ICCM 2007 · Vol. II · 1–4

Four proofs for the Cheeger inequality and graph partition algorithms

Fan Chung [*†]

Abstract

We will give four proofs of the Cheeger inequality which relates the eigenvalues of a graph with various isoperimetric variations of the Cheeger constant. The first is a simplified proof of the classical Cheeger inequality using eigenvectors. The second is based on a rapid mixing result for random walks by Lovász and Simonovits. The third uses PageRank, a quantitative ranking of the vertices introduced by Brin and Page. The fourth proof is by an improved notion of the heat kernel pagerank. The four proofs lead to further improvements of graph partition algorithms and in particular the local partition algorithms with cost proportional to its output instead of in terms of the total size of the graph.

2000 Mathematics Subject Classification: 05C50, 68R10.
Keywords and Phrases: Cheeger inequalities, Laplacian, random walks, heat kernel, Dirichlet eigenvalues, PageRank, graph partition algorithms

1. Introduction

One of the major tools in spectral graph theory is the Cheeger inequality. It concerns the relationship between two important graph invariants, the spectral gap and the Cheeger constant. For an (undirected) graph $G = (V, E)$, we denote by λ_G the spectral gap of the (normalized) Laplacian (see [5]). Let h_G denote the Cheeger constant of G, defined as the minimum value h such that any cut separating a set of volume x requires at least hx edges if x is less than half of the total volume. (The

*University of California, San Diego, USA
†Research supported in part by NSF Grants DMS 0457215 and ITR 0426858

detailed definitions will be defined later.) The Cheeger inequality states that for a connected graph G, we have

$$2h_G \geq \lambda_G \geq \frac{h_G^2}{2}.$$

The proof of the above (discrete) Cheeger inequality is quite similar to the proof in the continuous case in spectral geometry due to Jeff Cheeger [4]. The idea of the proof is to use eigenvectors to guide the search for good cuts. Indeed, the constructive proof gives the following expanded version of the Cheeger inequality:

$$2h_G \geq \lambda_G \geq \frac{\alpha_G^2}{2} \geq \frac{h_G^2}{2} \tag{11}$$

where α_G is the minimum Cheeger ratio of the size of the edge boundary and the volume of sets consisting of the vertices associated with the largest i coordinates of the eigenvector which is associated with λ_G. A somewhat simplified proof for (11) will be given in Section 3. In the search for the optimum cut (evaluated by the Cheeger constant) among an exponential number of possibilities, we can then focus on a linear number of choices for the cut using the order determined by the eigenvector. The above Cheeger inequality guarantees that the cut resulting from this efficient algorithm has the Cheeger ratio within a quadratic factor of the optimum. This spectral partition algorithm has been widely used in numerous applications in the generic divide-and-conquer approach by reducing problems recursively to smaller and more manageable sizes [12].

Nowadays we often are dealing with problems involving graphs or networks of prohibitively large sizes, such as the Web graph, various social networks, biological networks or many information networks arising from massive data sets [9]. It is often infeasible to compute eigenvectors for a graph with hundreds of thousands of vertices. For problems involving such large graphs, the total number of vertices and edges of the graph are no longer realistic parameters. Alternative approaches are needed in a way that computation can be carried out "locally" and the local optimum can be analyzed. As it turns out, one of the key ideas rests on random walks.

In the early 90's, Lovász and Simonovits derived an elegant result on rapid mixing of random walks in their work on approximating the volume of convex polytopes [13, 14]. Using this mixing result, Spielman and Teng in 2004 gave a graph partition algorithm which has running time proportional to the size of the output, and not depending on the total number of vertices [15]. Although it was not specifically described in the papers mentioned above, the basic thrust of this approach is a

Cheeger inequality,

$$2h_G \geq \lambda_G \geq \frac{\beta_G^2}{8} \geq \frac{h_G^2}{8} \tag{12}$$

where β_G is the minimum Cheeger ratio of the sets determined by the largest i values using the vector which is the probability distribution of a random walk starting at a vertex v after k steps over all vertices v, $i \leq n = |V(G)|$ and $k \leq (16 \log n)/\lambda_G^2$. Some local variations will be further discussed in Section 5.

A third Cheeger inequality relies on the notion of PageRank, which was first introduced by Brin and Page [3] as a method for quantitatively ranking Webpages by the Web search engines. The PageRank can be viewed as a combination of random walks, scaled by a parameter called the jumping constant. (The detailed definition will be given later.) In [1, 2], PageRank is used for developing a local partition algorithm which improves upon the partition algorithm using random walks. PageRank provides a way to deal with random walks of various lengths simultaneously in an organized fashion. The underlying theme here is again a Cheeger inequality, although it was not mentioned in [1, 2]. In Section 6, we will show that for a subset S in G with $\text{vol}(S) \leq \text{vol}(G)/2$, we have

$$h_S \geq \lambda_S \geq \frac{\gamma_S^2}{4 \log(\text{vol}(G))}$$

where λ_S is the Dirichlet eigenvalues on S and γ_S denotes the minimum of the Cheeger ratios determined by the PageRank involving only vertices in S with appropriately chosen parameters. As a result, the local partition algorithm using PageRank improves both the running time and performance of the previous algorithms by a factor of powers of $\log(\text{vol}(S))$ (see [1]).

The fourth Cheeger inequality involves a notion of pagerank based on the heat kernel of a graph. The heat kernel pagerank can be viewed as an exponential sum of random walks while PageRank is a geometric sum. The rate of diffusion of the heat kernel pagerank is controlled by a heat parameter $t \geq 0$. The heat kernel has many useful properties with close relations to the spectrum of the graph. We will use a rapid mixing inequality of the heat kernel to prove the following Cheeger inequality: For a subset S of volume $s \leq \text{vol}(G)^{2/3}$, we have

$$h_S \geq \lambda_S \geq \frac{\kappa_S^2}{8} \tag{13}$$

where κ_S denote the minimum Cheeger ratio of subsets S_i determined in a sweep of the heat kernel pagerank associated with vertices in S with $\text{vol}(S_i)$ at most $2s$ for some appropriately chosen parameters for the heat kernel pagerank. The details will be given in Section 7. This again leads to an improved local partition algorithm.

In this paper, we examine four Cheeger inequalities and their proofs using different analytic methods. In their entirety, it is of interest to observe the comparisons as well as pointing out the interconnections. The methods intertwine, ranging from their roots in differential geometry through spectral graph theory, random walks, PageRank, heat kernels to graph partition algorithms. These topics have been rapidly advancing and emerging as main tools for the age of information and the Internet.

2. Preliminaries

There are basically two ways to give definitions for the heat kernel and eigenvalues of a graph. One way is to use symmetric matrices throughout which in some situations simplify the arguments (as in [5]). However, in dealing with random walks, the transition probability matrix W is not necessarily symmetric. Namely,

$$W = D^{-1}A$$

where D is the diagonal degree matrix and A denotes the adjacency matrix. The normalized Laplacian $\mathcal{L}$ is defined by

$$\mathcal{L} = I - D^{-1/2}AD^{-1/2} = D^{1/2}(I - W)D^{-1/2}.$$

Since the main results will be stated in the language of random walks, we will use the unsymmetrical version. The spectral gap λ_G is the least nonzero eigenvalue of $\mathcal{L}$. The value λ_G can be expressed as the infimum of the Rayleigh quotient:

$$\lambda_G = \inf_g R(g) = \inf_g \frac{\sum_{u\sim v}(g(u) - g(v))^2}{\sum_u g^2(u)d_u}$$

where g ranges over functions defined on the vertices of G satisfying $\sum_u g(u)d_u = 0$ (see [5]).

For a set S, the volume of S, denoted by $\mathrm{vol}(S)$ is defined to be $\mathrm{vol}(S) = \sum_{v\in S} d_v$. The volume of G is written as $\mathrm{vol}(G) = \sum_v d_v$. Let π denote $\pi = (d_1/\mathrm{vol}(G), d_2/\mathrm{vol}(G), \ldots, d_n/\mathrm{vol}(G))$, indicated as a row vector. For a connected non-bipartite graph, π is the stationary distribution for the random walk W on G. In this paper, we consider only connected graphs.

For a set S, the Cheeger ratio of S is defined to be

$$h_S = \frac{|\partial S|}{\min\{\mathrm{vol}(S), \mathrm{vol}(G) - \mathrm{vol}(S)\}},$$

where the boundary of S is denoted $\partial S = \{\{u,v\} \in E \ : \ u \in S \text{ and } v \notin S\}$. Let $\bar{S}$ denote the complement of S. Then $h_S = h_{\bar{S}}$. The Cheeger constant of G is denoted by

$$h_G = \min_S h_S.$$

Brin and Page introduced the notion of PageRank which has been used as a major tool for conducting Web search. The definition of PageRank can be entirely described in graph-theoretical terms since its original application is in the Webgraph which has its vertex set consisting of all webpages and edge set consisting of all hyperlinks. In a graph G, the PageRank operator can be described as a combination of random walks on G and a *jumping constant* c, $0 \le c \le 1$. The PageRank matrix R_c is defined to be

$$R_c = c \sum_{k=0}^{\infty} (1-c)^k \mathbf{W}^k.$$

An equivalent definition is that the PageRank matrix satisfies the following recurrence:

$$R_c = cI + (1-c) R_c \mathbf{W}. \tag{21}$$

For a vertex u, let χ_u denote the $(0,1)$ indicator vector, i.e., $\chi_u(v) = 1$ if $v = u$ and 0 otherwise. The personalized PageRank has additional parameters, the *jumping constant* and a *preference vector*. For example, a typical preference vector associated with a vertex u is χ_u. All vectors here are taken to be row vectors unless mentioned otherwise. For a vertex u, the corresponding personalized PageRank, denoted by pr_u, is defined by $\mathrm{pr}_u = \chi_u R_c$, the matrix product of χ_u and R_c. The personalized PageRank can be viewed as quantitative ranking of all vertices with respect to u. The original definition for PageRank has the preference vector $(1/n, 1/n, \ldots, 1/n)$. In general, we can consider an arbitrary starting vector s, summing to 1, in place of χ_u.

We will consider a new notion of pagerank which is based on the heat kernel of a graph [8]. The heat kernel pagerank also has two parameters, the *heat* t and a seed distribution f. The heat kernel pagerank $\rho_{t,f}$ is the matrix product of f and the heat kernel H_t which is defined as follows for $t \ge 0$.

$$\begin{aligned}
H_t &= e^{-t}(I + tW + \frac{t^2}{2}W^2 + \ldots + \frac{t^k}{k!}W^k + \ldots) \\
&= e^{-t(I-W)} \\
&= e^{-tL} \\
&= I - tL + \frac{t^2}{2}L^2 + \ldots + (-1)^k \frac{t^k}{k!}L^k + \ldots.
\end{aligned}$$

Equivalently, H_t can be defined by the following heat equation:

$$\frac{\partial}{\partial t} H_t = -(I - W) H_t.$$

3. The proof of the Cheeger inequality using eigenvectors

The Cheeger inequality in (11) is the discrete analog of the classical Cheeger inequality that appeared in Jeff Cheeger's thesis in differential geometry [4, 16]. The proof for the discrete version is in fact quite similar to that in the continuous version. Several proofs for (11) can be found in various contexts (see [5, 11]). Here we will give a proof that is simpler than that in [5].

Theorem 1 *In a graph G, the Cheeger constant h_G and the spectral gap λ_G are related as follows:*

$$2h_G \geq \lambda_G \geq \frac{\alpha_G^2}{2} \geq \frac{h_G^2}{2}$$

where α_G is the minimum Cheeger ratio of subsets S_i consisting of vertices with the largest i values in the eigenvector associated with λ_G, over all i.

Proof: Suppose the Cheeger constant h_G is achieved by the set S, i.e., $h_G = h_S = |\partial S|/\text{vol}(S)$. By considering

$$g = \chi_S - \frac{\text{vol}(S)}{\text{vol}(G)} \mathbf{1},$$

we have

$$\lambda_G \leq R(g) \leq 2h_G.$$

Thus, the proofs for the Cheeger inequality mainly concerns a lower bound for λ_G in terms of Cheeger ratios.

Now let g denote an eigenvector achieving λ_G. Namely, $\lambda_G = R(g)$ and $\sum_v g(v) d_v = 0$. We order the vertices so that

$$g(v_1) \geq g(v_2) \geq \ldots \geq g(v_n).$$

Let $S_i = \{v_1, \ldots, v_i\}$ and define

$$\alpha_G = \min_i h_{S_i}.$$

Let r denote the largest integer such that $\text{vol}(S_r) \leq \text{vol}(G)/2$. Since $\sum_v g(v)d_v = 0$,

$$\sum_v g(v)^2 d_v = \min_c \sum_v (g(v) - c)^2 d_v \leq \sum_v (g(v) - g(v_r))^2 d_v.$$

We define the positive and negative part of $g - g(v_r)$, denoted by g_+ and g_-, respectively, as follows:

$$g_+(v) = \begin{cases} g(v) - g(v_r) & \text{if } g(v) \geq g(v_r), \\ 0 & \text{otherwise,} \end{cases}$$

$$g_-(v) = \begin{cases} |g(v) - g(v_r)| & \text{if } g(v) \leq g(v_r), \\ 0 & \text{otherwise.} \end{cases}$$

We consider

$$\begin{aligned} \lambda_G &= \frac{\sum_{u \sim v} (g(u) - g(v))^2}{\sum_v g(v)^2 d_v} \\ &\geq \frac{\sum_{u \sim v} (g(u) - g(v))^2}{\sum_v (g(v) - g(v_r))^2 d_v} \\ &\geq \frac{\sum_{u \sim v} \left((g_+(u) - g_+(v))^2 + (g_-(u) - g_-(v))^2 \right)}{\sum_v \left(g_+(v)^2 + g_-(v)^2 \right) d_u}. \end{aligned}$$

Without loss of generality, we assume $R(g_+) \leq R(g_-)$ and therefore we have $\lambda_G \geq R(g_+)$ since

$$\frac{a+b}{c+d} \geq \min\{\frac{a}{c}, \frac{b}{d}\}.$$

We here use the notation

$$\tilde{\text{vol}}(S) = \min\{\text{vol}(S), \text{vol}(G) - \text{vol}(S)\}$$

so that

$$|\partial(S_i)| \geq \alpha_G \tilde{\text{vol}}(S_i).$$

Then we have

$$\begin{aligned}
\lambda_G &\geq R(g_+)\\
&= \frac{\sum_{u\sim v}(g_+(u)-g_+(v))^2}{\sum_u g_+^2(u)d_u}\\
&= \frac{\big(\sum_{u\sim v}(g_+(u)-g_+(v))^2\big)\big(\sum_{u\sim v}(g_+(u)+g_+(v))^2\big)}{\sum_u g_+^2(u)d_u \sum_{u\sim v}(g_+(u)+g_+(v))^2}\\
&\geq \frac{\big(\sum_{u\sim v}(g_+(u)^2-g_+(v)^2)\big)^2}{2\big(\sum_u g_+^2(u)d_u\big)^2} && \text{by the Cauchy-Schwarz inequality,}\\
&= \frac{\big(\sum_i |g_+(v_i)^2-g_+(v_{i+1})^2| \ \ |\partial(S_i)|\big)^2}{2\big(\sum_u g_+^2(u)d_u\big)^2} && \text{by counting,}\\
&\geq \frac{\big(\sum_i |g_+(v_i)^2-g_+(v_{i+1})^2| \alpha_G |\tilde{\mathrm{vol}}(S_i)|\big)^2}{2\big(\sum_u g_+^2(u)d_u\big)^2} && \text{by the def. of } \alpha_G,\\
&= \frac{\alpha_G^2}{2}\frac{\big(\sum_i g_+(v_i)^2(|\tilde{\mathrm{vol}}(S_i)-\tilde{\mathrm{vol}}(S_{i+1})|)\big)^2}{\big(\sum_u g_+^2(u)d_u\big)^2}\\
&= \frac{\alpha_G^2}{2}\frac{\big(\sum_i g_+(v_i)^2 d_{v_i}\big)^2}{\big(\sum_u g_+^2(u)d_u\big)^2}\\
&= \frac{\alpha_G^2}{2}.
\end{aligned}$$

Therefore we have proved the Cheeger inequality in (11):

$$2h_G \geq \lambda_G \geq \frac{\alpha_G^2}{2} \geq \frac{h_G^2}{2}.$$

4. Isoperimetric properties of random walks

In this section, we will describe some useful isoperimetric properties for random walks, which originated from [13, 14]. Both the second and third proofs of Cheeger inequalities in (12) and (13) will use these properties.

For a function $f : V \to \mathbb{R}$, we define

$$f(u,v) = \begin{cases} \frac{f(u)}{d_u} & \text{if } u \text{ is adjacent to } v;\\ 0 & \text{otherwise.}\end{cases}$$

For a set T of pairs of vertices of G, we extend the domain of f to include all subsets of the edge set as follows.

$$f(T) = \sum_{(u,v)\in T} f(u,v).$$

For a set S of vertices of G, we define

$$S_{in} = \{(u,v) \ : \ v \in S, u \text{ is adjacent to } v\},$$
$$S_{out} = \{(u,v) \ : \ u \in S, u \text{ is adjacent to } v\}.$$

Also, we define

$$f(S) = \sum_{u \in S} f(u).$$

It follows from the definition that

$$f(S) = f(S_{out}).$$

For any positive value x, we further extend the domain of f to include the set of reals between 0 and $\text{vol}(G)$ by defining:

$$f(x) = \sup_{\sum_v c_{u,v} = x} c_{u,v} f(u,v),$$

over $c_{u,v}$ with $0 \le c_{u,v} \le 1$. Suppose we order the vertices so that

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

Let S_i denote the set consisting of $v_1, \ldots, v_i$. Let h_f denotes the least Cheeger ratio h_{S_i} over all S_i determined by f.

Then the following facts result directly from the above definitions.

Fact 1

(i) *If* $x = \sum_{j \le i} d_{v_j} + r d_{v_{i+1}}$ *for* $0 \le r < 1$, *then*

$$\begin{aligned} f(x) &= \sum_{j \le i} f(v_i) + r f(v_{i+1}) \\ &= (1-r) f(\text{vol}(S_i)) + r f(\text{vol}(S_{i+1})). \end{aligned}$$

(ii) $f(x)$ *is convex in* x.

Let $\mathbf{W}$ denote the lazy walk defined by

$$\mathbf{W} = \frac{I + W}{2}.$$

The following lemma on lazy walks is in the same spirit as that in Lovász and Simonovits [13, 14] and also in [1, 6]. For the sake of completeness, we include a simplified proof here.

Lemma 1 *For a subset* $S \subseteq V$, *we have*

$$\begin{aligned} f\mathbf{W}(S) &= \frac{f(S_{in}) + f(S_{out})}{2} \\ &\le \frac{f(\text{vol}(S)(1 + h_S)) + f(\text{vol}(S)(1 - h_S))}{2}. \end{aligned}$$

Proof: We note that

$$\begin{aligned} f\mathbf{W}(S) &= \frac{f(S)+f(S_{in})}{2} \\ &= \frac{f(S_{out})+f(S_{in})}{2} \\ &= \frac{f(S_{in}\cup S_{out})+f(S_{in}\cap S_{out})}{2} \\ &\le \frac{f(\mathrm{vol}(S)+|\partial S|)+f(\mathrm{vol}(S)-|\partial S|)}{2} \\ &\le \frac{f(\mathrm{vol}(S)(1+h_S))+f(\mathrm{vol}(S)(1-h_S))}{2}. \end{aligned}$$

Thus, from the definitions and the convexity of f, we have

Lemma 2 *For a function $f : V \to \mathbb{R}$, we have, for $0 \le x \le \mathrm{vol}(G)/2$,*

$$f\mathbf{W}(x) \le \frac{f(x(1+h_f))+f(x(1-h_f))}{2}. \tag{41}$$

5. A proof of the Cheeger inequality using random walks

Here we will give a short proof for the following lemma which is essentially the mixing result of Lovász and Simonovits in [13, 14].

Lemma 3 *In a graph G, a subset S of vertices with $\mathrm{vol}(S) \le \mathrm{vol}(G)/2$ satisfies that for any vertex u,*

$$|\mathbf{W}^k(u,S)-\pi(S)| \le \sqrt{\frac{\mathrm{vol}(S)}{d_u}}(1-\frac{\beta_k^2}{8})^k \tag{51}$$

where β_k is the minimum Cheeger ratio over sets determined by the i largest values of the distribution of the lazy random walk starting at u after k steps. In other words,

$$\beta_k = \inf\{h_f : f = \chi_u \mathbf{W}^{k'} \text{ for } u \in V \text{ and } k' \le k\}.$$

Proof: For a fixed u, we choose $f(v) = f_k(v) = \mathbf{W}^k(u,v) - \pi(v)$ as in (41). We will prove (51) by induction on k. For the case of $k = 0$, it is easy to see that (51) holds. By Lemma 1, we have

$$\begin{aligned} f_{k+1}(x) &= f_k\mathbf{W}(x) \\ &\leq \frac{f_k(x(1+\beta_{k+1})) + f_k(x(1-\beta_{k+1}))}{2} \\ &\leq \sqrt{\frac{x}{d_u}}(1-\frac{\beta_k^2}{8})^k \frac{\sqrt{1+\beta_{k+1}} + \sqrt{1-\beta_{k+1}}}{2} \\ &\leq \sqrt{\frac{x}{d_u}}(1-\frac{\beta_{k+1}^2}{8})^{k+1}. \end{aligned}$$

Here we use the fact that $\sqrt{1+y} + \sqrt{1-y} \leq 2 - y^2/4$ for $0 \leq y \leq 1$. To finish the proof of Lemma 3, we consider $g(v) = g_k(v) = -f_k(v)$. We can prove in a similar way that

$$g_k(v) \leq \sqrt{\frac{\mathrm{vol}(S)}{d_u}}(1-\frac{\beta_k^2}{8})^k.$$

The proof can be done by induction on k and we note that it is true for $k = 0$ since $g_0(x) \leq 1$. This finishes the proof for (51). □

A special case of (51) is the following:

$$|\mathbf{W}^k(u,v) - \pi(v)| \leq \sqrt{\frac{d_v}{d_u}}(1-\frac{\beta_k^2}{8})^k. \tag{52}$$

Theorem 2 *For a graph G, we have the following Cheeger inequality:*

$$2h_G \geq \lambda_G \geq \frac{\beta_G^2}{8} \geq \frac{h_G^2}{8},$$

where

$$\beta_G = \min\{\beta_t : t \leq \lceil \frac{16\log n}{\lambda_G^2} \rceil\}.$$

Let a left eigenvector of $I - W$ associated with λ_G be denoted by ϕ. Note that ϕ is orthogonal to the right eigenvector $\mathbf{1}$. On one hand, we have

$$\|\phi(\mathbf{W}^t - \mathbf{1}^*\pi)D^{-1/2}\| = \|\phi\mathbf{W}^tD^{-1/2}\| = \big(1 - \frac{\lambda_G}{2}\big)^t\|\phi D^{-1/2}\|.$$

On the other hand, we have

$$\begin{aligned}
\|\phi(\mathbf{W}^t - \mathbf{1}^*\pi)D^{-1/2}\|^2 &= \sum_v \big(\phi(\mathbf{W}^t - \mathbf{1}^*\pi)(v)\frac{1}{\sqrt{d_v}}\big)^2 \\
&= \sum_v \frac{1}{d_v}\Big(\sum_u \phi(u)\big(\mathbf{W}^t(u,v) - \pi(v)\big)\Big)^2 \\
&\le \sum_v \frac{1}{d_v}\Big(\sum_u |\phi(u)|\ |\mathbf{W}^t(u,v) - \pi(v)|\Big)^2 \\
&\le \sum_v \frac{1}{d_v}\Big(\sum_u |\phi(u)|(1-\frac{\beta_t^2}{8})^t\sqrt{\frac{d_v}{d_u}}\Big)^2 \\
&\le (1-\frac{\beta_t^2}{8})^{2t}\sum_v n\big(\sum_u \frac{\phi(u)^2}{d_u}\big) \\
&= (1-\frac{\beta_t^2}{8})^{2t} n^2 \|\phi D^{-1/2}\|^2.
\end{aligned}$$

Together we have

$$1 - \frac{\lambda_G}{2} \le (1-\frac{\beta_t^2}{8})n^{1/t}.$$

This implies:

$$2h_G \ge \lambda_G \ge \frac{\beta_t^2}{4} - \frac{2\log n}{t}.$$

Therefore for

$$\beta_G = \min\{\beta_t : t \le \lceil \frac{16\log n}{\lambda_G^2}\rceil\},$$

we have

$$2h_G \ge \lambda_G \ge \frac{\beta_G^2}{8} \ge \frac{h_G^2}{8}$$

as desired. □

6. Proving the Cheeger inequality using the PageRank

The third proof of the Cheeger inequality is based on the notion of PageRank. The personalized PageRank has two parameters, the jumping constant c and a preference vector.

For a vertex u, the personalized PageRank pr_u associated with u has the preference vector χ_u and we write $\mathrm{pr}_u = \chi_u R_c$. Although pr_u is defined on the vertex set V of $G = (V, E)$, we can extend the domain of $f = \mathrm{pr}_u$ to $V \cup 2^V \cup E \cup 2^E \cup [0, \mathrm{vol}(G)]$ as defined in Section 4.

We will use Lemma 1 and (21) by choosing $f = \mathrm{pr}_u - \pi$. We have, for any set S of vertices,

$$f(S) = c(1-\pi(S)) + (1-c)\big(\frac{1}{2}f(S_{in}\cap S_{out}) + \frac{1}{2}f(S_{in}\cup S_{out})\big).$$

We will establish the following inequality with a simpler proof than that in [1]:

Lemma 4 *For any positive integer k, a jumping constant $c, 0 \le c \le 1$, a vertex u and a subset S of vertices, the personalized pagerank pr_u satisfies*

$$\mathrm{pr}_u(S) - \pi(S) \le \Big(1-(1-c)^k + \sqrt{\frac{\mathrm{vol}(S)}{d_u}}\big(1-\frac{\gamma_u^2}{8}\big)^k(1-c)^k\Big)(1-\pi(S)),$$

where γ_u is the minimum Cheeger ratio determined by a sweep of pr_u.

Proof: It suffices to show that for any real value $x, 0 \le x \le \mathrm{vol}(G)/2$, $f(x) = (\mathrm{pr}_u(x) - \pi(x))/(1-\pi(x))$ satisfies

$$f(x) \le 1-(1-c)^k + \sqrt{\frac{x}{d_u}}\big(1-\frac{\gamma_u^2}{8}\big)^k(1-c)^k. \qquad (61)$$

First we observe that (61) holds for the case that $k=0$ since, for $x \ge d_u$, we have $f(x) \le 1$, and for $x \le d_u$,

$$f(x) = \frac{x}{d_u}f(d_u) \le \sqrt{\frac{x}{d_u}}.$$

Suppose that (61) holds for some $k \ge 0$. We wish to establish the inequality for $k+1$ for $x = \mathrm{vol}(S_i)$.

$$\begin{aligned} f(x) &\le c + (1-c)f\mathbf{W}(x) \\ &\le c+(1-c)\big(\frac{f(x(1-\gamma_u)) + f(x(1+\gamma_u))}{2}\big) \qquad \text{by Lemma 1} \\ &\le c+(1-c)\big(1-(1-c)^k + \frac{(\sqrt{x(1-\gamma_u)}+\sqrt{x(1+\gamma_u)})}{2\sqrt{d_u}}(1-\frac{\gamma_u^2}{8})^k(1-c)^k\big) \\ &\qquad \text{by induction} \\ &\le 1-(1-c)^{k+1} + \sqrt{\frac{x}{d_u}}(1-\frac{\gamma_u^2}{8})^{k+1}(1-c)^{k+1} \end{aligned}$$

by using the fact that $\sqrt{1+w}+\sqrt{1-w} \le 2(1-w^2/8)$ for $0 \le w \le 1$. We have proved (61). To show that $|f(x)|$ has the same upper bound, we consider $g(x) = -f(x)$ and apply the same inductive proof. □

We also need the following for later lower bound arguments:

Lemma 5 *For a subset S of vertices in G with $\mathrm{vol}(S) \le \mathrm{vol}(G)/2$, there is a subset $T \subset S$ with $\mathrm{vol}(T) \ge \mathrm{vol}(S)/2$ such that for any $u \in T$, the personalized pagerank pr_u satisfies that*

$$\mathrm{pr}_u(S) \ge 1 - \frac{(1-c)h_S}{c}.$$

Proof: We first consider

$$\begin{aligned}\chi_S D R_c(S) &= \chi_S D(cI + (1-c)D\mathbf{W}R_c)\chi_S^* \\ &= c\mathrm{vol}(S) - \frac{1-c}{2}\chi_S(D-A)R_c\chi_S^* + (1-c)\chi_S D R_c(S).\end{aligned}$$

Therefore we have

$$\begin{aligned}\chi_S D R_c(S) &= \mathrm{vol}(S) - \frac{1-c}{2c}\chi_S(D-A)R_c\chi_S^* \\ &= \mathrm{vol}(S) - \frac{1-c}{2c}\sum_{u\sim v}(\chi_S(u)-\chi_S(v))(\chi_S R_c^*(u) - \chi_S R_c^*(v)) \\ &= \mathrm{vol}(S) - \frac{1-c}{2c}\sum_{\{u,v\}\in\partial S}|\chi_u R_c(S) - \chi_v R_c(S)| \\ &\ge \mathrm{vol}(S) - \frac{1-c}{2c}|\partial S|.\end{aligned}$$

Here we use the fact that $f(D-A)g^* = \sum_{u\sim v}(f(u)-f(v))(g(u)-g(v)$. This implies that

$$\begin{aligned}\chi_S D R_c(\bar{S}) &= \mathrm{vol}(S) - \chi_S D R_c(S) \\ &\le \frac{1-c}{2c}|\partial S|.\end{aligned}$$

We consider a subset T' of S defined by

$$T' = \{v \in S \;:\; \chi_v R_c(\bar{S}) \ge \frac{1-c}{c}h_S\}$$

and we have

$$\begin{aligned}\chi_S D R_c(\bar{S}) &\ge \chi_{T'} D R_c(\bar{S}) \\ &= \sum_{u\in T'} d_v \chi_v R_c(\bar{S}) \\ &\ge \sum_{v\in T'} d_v \frac{1-c}{c}h_S \\ &\ge \frac{1-c}{c}\mathrm{vol}(T')h_S.\end{aligned}$$

Therefore, we have $\mathrm{vol}(T') \le \mathrm{vol}(S)/2$. We define $T = S \setminus T'$. Thus for u in T, we have

$$\mathrm{pr}_u(S) = 1 - \chi_u R_c(\bar{S}) \ge 1 - \frac{1-c}{c} h_S$$

as claimed. □

We are now ready to prove the following local version of the Cheeger inequality using PageRank.

Theorem 3 *For a subset S in G with $\mathrm{vol}(S) \le \mathrm{vol}(G)/2$, there is a subset T with $\mathrm{vol}(T) \ge \mathrm{vol}(S)/2$ such that for any u in T, we have*

$$h_S \ge \frac{\gamma_u^2}{32 \log(\mathrm{vol}(S))}$$

where γ_u denotes the Cheeger ratio determined by the PageRank pr_u with jumping constant $c = \gamma_u^2/(16 \log(\mathrm{vol}(S)))$.

Proof: By combining Lemma 4 and Lemma 5, there is a subset T of S with $\mathrm{vol}(T) \ge \mathrm{vol}(S)/2$ such that for $u \in T$, we have

$$\begin{aligned} & 1 - \frac{(1-c)h_S}{c} - \pi(S) \\ \le & \left(1 - (1-c)^k + \frac{\mathrm{vol}(S)}{d_u}(1 - \frac{\gamma_u^2}{8})^k (1-c)^k\right)(1 - \pi(S)). \end{aligned}$$

This implies

$$\frac{h_S(1-c)}{c(1-\pi(S))} \ge (1-c)^k \big(1 - \frac{\mathrm{vol}(S)}{d_u}(1 - \frac{\gamma_u^2}{8})^k\big). \tag{62}$$

We then choose k and c as follows:

$$k = \lceil \frac{16 \log(\mathrm{vol}(S))}{\gamma_u^2} \rceil, \quad c = \frac{1}{k}.$$

Then (62) implies

$$h_S \ge \frac{c}{2} \ge \frac{\gamma_u^2}{32 \log(\mathrm{vol}(S))},$$

as desired. □

The above local Cheeger inequality suggests that by choosing one or more random vertices near the seed, the PageRank associated with these vertices can be used to find good cuts with high probability. Of course, the above inequality does not involve any eigenvector. This correlation can be added as follows:

For a given subset $S \subseteq V$, A function $f : V \to \mathbb{R}$ is said to satisfy the *Dirichlet boundary condition* if $f(v) = 0$ for all $v \notin S$. The Dirichlet eigenvalue, denoted by λ_S is defined by

$$\lambda_S = \inf_f R(f) \tag{63}$$

where f ranges over all nontrivial function f which satisfy the Dirichlet boundary condition for S.

Theorem 4 *For a subset S in G with* $\mathrm{vol}(S) \le \mathrm{vol}(G)/2$ *and a constant* $c \le 1/2$*, we have*

$$h_S \ge \lambda_S \ge \frac{\gamma_S^2}{8\log(\mathrm{vol}(S))}$$

where γ_S denotes the minimum of the Cheeger ratios determined by the PageRank pr_u *for u in S and the jumping constant* $c \ge \gamma_S^2/(8\log(\mathrm{vol}(S)))$.

Proof: Let S denote the subset achieving the Cheeger constant h_G. From the definition in (63), we have $h_G = R(\chi_S) \ge \lambda_S$. To prove the lower bound, we consider $\mathbf{W}_S$ which has entries $\mathbf{W}_S(u,v)$ the same as $\mathbf{W}(u,v)$ if u, v are both in S and 0 otherwise. We define

$$R'_c = c\sum_k (1-c)^k \mathbf{W}_S^k$$

so that

$$R'_c = cI_S + (1-c)R'_c\mathbf{W}_S. \tag{64}$$

Clearly, by Lemma 4, we have

$$\begin{aligned}\chi_u R'_c \chi_S &\le \chi_u R_c \chi_S \\ &\le 1 - (1-c)^k + \sqrt{\frac{\mathrm{vol}(S)}{d_u}}(1 - \frac{\gamma_S^2}{8})^k (1-c)^k.\end{aligned}$$

Let φ denote the function achieving λ_S and $\varphi\chi_S = \sum_{u\in S}\varphi(u) = 1$. It is not hard to see we can choose that $\varphi(u) \ge 0$ for all $u \in S$. Then,

$$\begin{aligned}\varphi R'_c \chi_S &\le \sum_{u\in S}\varphi(u)\chi_u R'_c\chi_S \\ &\le 1 - (1-c)^k + \sqrt{\mathrm{vol}(S)}(1 - \frac{\gamma_S^2}{8})^k(1-c)^k.\end{aligned}$$

Note that from (64), we have

$$\begin{aligned}\varphi R'_c\chi_S &= \varphi\big(I_S - \frac{1-c}{c}R_S(I_S - \mathbf{W}_S)\big)\chi_S \\ &= 1 - \frac{(1-c)\lambda_S}{c + (1-c)\lambda_S}.\end{aligned}$$

By combining the preceding two inequalities, we get

$$(1-c)^k \le \frac{(1-c)\lambda_S}{c+(1-c)\lambda_S} + \sqrt{\mathrm{vol}(S)}(1-\frac{\gamma_S^2}{8})^k(1-c)^k. \tag{65}$$

We can now select k so that

$$\sqrt{\mathrm{vol}(S)}(1-\frac{\gamma_S^2}{8})^k \le 1/2.$$

Namely, k is chosen to be $k = 4\gamma_S^{-2}\log(\mathrm{vol}(S))$.

Then (65) implies that

$$\frac{(1-c)^k}{2} \le \frac{(1-c)\lambda_S}{c+(1-c)\lambda_S}.$$

Therefore

$$ck \ge \log\frac{1}{1-c/((1-c)\lambda_S)} \ge \frac{c}{(1-c)\lambda_S}.$$

We then have

$$\begin{aligned}\lambda_S &\ge \frac{1}{(1-c)k}\\ &\ge \frac{\gamma_S^2}{8\log(\mathrm{vol}(S))}.\end{aligned}$$

The proof of Theorem 4 is complete. □

7. The Cheeger inequality using the heat kernel

For a vertex u, we define the heat kernel pagerank $\rho_{t,u}$ to be

$$\rho_{t,u}(v) = \chi_u H_t \chi_v^*$$

where H_t is as defined in Section 2. For a positive value s, we define a s-local Cheeger ratio of $\rho_{t,u}$ to be the minimum Cheeger ratio of cuts that separate sets S_i which consist of the vertices v that have the i largest values of $\rho_{t,u}(v)/d_v$ and so that $\mathrm{vol}(S_i)$ is at most $2s$.

We will use the following rapid mixing result on $\rho_{t,u}$ (see [7]): In a graph G, for a positive value $t \ge 0$, a vertex u and a subset S of volume $s \le \mathrm{vol}(G)/4$, we have

$$\rho_{t,u}(S) - \pi(S) \le \sqrt{\frac{s}{d_u}}e^{-t\kappa_{t,u,s}^2/4} \tag{76}$$

where $\kappa_{t,u,s}$ denotes the minimum s-local Cheeger ratio of $\rho_{t,u}$.

We will prove the following local Cheeger inequality that relates the Cheeger ratio h_S of a subset S with the Dirichlet eigenvalue λ_S and the local Cheeger ratios.

Theorem 5 *In a graph G, for a subset S of volume $s \leq \mathrm{vol}(G)^{2/3}$, we have*

$$h_S \geq \lambda_S \geq \frac{\kappa_S^2}{8}$$

where κ_S denotes the minimum s-local Cheeger ratio of $\rho_{t,u}$ over all u in S and $t = \lfloor \log(\mathrm{vol}(G)/s)/\lambda_S \rfloor$.

Proof: We consider a left eigenvector φ of $I - W$ which is associated with the Dirichlet eigenvalue λ_S and satisfies $\sum_{u\in S} \varphi(u) = 1$. By using (76), we have

$$\begin{aligned} \varphi(H_t - \mathbf{1}^*\pi)\chi_S^* &= \sum_{u\in S} \varphi(u)\rho_{t,u}(S) - \pi(S) \\ &\leq \sqrt{\frac{s}{d_u}} e^{-t\kappa_{t,S}^2/4}. \end{aligned}$$

We consider

$$H'_t = e^{-t} \sum_{k=0}^{\infty} \frac{t^k W_S^k}{k!}$$

where W_S is as defined in the previous section. It is not difficult to see that $H_t(u,v) \geq H'_t(u,v)$ for all u and v. We have

$$\begin{aligned} \varphi(H_t - \mathbf{1}^*\pi)\chi_S^* &\geq \varphi(H'_t - \mathbf{1}^*\pi)\chi_S^* \\ &= e^{-t\lambda_S} - \frac{s}{\mathrm{vol}(G)}. \end{aligned}$$

We choose t satisfying

$$e^{-t\lambda_S} \geq 2\frac{s}{\mathrm{vol}(G)}.$$

By considering $t \leq \log(\mathrm{vol}(G)/s)/\lambda_S$, we have

$$\begin{aligned} \frac{1}{2}e^{-t\lambda_S} &\leq \varphi(H_t - \mathbf{1}^*\pi)\chi_S^* \\ &\leq \sqrt{\frac{s}{d_u}} e^{-t\kappa_{t,S}^2/4}. \end{aligned}$$

This implies,

$$\begin{aligned} \lambda_S &\geq \frac{\kappa_{t,S}^2}{4} - \frac{\log s}{2t} \\ &\geq \frac{\kappa_{t,S}^2}{4} - \lambda_S \end{aligned}$$

if $t \geq \frac{\log s}{2\lambda_S}$. Such t exists if $s \leq \mathrm{vol}(G)^{2/3}$. This leads to

$$\lambda_S \geq \frac{\kappa_{t,S}^2}{8}$$

for $t = \lfloor \log(\mathrm{vol}(G)/s)/\lambda_S \rfloor$. The proof of Theorem 5 is complete. □

For applications, the following modified Cheeger inequality can be derived along the similar spirit of Theorem 4. Its proof is contained in [7] and will be omitted.

Theorem 6 *In a graph G for a subset of volume s, $s \leq \mathrm{vol}(G)/4$ and Cheeger ratio $h_S \leq \phi^2/4$, there is a subset $S' \subset S$ with $\mathrm{vol}(S') \geq s/2$ such that for any $u \in S'$, the sweep by using the heat kernel pagerank $\rho_{t,u}$, with $t = \lceil \phi^{-2}/4 \rceil$, will find a set T with s-local Cheeger ratio at most $\phi\sqrt{\log s}$.*

8. Graph partition algorithms

The four Cheeger inequalities and their constructive proofs lead to a number of graph partition algorithms. These algorithms are one-sweep algorithms so that the output is a subset consisting of vertices associated with the highest i values determined by some function (which sometimes is adjusted by the degrees of the vertices). Except for the spectral partition algorithm, all the other three methods, using random walks, PageRank and heat kernel pagerank yield local partition algorithms.

Before proceeding to discuss local partition algorithms, we shall mention the mode of computing. The input of all our algorithms of course includes the whole graph (which can be very large) which is computed *off-line*. For example, for search engines, such as Google, Yahoo or Baidu, the vertices and edges of the Webgraph are meticulously, exhaustively and continuously computed in an off-line mode and then compiled and stored (in some appropriate data structure which we will not get into here). When one uses the search engine by inputting a word or key phrase (as the seed), the search is in on-line mode since one expects the results of the search in a fraction of a second. The output of an on-line algorithm is usually quite small in comparison with the whole size of the data corpus. The local algorithms are all on-line algorithms which have running time proportional to the sizes of their outputs. In order to achieve this, the algorithmic steps for a local algorithm usually rely on local operations such as propagating through the neighbors iteratively.

The spectral partition algorithm.
The one sweep algorithm uses the eigenvector that achieves the spectral gap as stated in Theorem 1. The Cheeger inequality in (11) guarantees

that the output has its Cheeger ratio α_G which is no more than $2\sqrt{h_G}$. The running time of this algorithm is basically the same as the running time for matrix multiplication, of order $O(n^\omega)$, $\omega = 2.376\ldots$ (see [10]) where n is the number of vertices in G.

Graph partition algorithm using random walks
Spielman and Teng [15] gave a graph partition algorithm which, for a graph G containing a set of vertices such that $h_S \leq \Phi$ and $\mathrm{vol}(S) \leq \mathrm{vol}(G)/2$, can find a subset T with $\mathrm{vol}(T) \geq \mathrm{vol}(S)/2$ and $h_T \leq \phi$ where $\Phi = O(\phi^3/\log^2(\mathrm{vol}(G)))$. The partition algorithm contains quick approximations of distributions after steps of random walks. The algorithmic analysis uses the ideas of Lovász and Simonovits in [13, 14]. The running time of their algorithm is $O(\mathrm{vol}(G)\log^6(\mathrm{vol}(G))/\phi^5)$. A main subroutine NIBBLE in their algorithm has inputs including a seed, a target volume k and a target Cheeger ratio Φ. The Nibble, which runs in time $O(k\log^4(\mathrm{vol}(G))/\phi^5)$, outputs a subset T with $h_T \leq \phi$, where $\Phi = O(\phi^3/\log^2(\mathrm{vol}(G)))$, and the volume is within a factor of 2 of the target volume.

Graph partition algorithm using PageRank
The local partition algorithm of Andersen, Chung and Lang uses a fast approximation of PageRank [1]. The algorithm basically follows from the modified Cheeger inequality in Theorem 3 and improves the previous algorithm using random walks both in its performance and in running time. For any set S with Cheeger ratio Φ, there are a large number of starting vertices within S for which a sweep over an appropriate PageRank vector find a subset with Cheeger ratio $O(\sqrt{\Phi\log\mathrm{vol}(S)})$, in time $O(\mathrm{vol}(S)\log^2(\mathrm{vol}(G))/\Phi)$. The local graph partition algorithm can be used iteratively to derive a general graph partition algorithm which will find a cut which has Cheeger ratio at most $\sqrt{h_G\log(\mathrm{vol}(G))}$ in time $O(\mathrm{vol}(G)\log^4(\mathrm{vol}(G))/h_G)$. There is a second paper by the same authors [2] that further simplifies the local partition algorithm by focusing on cuts which are associated with "sharp" drops in the sorted PageRank functions. The algorithmic analysis is also simpler although the computational complexity stays of the same order.

Graph partition algorithm using heat kernel pagerank
The Cheeger inequalities in Theorems 5 and 6 suggest a graph partition algorithm in a straightforward way. This improves the previous algorithm by removing the factor of $\log(\mathrm{vol}(G))$ in the estimate of the Cheeger ratio, in comparison with the previous algorithm using PageRank. In order to obtain an even more efficient algorithm, we need to get an efficient approximation of the heat kernel pagerank which the formulation as an exponential sum can easily yield. For any set S with Cheeger ratio Φ, there are a large number of starting vertices within S

so that the associated heat kernel pagerank can be use to find a cut with Cheeger ratio at most $O(\sqrt{\Phi \log(\mathrm{vol}(S))})$ (see [7]). The running time is basically $O(\mathrm{vol}(S) \log(\mathrm{vol}(S))/\Phi)$ for approximating and sorting the heat kernel pagerank with a support no more than the target volume.

References

[1] R. Andersen, F. Chung and K. Lang, Local graph partitioning using pagerank vectors, *Proceedings of the 47th Annual IEEE Symposium on Founation of Computer Science (FOCS'2006)*, 475–486.

[2] R. Andersen, F. Chung and K. Lang, Detecting sharp drops in PageRank and a simplified local partitioning algorithm, *Theory and Applications of Models of Computation, Proceedings of TAMC 2007*, 1–12.

[3] S. Brin and L. Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems*, **30 (1-7)**, (1998), 107–117.

[4] J. Cheeger, A lower bound for the smallest eigenvalue of the Laplacian, Problems in Analysis (R. C. Gunning, ed.), Princeton Univ. Press (1970), 195–199.

[5] F. Chung, *Spectral Graph Theory*, AMS Publications, 1997.

[6] F. Chung, Random walks and local cuts in graphs, LAA **423** (2007), 22–32.

[7] F. Chung, The heat kernel as the pagerank of a graph, *PNAS*, to appear.

[8] F. Chung and S.-T. Yau, Coverings, heat kernels and spanning trees, *Electronic Journal of Combinatorics* **6** (1999), #R12.

[9] F. Chung and L. Lu, *Complex Graphs and Networks*, CBMS Regional Conference Series in Mathematics, 107, AMS Publications, RI, 2006. viii+264 pp.

[10] D. Coppersmith and S. Winograd, Matrix multiplication via arithmetic progressions, *J. Symbolic Comput.* **9** (1990), 251–280.

[11] M. Jerrum and A. J. Sinclair, Approximating the permanent, *SIAM J. Computing* **18** (1989), 1149-1178.

[12] R. Kannan and S. Vempala and A. Vetta, On clusterings: Good, bad and spectral, *JACM* **51** (2004), 497–515.

[13] L. Lovász and M. Simonovits, The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume, *31st IEEE Annual Symposium on Foundations of Computer Science*, (1990), 346–354.

[14] L. Lovász and M. Simonovits, Random walks in a convex body and an improved volume algorithm, *Random Structures and Algorithms* **4** (1993), 359–412.

[15] D. Spielman and S.-H. Teng, Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems, *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, (2004), 81–90.

[16] R. M. Schoen and S. T. Yau, *Differential Geometry*, International Press, Cambridge, Massachusetts, 1994 .