captures this structure with $A(u, v) = 1$ if $(u, v) \in E$ and $A(u, v) = 0$ otherwise. The degree of a vertex u is d_u, and D is a diagonal matrix such that $D(u, u) = d_u, u \in V$.

The random walk matrix W is defined for every pair of vertices (u, v) to be a uniform distribution of the outgoing edges from u and 0 if the pair is not an edge in the graph.

$$W(u, v) = \begin{cases} \frac{1}{d_u}, & \text{if} (u, v) \in E \\ 0 & \text{otherwise} \end{cases}$$

W can more compactly be represented by $W = D^{-1}A$. A random walk on W will converge assuming that it is connected and non-bipartite meaning it cannot be partitioned into two vertex sets. We assume that the all graphs considered obey these two properties. Additionally, the lazy random walk is defined as $W' = \frac{W+I}{2}$ where the difference is that there is equal probability of staying at the same vertex versus moving.

The volume of a set of vertices is defined to be the sum of the degrees of vertices in the set or $\text{vol}(S) = \sum_{u \in S} d_u$. Note that this sum double counts contributions from edges (u, v) such that $u \in S$ and $v \in S$. The edge boundary of S, $\partial(S)$, is defined to be the set of edges leaving S.

$$\partial(S) = \{(u, v) \in E : u \in S \text{ and } v \notin S\}$$

.

Given a partition S, let $\overline{S} = V \setminus S$ be the complement of S. These definitions are useful for representing the conductance (also called Cheeger ratio) of a set ϕ_S. This ratio captures the spread of some a random walk out of S. Namely, a larger conductance means better mixing. Correspondingly, a lower conductance represents a large set of vertices weakly connected to the rest of the graph constituting a sparse cut.

$$\phi_S = \frac{|\partial(S)|}{\min(\text{vol}(S), \text{vol}(\overline{S}))}$$

2.2. **Sparse Cuts.** The global conductance of a graph is minimum, bottleneck conductance $\phi_G = \min_{S \subseteq V} \phi_S$, and the local conductance of a set S is $\phi_S^* = \min_{T \subseteq S} \phi_T$. Note that ϕ_S does not necessarily equal ϕ_S^* because ϕ_S^* is computed from cuts of subsets. The classical graph partitioning optimization problem involves finding S that produces $\phi_S = \phi_G$, the minimum conductance in the graph. Because this problem is NP-complete, we are more interested in algorithms that find cuts with approximately minimal conductance [4]. As previously mentioned, a common way to do this is to compute a ranking of the vertices and perform a sweep of the ranking vector. Namely given an ranking function $f : V \to \mathbb{R}$, a sweep finds the minimum conductance of cuts computed by ordering the vertices by decreasing rank weighted by volume.

$$\frac{f(v_1)}{d_{v_1}} \geq \frac{f(v_2)}{d_{v_2}} \geq \ldots \geq \frac{f(v_n)}{d_{v_n}}$$

Each cut is then computed by choosing i to select the first $v_1, v_2, \ldots, v_i$ vertices to compose the set S_i. Using these cuts, we can compute the minimum conductance over the n possible cuts $\min_i \phi_{S_i}$. Because these algorithms give cuts that approximate the minimum conductance, we need tools to give bounds on the conductance of these cuts. These bounds are given by analyzing Cheeger Inequalities. Cheeger Inequalities are used to relate the conductance of the graph, eigenvalues, and convergence of random walks. The most famous of these bounds is called Cheeger's Inequality:

Theorem 1.

$$\frac{\phi_G^2}{2} \leq \nu_2 \leq 2\phi_G$$

which bounds the second smallest eigenvalue of a normalized Laplacian matrix N by the minimum conductance of the graph. When applied to random walk matrices, $N = I - D^{-1/2}W'D^{1/2}$ and the eigenvalues of N map to the eigenvalues of W', the lazy random walk matrix, by $\nu_i = (1 - \lambda_i)/2$. Note the i smallest eigenvalue of N is related to the i largest eigenvalue of W'. This inequality is the basis the analysis for one sweep sparse cuts algorithms that use the second eigenvector as a ranking because it relates the second eigenvalue to the squared minimum conductance of the graph.

Cheeger's Inequality also relates the conductance of the graph to the distance of random walks from their stationary distribution which is useful in the analysis of random walks for sparse cuts. $p_{t,u}(v)$ is the probability mass at vertex v on time step t of a lazy random walk starting from vertex u. $p_{t,u} = (W')^t\chi_u$ where χ_u is the indicator vector for vertex u. The distance from the stationary distribution for each vertex v given a seed vertex u is controlled by its second largest eigenvalue λ_2.

$$|p_{t,u}(v) - \pi(v)| \leq \sqrt{\frac{d_v}{d_u}}\lambda_2^t$$

.

This can be converted to an expression using eigenvalues of the normalized Laplacian matrix.

$$|p_{t,u}(v) - \pi(v)| \leq \sqrt{\frac{d_v}{d_u}}\left(1 - \frac{\nu_2}{2}\right)^t$$

By applying Cheeger's inequality, we achieve a bound on the convergence as a function of the minimum conductance of the graph.

Theorem 2.

$$|p_{t,u}(v) - \pi(v)| \leq \sqrt{\frac{d_v}{d_u}}\left(1 - \frac{\phi_G^2}{4}\right)^t$$

Although this result is useful for the classical sparse cuts problem, it depends on the global conductance of the graph instead local conductance, something we attempt to avoid in our model. For the large graphs we are interested in, this bound may be loose because ϕ_G is tiny due to a bottleneck in a distant part of the graph that is never encountered by a random walk. This motivates studying Local Cheeger Inequalities that relate the convergence of random walks to local conductances they actually encounter.

2.3. **Local Sparse Cuts.** The following Theorem by Lovász and Simonovits spawned research into local bounds on random walks using the idea of computing conductance over visited cuts [1].

Theorem 3. *Let vol(S) $\leq$ vol(G)/2. Then, for any seed u,*

$$|p_{t,u}S - \pi(S)| \leq \sqrt{\frac{vol(S)}{d_u}}\left(1 - \frac{\phi_{t,u}}{8}\right)^t.$$

where $\phi_{t,u}$ is the minimum conductance of a threshold cut of the vectors

$$D^{-1}p_0, D^{-1}p_1, ..., D^{-1}p_{t-1}.$$

This theorem forms the basis for Spielman and Teng's sparse cuts algorithm [10]. It follows directly from this theorem by running the lazy random walk for t steps, computing the conductance of each of the n cuts along the way and returning the minimum conductance cut it sees. By reversing Lovász and Simonovits to become a bound on $\phi_{t,u}$, they obtain an upper bound on the conductance of the resulting set.

Although this makes progress towards our goal of an efficient local partitioning algorithm, it still considers t sweeps of probability vectors through t steps of a lazy random walk. And each ordered probability vector sweep has n locations to perform a partition. This means that the number of considered cuts is $O(tn)$. Because this runtime is still a function of n, it is not tractable for our model. Additionally, there is room to reduce the number of sweeps considered from t back to just one sweep.

Local graph partitioning is a tractable variant of the graph partitioning problem that allows us to efficiently perform partitioning on large graphs. The problem is given seed vertex u along with a target volume s to find a cut around u with volume at most s that has minimal conductance. Because we are restricting partitions to this volume, researchers have been able reduce the runtime of computing random walks and number of cuts considered to both be functions of s rather than n. This is done by considering other random walks.

2.4. **PageRank.** PageRank is random walk that models the probability of ending at a vertex when a walk randomly moves for a number of steps governed by a geometric random variable. We will consider the Personalized PageRank which assumes an initial seed u. α is a jumping constant representing the likelihood of ending the walk at each step. This can be understood recursively by the following recurrence.

$$\mathrm{pr}_{\alpha,u} = \alpha\chi_u + (1-\alpha)\mathrm{pr}_{\alpha,u}W\chi_u.$$

Equivalently, PageRank can be computed explicitly via an infinite geometric sum over the number of hops of the walk. This also gives the closed form solution.

$$\mathrm{pr}_{\alpha,u} = \alpha\sum_{k=0}^{\infty}(1-\alpha)^k W^k \chi_u = \alpha(I-(1-\alpha)W)^{-1}\chi_u$$

The algorithm for computing local spare cuts with PageRank is simple. Simply compute the PageRank vector $\mathrm{pr}_{\alpha,u}$ for a chosen $\alpha = O(\phi)$ where ϕ is a guess at the minimum conductance and keeping in mind the maximum volume s. The perform a sweep of $\mathrm{pr}_{\alpha,u}$ to obtain a local, low conductance cut of volume less than s.

PageRank can be ϵ-approximated in sub-linear time of the size. The current best runtime is $O(\frac{\log s \log 1/\epsilon}{\log \epsilon})$ [7]. Additionally, sorting the vector via a comparison sort runs in $O(s \log s)$

assuming for comparing volume to vertices that edges are unweighted. This gives combined runtime of $O(\frac{\log s \log 1/\epsilon}{\log \epsilon} + s \log s)$ [7].

The final step is to analyze the quality of cuts PageRank produces is to determine a Cheeger inequality mixing result. Because we are only considering cuts of size at most s, conductance that appears in the inequality is only useful for the analysis if it is also limited to that volume. Therefore, the s-local conductance of the PageRank vector with seed u and parameter α is $\sigma_{\alpha,u,s}$ and is defined to be the minimum conductance over a sweep of $\mathrm{pr}_{\alpha,u}$ only considering cuts S_i with $\mathrm{vol}(S_i) \leq s$. This gives the following inequality:

$$\mathrm{pr}_{\alpha,u}(S) - \pi(S) \leq \alpha t + \sqrt{s}\left(1 - \frac{\sigma_{\alpha,u,s}}{8}\right)^t$$

for any integer $t \geq 0$. Note that whenever a function f is applied to a vertex set S, $f(S) = \sum_{v \in S} f(v)$. Without going into the analysis, this inequality can be reversed to give $\sigma_{\alpha,u,s} \leq O(\sqrt{\phi \log s})$ which is the best known cut bound obtained by local sparse cut algorithms. Heat Kernel PageRank showed promise at improving on this bound, but it's result follows from analysis of an incorrect Cheeger inequality.

2.5. **Heat Kernel PageRank.** Heat Kernel PageRank is a variant on PageRank. It is inspired by the heat equation from spectral geometry and is similar to the original PageRank except the underlying geometric distribution is replaced with a Poisson distribution. Instead of α, it has a parameter t which represents the temperature of the Heat Kernel and also the expected number of steps of the walk. It still uses u to seed the random walk. Instead of a recurrence, it satisfies the heat equation [2].

$$\frac{\partial}{\partial t}\rho_{t,u} = -\rho_{t,u}(I - W)$$

And the Heat Kernel PageRank can be also be written as the infinite exponential sum of the number of steps of the walk given by the Poisson distribution also resulting in a closed form solution.

$$\rho_{t,u} = e^{-t}\sum_{k=0}^{\infty}\frac{t^k}{k!}W^k\chi_u = e^{-t(I-W)}\chi_u$$

The strategy for using Heat Kernel PageRank to find local sparse cuts is the same as in the case of regular PageRank, compute and perform a sweep of $\rho_{t,u}$ to find low conductance cut. The state of the art method for computing a local ϵ-approximation of Heat Kernel PageRank is a little slower but still runs in sub-linear time of $O(\frac{\log s \log 1/\epsilon}{\epsilon^3 \log\log \epsilon})$ [8]. Therefore the combined runtime of the algorithm is $O(\frac{\log s \log 1/\epsilon}{\epsilon^3 \log\log\epsilon} + s\log s)$.

The difference is that Chung was able to obtain a different Cheeger inequality than the one used in PageRank analysis [3]. It still gives an upper bound on the distance of the Heat Kernel random walk from the stationary distribution using the s-local conductance of a sweep of the Heat Kernel PageRank vector. This conductance is called $\kappa_{t,u,s}$. A differing technical detail is that the s-local conductance when talking about $\kappa_{t,u,s}$ considers cuts of volume less than $2s$ instead of just s. She obtains the following inequality:

$$\rho_{t,u}(S) - \pi(S) \leq \sqrt{\frac{s}{d_u}} e^{-t\kappa_{t,u,s}^2/4}$$

.

Notice that it has slightly different from than the inequality used for PageRank, specifically it lacks an additive term that is a function of its parameter. Assuming this inequality is true, it follows that the quality of cuts produced by the Heat Kernel PageRank is $O(\sqrt{\phi})$. This time, we will present how this bound is obtained because it is relevant to our discussion.

Let $\phi^*_{u,s}$ be the minimum Cheeger Ratio of any cut seeded at u of volume less than $2s$. Chung proves a lower bound on this distance from the stationary distribution using this conductance.

$$\frac{1}{2} e^{-2t\phi^*_{u,s}} \leq \rho_{t,u}(S) - \pi(S)$$

These two can be combined to give

$$\frac{1}{2} e^{-2t\phi^*_{u,s}} \leq \sqrt{\frac{s}{d_u}} e^{-t\kappa_{t,u,s}^2/4}.$$

After, rearranging we obtain

$$2\sqrt{\phi^*_{u,s} + \frac{2\log s}{t}} \geq \kappa_{t,u,s}.$$

When we choose $t = 2\log s/\phi_{u,s}$, the resulting bound on the Heat Kernel PageRank conductance is $O(\sqrt{\phi_{u,s}}) \geq \kappa_{t,u,s}$ giving an improvement of $O(\sqrt{\log s})$ over the resulting PageRank conductance.

3. Implications on Hardness on Approximation

The correctness of the resulting conductance from the Heat Kernel PageRank we first questioned because of its implications on complexity theory if it were correct. Specifically, this result affect the Small-Set Expansion Problem [5]. The Small-Set Expansion Problem is an optimization problem to determine the expansion profile $\phi(\delta)$ of a graph which is the minimum conductance of a subset with a certain size, here size referring to the number of vertices in the set instead of volume. Formally, $\phi(\delta) = \min_{|S|=\delta n} \phi_S, \forall \delta \in [0, 1/2]$. The approximation variant involves finding a constant approximation to this expansion profile. Like many problems when dealing with approximation algorithms, this can be restated by an equivalent decision problem called the Gap-Small-Set Expansion Problem. Given (η, δ), decide $\phi(\delta) \geq 1 - \eta$ versus $\phi(\delta) < \eta$. Note that this is called a promise problem because the algorithm is promised that exactly one of those conditions hold in the input, and it just has to decide which one.

The Gap-Small-Set Expansion Problem has recently receive attention because of its ties to the Unique Games Conjecture. Informally, the Unique Games Problem is a famous problem in the complexity of approximation algorithms because it reduces to many classical NP-hard problems like Vertex Cover and Max Cut. Although researchers have not yet found a reverse reduction from one of these problems to Unique Games, the Unique Games Conjecture states that this reduction exists and therefore the Unique Games Problem is NP-hard. Although many unproven conjectures still surround it, Unique Games is often described at the 3-SAT of approximation algorithms.

Gap-Small-Set Expansion was shown to reduce to the Unique Games problem, becoming one of the first nontrivial problems to do so. Although a reduction from Unique Games to Gap-Small-Set Expansion has not been found, researchers conjecture that it exists making the two problems equivalent. If this conjecture and the Unique Games Conjecture are both true, it would imply that Gap-Small-Set Expansion is NP-hard.

The surprising thing about the result of Heat Kernel PageRank is that it is able to solve the Gap-Small-Set Expansion problem. Recall that Heat Kernel PageRank takes a parameter ϕ when choosing t and is able to produce a cut that is $O(\sqrt{\phi})$ for a specified volume s and seed u. This can be adapted to Gap-Small-Set Expansion by relating η to ϕ and δ to s to apply the algorithm. And the resulting cut is a constant approximation of the input so it is able to distinguish between sets with high conductance and low conductance for a given size δ, solving the Gap-Small-Set Expansion problem in polynomial time of the target size.

This result implies that Gap-Small-Set Expansion is easy and thus not NP-hard. Although this doesn't immediately imply that the Unique Games Problem is easy because this reduction is only conjecture, it gives evidence against the Unique Games Conjecture.

4. Heat Kernel Cheeger Inequality

Because of the strength of the result, the Cheeger Inequality used to derive the approximation ratio deserved more inspection. We found this inequality to be false through a counterexample. Next, we analyzed Chung's proof to find the mistake. And finally, we attempted to weaken the statement to see if we could obtain an inequality similar to the one used in PageRank.

4.1. **Counterexample.** At first glance, the inequality seem plausible because of its similar structure to the other Cheeger inequalities. But its restriction to only consider cuts that have a limited volume without adding an error term causes it be sometimes overestimate the spread of probability mass. Consider the counterexample in Figure 5.1. This is Dumbbell Graph D_n that represents two complete graphs of n vertices each connected by one edge. The complication is that the connecting edge has an extremely small weight $\delta << 1$. This simulates the effect of almost disconnecting the two complete graphs. The target set is S with volume $s = 10$ and the seed is labeled u. The edge gives the graph an extremely low global conductance of $\phi_G = O(\frac{\delta}{n^2}) \approx 0$ by considering a cut that naturally partitions the two complete graphs. Note that the stationary distribution for the graph is $\pi(v) \approx 1/12$.

Meanwhile for a reasonably large choice of t that is still significantly less than $1/\delta$, $\rho_{t,u} \approx 1/6$ for u in the left complete graph while $\rho_{t,u} \approx 0$ because the random walk does not have

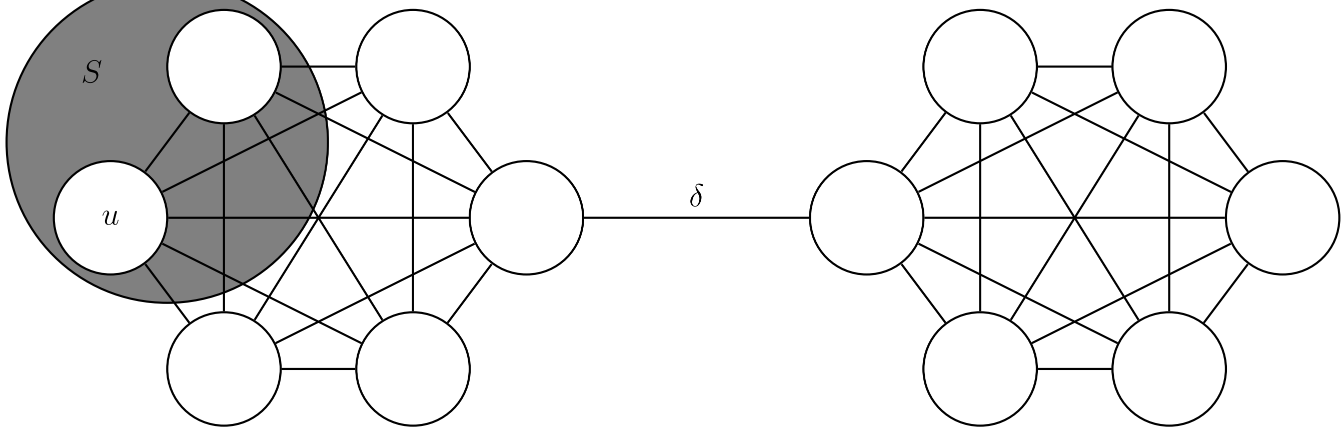

FIGURE 4.1. Modified Dumbbell Graph D_6 Counterexample

high enough temperature to significantly cross the weak edge. Thus, the distance from the stationary distribution for set S is $1/6$. Finally, $\kappa_{t,u,s} \approx 2/5$ because sweeps of $\rho_{t,u}$ can essentially take any four vertices in the left complete graph to maintain a volume less than $2s$. But notice that $\kappa_{t,u,s}$ is significantly larger than ϕ_G. By limiting the volume of cuts on the sweep, the cut that separates the two complete graphs cannot be considered because it is too large. Thus the right side of the inequality is small due to a reasonably sized choice of t that is not counteracted much by $\kappa_{t,u,s}$ and certainly smaller than $1/6$. For concrete values to see this, choose $\delta = 1/1000$ and $t = 100$. Note that this effect can be replicated in an unweighted graph by replacing the edge with weight δ with a long path while increasing the sizes of the complete graphs forming a Bolas Graph B_n.

Analytically, this counterexample can be analyzed by realizing that $\kappa_{t,u,s}$ is unable to be depending on δ while $\rho_{t,u}$ is extremely dependent on δ and t. Thus, an adversary attempting to disprove this inequality could increase t to drive the right side to zero while decreasing δ to maintain a constant value on the left side.

This counter example gives us the intuition that the convergence of the random walk can be extremely dependent on the structure of the graph outside of a certain volume. In our example, the random walk seems to think it converged to the stationary distribution because it had a high local conductance when really a harsh bottleneck was just outside the set S we were considering. This prevented the random walk from quickly converging to the true stationary distribution. But because the inequality cannot capture this low conductance cut because it has too large of a volume, the upper bound believed the walk converged significantly more quickly than it actually has.

4.2. **Review of the Proof.** The incorrect inequality Chung uses is the following.

Theorem 4. *In a graph G with a subset S with volume $s \leq vol(G)/4$, for any vertex $u \in G$,*

$$\rho_{t,u}(S) - \pi(S) \leq \sqrt{\frac{s}{d_u}} e^{-t\kappa_{t,u,s}^2/4}$$

The proof follows the form where she attempts to upper bound the maximum the distance from the stationary distribution for any set of volume x by the right exponential of the inequality. First, she derives an inequality for an arbitrary ranking function $f : V \to \Re$. Define $f(u,v) = f(u)/d_v$ if $(u,v) \in E$ otherwise 0. Define $f(x)$ for some volume x to be the maximum summed ranking using f for a set that achieves volume x. Formally when x is an integer,

$$f(x) = \max_{T \subseteq V \times V \text{ and } |T| = x} \sum_{(u,v) \in T} f(u,v).$$

$f(x)$ is interpolated for non-integer x such that for real $x = k + r$ and $r \in [0,1]$, $f(x) = (1-r)f(k) + rf(k+1)$. Notice this definition causes $f(x)$ to be concave in x because on increasing x, the function adds an edge that has ranking smaller than the previously added edges. Recall that W' is the lazy walk matrix.

$$\begin{aligned} f(S)W' &= \tfrac{1}{2}\left(f(S) + \sum_{u \to v \text{ and } v \in S} f(u,v)\right) \\ &= \tfrac{1}{2}\left(\sum_{u \in S \text{ or } v \in S} f(u,v) + \sum_{u \in S \text{ and } v \in S} f(u,v)\right) \\ &\leq \tfrac{1}{2}(f(\mathrm{vol}(S) + |\partial S|) + f(\mathrm{vol}(S) - |\partial S|)) \\ &= \tfrac{1}{2}(f(\mathrm{vol}(S)(1+\phi_S)) + f(\mathrm{vol}(S)(1-\phi_S))) \end{aligned}$$

So far the analysis holds cleanly for any f. Now we choose $f = \rho_{t,u} - \pi$ to rank by the distance from the stationary distribution. This allows us to rewrite the inequality above in terms of the target volume x and the minimum s-local conductance over the sets we are considering which are constructed, given our new f, by sweeps of $\rho_{t,u}$. By the concavity of $f(x)$, we can replace ϕ_S with $\kappa_{t,u,s} \leq \phi_S$ to obtain this inequality because decreasing ϕ_s increases the sum.

$$f(x)W' \leq \frac{1}{2}(f_t(x(1+\kappa_{t,u,s})) + f_t(x(1-\kappa_{t,u,s}))))$$

Note this only holds for $x \in [0, 2s]$ otherwise $\kappa_{t,u,s} \leq \phi_S$ may not be true. In fact for $x > 2s$, this inequality may be strongly incorrect. For example, the real conductance of a larger set S could be extremely small, and we replace it with the larger $\kappa_{t,u,s}$. This is the same phenomenon that appeared in the counterexample. While this doesn't immediately affect the proof, this boundary condition affects $x \in [0, 2s]$ as we show next.

Now for $x \in [0, 2s]$, we apply the fact that $\rho_{t,u}$ satisfies the heat equation to obtain

$$\begin{aligned}
\frac{\partial}{\partial t} f_t(x) &= -\rho_{t,u}(I-W)(x) \\
&= -2\rho_{t,u}(I-W')(x) \\
&= -2f_t(x) + 2f_t(x)W' \\
&\le -2f_t(x) + f_t(x(1+\kappa_{t,u,s})) + f_t(x(1-\kappa_{t,u,s}))) \\
&\le 0
\end{aligned}$$

The last inequality proves $f_t(x)$ is decreasing in t by applying the fact that $f_t(x)$ is concave in x to the inequality above it. The final goal is to upper bound $f_t(x)$ by $g_t(x) = \sqrt{\frac{s}{d_u}} e^{-t\kappa_{t,u,s}^2/4}$ to prove the theorem. To do this, Chung notes the initial conditions of the two functions satisfy $f_0(x) \le g_0(x)$, $f_t(0) = g_t(0)$, and $\frac{\partial}{\partial t} f_t(x)|_{t=0} \le \frac{\partial}{\partial t} g_t(x)|_{t=0}$. Finally, she uses the bound we obtained on the partial derivative of $f_t(x)$ with respect to t to attempt that $g_t(x)$ decays less rapidly.

$$\begin{aligned}
\frac{\partial}{\partial t} g_t(x) &= -\frac{\kappa_{t,u,s}^2}{4} g_t(x) \\
&\ge \left(-2 + \sqrt{1+\kappa_{t,u,s}} + \sqrt{1-\kappa_{t,u,s}}\right) g_t(x) \\
&= -2g_t(x) + g_t(x(1+\kappa_{t,u,s})) + g_t(x(1-\kappa_{t,u,s})) \\
&\ge -2f_t(x) + f_t(x(1+\kappa_{t,u,s})) + f_t(x(1-\kappa_{t,u,s})) \\
&\ge \frac{\partial}{\partial t} f_t(x)
\end{aligned}$$

The jump to the second step follows from noting that $-y^2/4 \ge -2 + \sqrt{1+y} + \sqrt{1-y}$ for any $y \in [-1,1]$. However this is the flaw in the proof. Chung assumes almost inductively when comparing the partial derivatives at x that $f_t(x(1+\kappa_{t,u,s})) \le g_t(x(1+\kappa_{t,u,s}))$. But when looking at x near but below $2s$, this inequality may not be true because it reaches outside the boundary. Specifically, we showed earlier that outside $2s$, $f(x)W'$ may be significantly larger than what the inequality shows. Thus, the decay of $f_t(x)$ with respect to t may be stagnant while $g_t(x)$ is still decaying rapidly. So $f(x) \not\le g(x)$ for $x > 2s$. Additionally, this affects results in the boundary when invoked as $f_t(x(1+\kappa_{t,u,s})) \le g_t(x(1+\kappa_{t,u,s}))$. Therefore, we cannot prove $f_t(x) \le g_t(x)$.

If this wasn't an issue, the proof would be completed by the following:

$$\begin{aligned}
\rho_{t,u}(S) - \pi(S) &\le \rho_{t,u}(s) - \pi(s) \\
&\le f_t(s) \\
&\le \sqrt{\frac{s}{d_u}} e^{-t\kappa_{t,u,s}^2/4}
\end{aligned}$$

4.3. **Weaker Results.** As demonstrated both in the counterexample and proof analysis, the theorem breaks down when there is a large disparity between the local conductance of a certain volume and the conductance of the larger region it is inclosed in. This is because although locally it seems like the walk is close to its stationary distribution, the small conductance prevents the probably mass from escaping to converge to the true stationary distribution.

Notice that the proof is correct if we replace $\kappa_{t,u,s}$ with $\kappa_{t,u}$, the minimum conductance of all possible cuts of $\rho_{t,u}$ including cuts with volume greater than $2s$. Because $\kappa_{t,u}$ does not have a maximum allowed volume, it doesn't have boundary conditions on the inequalities. Thus the argument follows nicely as intended and gives the following bound.

$$\rho_{t,u}(S) - \pi(S) \leq \sqrt{\frac{s}{d_u}} e^{-t\kappa_{t,u}^2/4}$$

This turns out to be just a weaker version of a Cheeger's inequality for the heat kernel random walk. Chung has a proof of a stronger form of it in her paper.

$$\rho_{t,u}(S) - \pi(S) \leq \sqrt{\frac{s}{d_u}} e^{-t(\kappa_{t,u}^2 + \kappa_{t,v}^2)/4}$$

Additionally because this conductance considers all cuts, it doesn't give the algorithmic benefit of reducing the runtime to a dependence on just s. Thus, the runtime of the Heat Kernel PageRank would have to be a function of n to use this inequality.

We noticed that the Cheeger Inequality for PageRank avoided these problems but still used a conductance derived from a sweep of limited volume. This inequality differed slightly in that it had an additive term. Our intuition is that we should be able to achieve a similar result using the Heat Kernel PageRank. And this inequality would likely give cuts that perform the same as PageRank with conductance $O(\sqrt{\phi \log s})$. One immediate additive term we can use is $1/2 - s/\mathrm{vol}(G)$ to give the inequality

$$\rho_{t,u}(S) - \pi(S) \leq \sqrt{\frac{s}{d_u}} e^{-t\kappa_{t,u,s}^2/4} + \left(\frac{1}{2} - \frac{s}{\mathrm{vol}(G)}\right).$$

This term compares the worst case difference between the stationary distribution of a subgraph of volume $2s$ and the full graph. This compensates for the fact that little probability mass may have left the volume of $2s$ as seen in our counterexample. Note that this bound is exceptionally weak if $s << \mathrm{vol}(G)/2$ but performs similarly to the non-volume restricted case when $s \approx \mathrm{vol}(G)/2$. This additive term fixes the boundary conditions by ensuring $f_t(x) \leq g_t(x)$ even for $x > 2s$. Future work should look at additional terms like this that may instead depend on t to obtain a tighter bound.

5. Conclusion

We surveyed the problem of local graph partitioning by giving a background of approximate graph partitioning in general, discussing Cheeger Inequalities and their relation to random walks, and reviewing the PageRank and Heat Kernel PageRank random walks for use in fast, local sparse cut algorithms. We highlighted a recent result from the Heat Kernel PageRank that would improve on the best known cut approximation by removing its dependence on s and noted the implications on complexity theory if it is correct. Finally, we provided a counterexample for a Cheeger inequality used to obtain the bound, explained the flaw in the proof, and provided alternate, weaker Cheeger inequalities that may be useful instead to analyze the performance of Heat Kernel PageRank local cuts.

6. Acknowledgements

I would like to thank Lorenzo Orecchia and Siggi Kjartansson for their discussions and collaboration when researching this topic.

References

[1] 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.

[2] F. Chung, " A local graph partitioning algorithm using heat kernel pagerank,". Internet Mathematics 6.3 (2009): 315-330.

[3] F. Chung, " The heat kernel as the pagerank of a graph," Proceedings of the National Academy of Sciences, vol. 104, no. 50, pp. 19735-19740, Dec. 2007. [Online]. Available: http://www.pnas.org/cgi/doi/10.1073/pnas.0708838104

[4] A. Louis, P. Raghavendra, P. Tetali, and S. Vempala, "Many sparse cuts via higher eigenvalues." In Proceedings of the 44th symposium on Theory of Computing (pp. 1131-1140). ACM, 2012.

[5] P. Raghavendra and D. Steurer, "Graph Expansion and the Unique Games Conjecture, ". Proceedings of the 42nd ACM symposium on Theory of computing. ACM, 2010.

[6] R. Andersen, F. Chung, and K. Lang, "Local Graph Partitioning using PageRank Vectors, " 2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06), pp. 475-486, 2006. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4031383

[7] C. Borgs, M. Brautbar, J. Chayes, and S. Teng, "A Sublinear Time Algorithm for PageRank Computations, ". Algorithms and Models for the Web Graph. Springer Berlin Heidelberg, 2012. 41-53. [Online]. Available: http://research.microsoft.com/en-us/um/people/borgs/Papers/SublinearPR.pdf

[8] F. Chung and O. Simpson, "Solving Linear Systems with Boundary Conditions Using Heat Kernel Pagerank, ". Algorithms and Models for the Web Graph. Springer International Publishing, 2013. 203-219. [Online]. Available: http://www.math.ucsd.edu/ fan/wp/hklinear_waw.pdf

[9] S. Arora, S. Rao, and U. Vazirani, "Expander Flows, Geometric Embeddings and Graph Partitioning, ". Journal of the ACM (JACM) 56.2 (2009): 5.

[10] D. Spielman and S. Teng, "Nearly-Linear Time Algorithms for Graph Partitioning, Graph Sparsification, and Solving Linear Systems, ". Proceedings of the thirty-sixth annual ACM symposium on Theory of computing. ACM, 2004. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.205.2406&rep=rep1&type=pdf