# A Generalized Cheeger Inequality

Ioannis Koutis[a], Gary Miller[b], Richard Peng[c]

[a] *Department of Computer Science, New Jersey Institute of Technology*
[b] *Computer Science Department, Carnegie Mellon University*
[c] *Cheriton Schoold of Computer Science, University of Waterloo*

**Abstract**

The generalized conductance $\phi(G, H)$ between two weighted graphs $G$ and $H$ on the same vertex set $V$ is defined as the ratio

$$\phi(G, H) = \min_{S \subseteq V} \frac{cap_G(S, \bar{S})}{cap_H(S, \bar{S})},$$

where $cap_G(S, \bar{S})$ is the total weight of the edges crossing from vertex set $S \subseteq V$ to $\bar{S} = V - S$. We show that the minimum generalized eigenvalue $\lambda(L_G, L_H)$ of the pair of Laplacians $L_G$ and $L_H$ satisfies

$$\phi(G, H) \geq \lambda(L_G, L_H) \geq \phi(G, H)\phi(G)/16,$$

where $\phi(G)$ is the standard conductance of $G$. A generalized cut that meets this bound can be obtained from the generalized eigenvector corresponding to $\lambda(L_G, L_H)$.

*Keywords:* Spectral graph theory, Generalized cuts, Cheeger inequality

## 1. Introduction

The discrete version of the Cheeger inequality [2] relates graph connectivity with the second eigenvalue of the normalized graph Laplacian [3]. It has been a driving force in spectral graph theory, algorithm design and machine learning (for example, see [4, 5, 6, 7, 8, 9, 10]). More recently, there have been

⋆This work has previously appeared as non-reviewed supplementary material in [1].
*URL:* ikoutis@njit.edu (Ioannis Koutis), glmiller@cs.cmu.edu (Gary Miller), y5peng@uwaterloo.ca (Richard Peng)

improvements to the basic inequality that take into account higher order spectral gaps [11], or extend it to inequalities reflecting multiway graph cuts [12].

The departure point of this article is the observation that the eigenvalues of the normalized Laplacian can be viewed as the *generalized eigenvalues* of a pair of graph Laplacians $(L_G, L_K)$, where $G$ is the given graph and $K$ is a complete weighted graph whose edge weights depend solely on the vertex degrees of $G$. In turn, the generalized eigenvalue problem $(L_G, L_K)$ is a relaxation of a simultaneous cut problem on $(G, K)$, known as the sparsest cut problem. In this work we present a generalization of the standard Cheeger inequality to arbitrary pairs of graphs $(G, H)$. The new inequality recovers, up to a constant factor, the original Cheeger inequality for the case when $H = K$. Up to our knowledge, a similar question was previously considered by Trevisan [13]; this is further discussed in Section 2.4.

## 2. Background and Definitions

Let $G = (V, E, w)$ be a weighted graph, where $V$ is the set of vertices $E \subseteq V \times V$ is the set of edges, and $w : E \to \mathbb{R}^+$ are positive weights on the edges. For $v \in V$ and $S \subseteq V$ we let

$$vol(v) = \sum_{(v,w)\in E} w(v,u) \qquad \text{and} \qquad vol(S) = \sum_{v\in S} vol(S).$$

In order to avoid trivial considerations we assume that for every $v$, we have $vol(v) > 0$. We also denote by $cap(S, \bar{S})$ the total weight of edges with exactly one endpoint in $S$ and one endpoint in $\bar{S} = V - S$. The **sparsity** of a cut $(S, \bar{S})$ is defined as

$$\phi_S(G) = \frac{cap(S, \bar{S})}{\min\{vol(S), vol(\bar{S})\}}.$$

The **conductance** of $G$ is defined as

$$\min_{\substack{S\subset V,\\ S\neq\emptyset}} \phi_S(G)$$

The Laplacian of $G$ is defined by

$$L(u, v) = -w(u, v) \qquad \text{and} \qquad L(u, u) = \sum_{v\neq u} L(u, v).$$

The normalized Laplacian $\tilde{L}$ of $G$ is the matrix $D^{-1/2}LD^{-1/2}$ where $D$ is the diagonal of $L$. It is well understood that the normalized Laplacian of a connected graph is positive semi-definite with a unique zero eigenvalue.

### *2.1. The standard Cheeger inequality and cut algorithm*

If $\lambda_2$ is the second eigenvalue of the normalized Laplacian, then the Cheeger inequality relates it to $\phi(G)$ as follows:

$$\lambda_2 \geq \phi(G)^2/2. \tag{1}$$

At least one proof of the Cheeger inequality, due to Mihail [14], actually shows something stronger. Namely, for any vector $y \perp Null(\tilde{L}_G)$, we can find a set $S_y$ such that

$$\phi_{S_y}(G) < 2(y^T\tilde{L}_G y)^{1/2}. \tag{2}$$

The cut can be found by letting $S_y$ consist of the vertices corresponding to the $k$ smallest entries of $D^{-1/2}y$, for some $1 \leq k \leq n$. In particular, by letting $y$ to be a standard approximation to the second eigenvector of the normalized Laplacian, then we have that $y^T\tilde{L}_G y = \Theta(\lambda(G))$. It thus follows that we can compute a cut with sparsity $O(\sqrt{\phi})$, in polynomial time. Some further algorithmic details are discussed in Section 4.

### *2.2. Generalized cuts for graph pairs*

We will now consider pairs of weighted graphs $(G, H)$ on the same vertex set $V$. We assume that $G$ is connected. We define the generalized sparsity of a cut $(S, \bar{S})$ as:

$$\phi_S(G, H) = \frac{cap_G(S, \bar{S})}{cap_H(S, \bar{S})}.$$

We define the **generalized conductance** between $G$ and $H$ as follows:

$$\phi(G, H) = \min_{\substack{S \subset V, \\ S \neq \emptyset}} \phi_S(G, H).$$

To see the utility of this definition, we observe that the sparsest cut problem can be captured within a factor of 2 as a generalized cut problem between two graphs. This is also known as the non-uniform sparsest cut problem. To this end let us define the **demand graph** $D_G = (V, E', w')$ with every edge being present in $E'$ and the weights specified by

$$w'(u,v) = \frac{vol(u)vol(v)}{vol(V)}.$$

Let $S \subseteq V$. Observe that by construction we have

$$cap_{D_G}(S, \bar{S}) = \frac{vol(S)vol(\bar{S})}{vol(V)}.$$

Note now that

$$\min\{vol(S), vol(\bar{S})\} \geq \frac{vol(S)vol(\bar{S})}{vol(V)} \geq \min\{vol(S), vol(\bar{S})\}/2.$$

From this it can be seen that

$$\frac{\phi(G)}{2} \leq \phi(G, D_G) \leq \phi(G). \tag{3}$$

A number of other problems can be viewed as generalized cut problems.

For example, consider the **isoperimetric number** defined by:

$$h(G) = \min_{\substack{S \subset V,\\ S \neq \emptyset}} \frac{cap_G(S, \bar{S})}{\min\{|S|, |\bar{S}|\}}.$$

If $K_n$ is the complete graph on $n$ vertices with edges weighted by $1/n$, i.e. the identity over the space of sets orthogonal to the constant vectors, it can be verified that we have

$$\frac{\phi(G, K_n)}{2} \leq h(G) \leq \phi(G, K_n).$$

Another example is the min $s$-$t$ cut problem which looks for a cut of minimum value among all possible cuts that separate $s$ and $t$. If we denote that value by $\mu_{s,t}$, and let $G_{s,t}$ be the Laplacian of the edge $(s,t)$, we have

$$\mu_{s,t} = \phi(G, G_{s,t}).$$

*2.3. The minimum generalized eigenvalue of a pair of Laplacians*

Let $(G, H)$ be a pair of graphs, where $G$ is connected. Let $\mathbf{1}$ be the constant vector. It is well understood that

$$\mathbf{1} = Null(L_G) \subseteq Null(L_H).$$

Hence, we consider the generalized eigenvalue problem

$$L_G x = \lambda L_H x, \tag{4}$$

where $x$ is constrained to satisfy $x^T \mathbf{1} = 0$. Then, equation 4 is equivalent to

$$L_G^+ L_H x = \lambda^{-1} x \Rightarrow My = \lambda^{-1} y,$$

where $M = (L_G^+)^{1/2} L_H (L_G^+)^{1/2}$ and $y = L_G^{1/2} x$. We have $\mathbf{1} \in Null(M)$. Since $M$ is symmetric, it follows that $M$ has a maximum eigenvalue $\lambda^{-1}$ with a corresponding eigenvector $y$, such that $y^T \mathbf{1} = 0$, which in turn implies that $y$ can be written as $L_G^{1/2} x$ for some vector $x$ satisfying the constraint $x^T \mathbf{1} = 0$. Thus, under that constraint, there is a minimum $\lambda$ that satisfies equation 4. Let $\lambda(G, H)$ denote that minimum eigenvalue. By an application of the Courant-Fisher theorem [15] on $M$ we get

$$\lambda(G, H) = \min_{x^T \mathbf{1} = 0} \frac{x^T L_G x}{x^T L_H x}. \tag{5}$$

Let now $d$ be the vector containing the degrees of the vertices in $G$. If $x$ is any vector, the map $y = x - \frac{x^T d}{(\mathbf{1}^T d)} \cdot \mathbf{1}$ satisfies $y^T d = 0$. The map is clearly invertible. This implies that there is a 1-1 map between vectors $x$ with $x^T \mathbf{1} = 0$ and vectors $y$ with $y^T d = 0$. Furthermore, for each pair $(x, y)$ we have $x^T L x = y^T L y$ for any Laplacian matrix $L$, because $(y - x)$ is in the null space of $L$. We thus have

$$\lambda(G, H) = \min_{y^T \mathbf{d} = 0} \frac{y^T L_G y}{y^T L_H y}. \tag{6}$$

### *2.4. Cuts and Eigenvalues*

The value of a cut between $S$ and $\bar{S}$ can be expressed in terms of the graph Laplacian as:

$$cap_G(S, \bar{S}) = x_S^T L_G x_S,$$

where $x_S$ is characteristic vector of $S$, i.e. the vector with ones in its entries corresponding to $S$ and zeros in all other entries. It follows that the generalized conductance can be expressed as an optimization problem over the discrete 0-1 vectors:

$$\phi(G, H) = \min_{\substack{x \in \{0,1\}^n \\ x^T \mathbf{1} \neq 0}} \frac{x^T L_G x}{x^T L_H x}.$$

**Theorem 1.** *We have* $\lambda(G,H) \leq \phi(G,H)$.

*Proof.* Let $x = \{0,1\}^n$ and $y = x - \frac{1}{n}(x^T\mathbf{1})\mathbf{1}$. Note that for any Laplacian $L$ we have $x^T L x = y^T L y$ because $y - x$ is in the null space of $L$. Because $y^T\mathbf{1} = 0$, by equation 5, we have

$$\lambda(G,H) \leq \frac{y^T L_G y}{y^T L_H y} = \frac{x^T L_G x}{x^T L_H x}.$$

The claims follows by letting $x$ be the characteristic vector of the cut attaining $\phi(G,H)$. □

**Remark-1:** The eigenvalue $\lambda(G,H)$, as expressed in 5, can be viewed as a relaxation of $\phi(G,H)$ over the reals.

The minimum eigenvalue $\lambda_2$ of the normalized Laplacian of $G$ is equal to the minimum eigenvalue of the generalized problem $L_G x = \lambda D x$, under the constraint $x^T d = 0$, where $d$ is the vector containing the degrees of the vertices in $G$. Then, due to Lemma 2, $\lambda_2$ is equal to $\lambda(G, D_G)$ and thus it can be seen as a relaxation of $\phi(G, D_G)$ which is within a factor of 2 from $\phi(G)$ (equation 3). Thus the Cheeger inequality characterizes $\phi(G, D_G)$ in terms of $\lambda(G, D_G)$. We aim to prove a similar characterization for the generalized conductance of any pair of graphs.

**Remark-2:** In [13], Trevisan asked whether the Cheeger inequality can be extended to the generalized cut on a pair of graphs $(G,H)$, for arbitrary $H$. They showed that under a complexity assumption known as the Unique Games Conjecture, it is impossible to find a cut of sparsity $O(\sqrt{\phi(G,H)})$ in polynomial time. That indicates that an analog of the Cheeger inequality and the associated algorithm do not exist. Our result is compatible with Trevisan's work, as it provides a different type of bound.

## 3. Generalized Cheeger Inequality

We now present and prove our main theorem; the proof is based on lemmas that are proved separately, in Section 3.1.

**Theorem 2.** *Let $G$ and $H$ be any two weighted graphs and $d$ be the vector containing the degrees of the vertices in $G$. For any vector $x$ such that $x^T d = 0$, we have*

$$\frac{x^T L_G x}{x^T L_H x} \geq \phi(G, D_G) \cdot \phi(G, H)/8,$$

*where $D_G$ is the demand graph of $G$. If we let $x$ be an eigenvector corresponding to the minimum eigenvalue $\lambda(G, H)$ we get that*

$$\lambda(G, H) \geq \phi(G, D_G) \cdot \phi(G, H)/8.$$

We first introduce auxiliary notation. Let $V^-$ denote the set of $u$ such that $x_u \leq 0$ and $V^+$ denote the set such that $x_u > 0$. Then we can divide $E_G$ into two sets: $E_G^{same}$ consisting of edges with both endpoints in $V^-$ or $V^+$, and $E_G^{dif}$ consisting of edges with one endpoint in each. We also define $E_H^{dif}$ and $E_H^{same}$ similarly.

*Proof.* Let

$$S_G = \sum_{uv \in E_G^{same}} w_G(u, v) \left| x_u^2 - x_v^2 \right| + \sum_{uv \in E_G^{dif}} w_G(u, v) \left( x_u^2 + x_v^2 \right). \tag{7}$$

and

$$\begin{aligned} A &= \sum_{uv \in E_G^{same}} w_G(u, v)(x_u - x_v)^2 + \sum_{uv \in E_G^{dif}} w_G(u, v)(x_u^2 + x_v^2). \\ B &= \sum_{uv \in E_G^{same}} w_G(u, v)(x_u + x_v)^2 + \sum_{uv \in E_G^{dif}} w_G(u, v)(x_u^2 + x_v^2). \end{aligned}$$

We define two vectors on the edges of $G$:

- $u_A(u, v) = \sqrt{w_G(u, v)} |x_u - x_v|$ if $uv \in E_G^{same}$
- $u_A(u, v) = \sqrt{w_G(u, v)(x_u^2 + x_v^2)}$ if $uv \in E_G^{dif}$

and similarly

- $u_B(u, v) = \sqrt{w_G(u, v)} |x_u + x_v|$ if $uv \in E_G^{same}$
- $u_B(u, v) = \sqrt{w_G(u, v)(x_u^2 + x_v^2)}$ if $uv \in E_G^{dif}$

We have $A = \langle u_A, u_A \rangle$ and $B = \langle u_B, u_B \rangle$. By the Cauchy-Schwarz inequality we have

$$AB \geq \langle u_A, u_B \rangle^2.$$

This gives that:

$$AB \geq \left( \sum_{uv \in E_G^{same}} w_G(u,v)|x_u^2 - x_v^2| + \sum_{uv \in E_G^{dif}} w_G(u,v)(x_u^2 + x_v^2) \right)^2 = S_G^2. \quad (8)$$

We also have

$$\begin{aligned} x^T L_G x &= \sum_{uv \in E_G} w_G(u,v)(x_u - x_v)^2 \\ &= \sum_{uv \in E_G^{same}} w_G(u,v)(x_u - x_v)^2 + \sum_{uv \in E_G^{dif}} w_G(u,v)(x_u - x_v)^2 \\ &\geq A \end{aligned} \quad (9)$$

The last inequality follows by $x_u x_v \leq 0$ as $x_u \leq 0$ for all $u \in V^-$ and $x_v \geq 0$ for all $v \in V^+$. We also have $(x_u + x_v)^2 \leq 2x_u^2 + 2x_v^2$, since $2x_u^2 + 2x_v^2 - (x_u + x_v)^2 = (x_u - x_v)^2 \geq 0$. Thus, we get

$$\begin{aligned} B \leq& 2 \left( \sum_{uv \in E_G^{same}} w_G(u,v)(x_u^2 + x_v^2) + \sum_{uv \in E_G^{dif}} w_G(u,v)(x_u^2 + x_v^2) \right) \\ &= 2x^T D x = 2x^T L_{D_G} x, \end{aligned} \quad (10)$$

where $D$ is the diagonal of $L_G$ and the last equality comes from Lemma 2 and the assumption that $x^T d = 0$.

Applying inequalities 9, 10 and 8 we get

$$\frac{x^T L_G x}{x^T L_H x} \geq \frac{A}{x^T L_H x} \geq \frac{1}{2} \cdot \frac{A}{x^T L_H x} \cdot \frac{B}{x_{D_G}^L x} \geq \frac{1}{2} \cdot \frac{S_G}{x^T L_H x} \cdot \frac{S_G}{x^T L_{D_G} x} \quad (11)$$

Finally with two applications of Lemma 3 on the pairs graph $(G, H)$ and $(G, D_G)$, we have

$$\frac{x^T L_G x}{x^T L_H x} \geq \frac{1}{2} \cdot \frac{S_G}{x^T L_H x} \cdot \frac{S_G}{x_{D_G}^L x} \geq \frac{1}{8} \phi(G,H) \phi(G, D_G). \quad (12)$$

□

**Remark-1:** By setting $H = D_G$ and using equation 3, we get that

$$\lambda_2 = \lambda(G, D_G) \geq \phi(G, D_G)^2/8 \geq \phi(G)^2/32,$$

which recovers the original Cheeger inequality up to a factor of 16.

**Remark-2:** Let $G$ be the cycle graph on $n$ vertices, and $K$ be the complete graph on $n$ vertices. Using the well-known fact that $\lambda_2(G) = \Theta(1/n^2)$ we have

$$\lambda_2(G) = \min_{x^T \mathbf{1}=0} x^T L_G x = \min_{x^T \mathbf{1}=0} n \cdot \frac{x^T L_G x}{x^T L_K x} = \Theta(1/n^2),$$

where we used the identity $x^T L_K x = n$ for all vectors $x$ orthogonal to $\mathbf{1}$. It thus follows that

$$\lambda(G, K) = \min_{x^T \mathbf{1}=0} \frac{x^T L_G x}{x^T L_K x} = \Theta(1/n^3)$$

We also have $\phi(G, K) = \Theta(1/n^2)$, $\phi(G, D_G) = \Theta(1/n)$. It follows that the generalized Cheeger inequality is tight up to constants even when $H = K$.

### *3.1. Lemmas*

We present and prove lemmas used in the proof of Theorem 2.

**Lemma 1.** *For all $a_i, b_i > 0$ we have*

$$\frac{\sum_i a_i}{\sum_i b_i} \geq \min_i \left\{ \frac{a_i}{b_i} \right\}.$$

**Lemma 2.** *Let $G$ be a graph, $d$ be the vector containing the degrees of the vertices, and $D$ be corresponding diagonal matrix. For all vectors $x$ where $x^T d = 0$ we have*

$$x^T D x = x^T L_{D_G} x,$$

*where $D_G$ is the demand graph for $G$.*

*Proof.* Let $d$ be the vector consisting of the entries along the diagonal of $D$. By definition, we have

$$L_{D_G} = D - \frac{dd^T}{vol(V)}.$$

The lemma follows. □

The following lemma is similar to one used in the proof of Cheeger's inequality [3]:

**Lemma 3.** *Let $G$ and $H$ be any two weighted graphs on the same vertex set $V$ partitioned into $V^-$ and $V^+$. For any vector $x$ we have*

$$\frac{S_G}{x^T L_H x} \geq \frac{\phi(G,H)}{2}, \tag{13}$$

*where, as defined in equation 7,*

$$S_G = \sum_{uv \in E_G^{same}} w_G(u,v) \left| x_u^2 - x_v^2 \right| + \sum_{uv \in E_G^{dif}} w_G(u,v) \left( x_u^2 + x_v^2 \right).$$

*Proof.* We begin with two inequalities:

Note that $2x_u^2 + 2x_v^2 - (x_u - x_v)^2 = (x_u + x_v)^2 \geq 0$ gives:

$$(x_u - x_v)^2 \leq 2x_u^2 + 2x_v^2.$$

Also, suppose $uv \in E_H^{same}$ and without loss of generality that $|x_u| \geq |x_v|$. Then letting $y = |x_u| - |x_v|$, we get:

$$\begin{aligned} |x_u^2 - x_v^2| &= (|x_v| + y)^2 - |x_v|^2 \\ &= y^2 + 2y|x_v| \\ &\geq y^2 = (x_u - x_v)^2. \end{aligned}$$

The last equality follows because $x_u$ and $x_v$ have the same sign.

We then use the above inequalities to upper bound the $x^T L_H x$ term.

$$\begin{aligned} x^T L_H x &= \sum_{uv \in E_H^{same}} w_H(u,v) (x_u - x_v)^2 + \sum_{uv \in E_H^{dif}} w_H(u,v) (x_u - x_v)^2 \\ &\leq \sum_{uv \in E_H^{same}} w_H(u,v) |x_u^2 - x_v^2| + \sum_{uv \in E_H^{dif}} w_H(u,v) \left( 2x_u^2 + 2x_v^2 \right) \\ &\leq 2 \left( \sum_{uv \in E_H^{same}} w_H(u,v) |x_u^2 - x_v^2| + \sum_{uv \in E_H^{dif}} w_H(u,v) \left( x_u^2 + x_v^2 \right) \right) \\ &= 2S_H \end{aligned} \tag{14}$$

Here $S_H$ is analogous to the quantity $S_G$ defined for $G$ in the Lemma statement. We thus have:

$$\frac{S_G}{x^T L_H x} \geq \frac{S_G}{2S_H} \tag{15}$$

We can now decompose the sum $S_G$ further into parts for $V^-$ and $V^+$:

$$\begin{aligned} S_G &= \sum_{uv \in E_G^{same}} w_G(u,v) \left| x_u^2 - x_v^2 \right| + \sum_{uv \in E_G^{dif}} w_G(u,v) \left( x_u^2 + x_v^2 \right) \\ &= N_G + P_G \end{aligned}$$

where we set

$$N_G = \sum_{u \in V^-, v \in V^-} w_G(u,v) \left| x_u^2 - x_v^2 \right| + \sum_{u \in V^-, v \in V^+} w_G(u,v) \, x_u^2$$

$$P_G = \sum_{u \in V^+, v \in V^+} w_G(u,v) \left| x_u^2 - x_v^2 \right| + \sum_{u \in V^-, v \in V^+} w_G(u,v) \, x_v^2.$$

We similarly write $S_H = N_H + P_H$ and by applying Lemma 1 we get:

$$\frac{S_G}{S_H} = \frac{N_G + P_G}{N_H + P_H} \geq \min \left\{ \frac{N_G}{N_H}, \frac{P_G}{P_H} \right\}$$

By symmetry in $V^-$ and $V^+$, it suffices to show that

$$\frac{N_G}{N_H} \geq \phi(G,H). \tag{16}$$

Let $V^- = \{u_1, \ldots, u_k\}$ and without loss of generality assume that

$$x_{u_1} \leq x_{u_2} \leq \ldots \leq x_{u_k} \leq 0.$$

Let $r_t = x_{u_t}^2 - x_{u_{t+1}}^2$, for $t = 1, \ldots, k-1$, and $r_k = x_k^2$. Also, let $S_i$ denote the set of nodes $\{u_1, \ldots, u_i\}$.

Consider now a term $|x_{u_i}^2 - x_{u_j}^2|$ where $x_{u_i} \leq x_{u_j}$. We can re-write it as

$$w_G(u_i,u_j)|x_{u_i}^2 - x_{u_j}^2| = w_G(u_i,u_j)(x_{u_i}^2 - x_{u_j}^2) = w_G(u_i,u_j) \sum_{t=i}^{j-1} r_t.$$

Similarly for a term $x_{u_i}^2$ associated with an edge from $u_i \in V^-$ to $v \in V^+$ we have

$$w_G(u_i,v) x_{u_i}^2 = w_G(u_i,v) \sum_{t=i}^{k} r_t.$$

We re-write every term of $N_G$ as suggested above. It can be seen that $r_i$ will appear multiplied by $w_G(e)$ for each edge $e$ whose one endpoint is in $S_i$ and the other endpoint in $V - S_i$. Then the coefficient of $r_i$ in $L_G$ will be equal to $cap(S_i, V - S_i)$. It follows that we have

$$N_G = \sum_{i=1}^{k} cap_G(S_i, \bar{S}_i) r_i$$

and similarly for $H$

$$N_H = \sum_{i=1}^{k} cap_H(S_i, \bar{S}_i) r_i.$$

By applying Lemma 1 and the definition of $\phi(G, H)$, we have

$$\frac{N_G}{N_H} = \frac{\sum_{i=1}^{k} cap_G(S_i, \bar{S}_i) r_i}{\sum_{i=1}^{k} cap_H(S_i, \bar{S}_i) r_i} \geq min_i \frac{cap_G(S_i, \bar{S}_i)}{cap_H(S_i, \bar{S}_i)} \geq \phi(G, H).$$

This proves equation 16. Then by substituting in inequality 15 the Lemma follows. □

## 4. Computation

In this section we –somewhat informally– discuss the computation of an approximation to the minimum cut for the pair $(G, H)$. To simplify our notation let us denote $L_G$ and $L_H$ by $A$ and $B$ respectively.

Suppose $x$ is an arbitrary vector not in the null space of $B$. Let $S_{x,i}$ be the set of nodes $u$ such that $x(u)$ is among the $i$ smallest entries of $x$. The combination of Lemmas 3 implicitly show that

$$\min_i \frac{cap_G(S_{x,i}, \bar{S}_{x,i})}{cap_H(S_{x,i}, \bar{S}_{x,i})} \leq \frac{8}{\phi(G, D_G)} \cdot \frac{x^T A x}{x^T B x}.$$

That means that given $x$ one can compute a cut with sparsity at most

$$\frac{x^T A x}{x^T B x} \cdot \frac{8}{\phi(G, D_G)}$$

by sorting $x$, and then sweeping $x$ for computing the smallest of the $n - 1$ generalized cuts defined by $x$, exactly as in the case of the standard Cheeger inequality.

To obtain the best possible approximation within this context, we would like to minimize $(x^T Ax/x^T Bx)$; it is well understood that the minimizer of this Rayleigh ratio is the associated eigenvector $y$. This suggests, similar to the discussion in Section 2.1, that we can find in polynomial time a cut $(S, \bar{S})$ which is at most $1/\phi(G, D_G)$ larger than the ratio $(x^T Ax/x^T Bx)$.

**Faster approximate computation.** We say that $x$ is an $(1+\epsilon)$-approximate eigenvector if it satisfies

$$\frac{x^T Ax}{x^T Bx} \leq (1+\epsilon)\lambda_{\min}(A, B). \tag{17}$$

The computation of an approximate eigenvector can be done in near-linear time. We informally describe the steps. Given any positive definite matrix $A$, one can use the inverse power iteration $y_{i+i} = A^{-1}y_i$, where $y_0$ is a random vector, to find a vector $x$ such that

$$\frac{x^T Ax}{x^T x} \leq (1+\epsilon)\lambda_{\min}(A). \tag{18}$$

The number of rounds required for this is $O(\log n/\epsilon)$; for a proof see [16]. Analogously, given a pair of positive definite matrices $(A, B)$, one can perform power iteration with the matrix $A^{-1}B$ to find a vector $x$ such that

$$\frac{x^T Ax}{x^T Bx} \leq (1+\epsilon)\lambda_{\min}(A, B).$$

The proof is similar to the simple eigenvalue problem case, using only the additional fact that the generalized eigenvectors of the pair $(A^{-1}, B^{-1})$ are the usual eigenvectors of the matrix $A^{-1}B$, in addition with the fact that the eigenvectors are $A$-orthogonal and $B$-orthogonal [15]. Note that the iteration can be extended to the case when $A$ has a known null space (as in the case of Laplacians), by simply operating on vectors orthogonal to the null space.

Additionally observe that each step of power iteration $A^{-1}By_i$ can be implemented as a linear system solve $Az = By_i$. Instead of solving exactly a linear system with the Laplacian $A$, one can use a more efficient iterative solver, and compute a solution $\tilde{z}$ that satisfies $||\tilde{z} - z||_A \leq (1+\epsilon/4)||A^{-1}y_i||_A$. Using fast Laplacian solvers, this can be computed in time near-linear time [17]. In such solvers, the approximate solution of a linear system $Ay = b$ implements implicitly a matrix-vector multiplications $\tilde{A}^{-1}y$, where $\tilde{A}^{-1}$ is spectrally close to $A^{-1}$. Spielman and Teng [16] observe that this is sufficient for the computation of an approximate eigenvector that satisfies inequality 18. This extends to the generalized problem with Laplacians. Finally, a little more care has to be taken for the case of Laplacian solvers that are randomized. In that case, $O(\log(1/p))$ different runs of the inverse power method are needed to get a good approximate eigenvector with probability at least $1 - p$. Overall, with the use of fast Laplacian solvers [17], the running time required to compute a 2-approximate eigenvector is $O(n \log^2 n \log(1/p))$, where $n$ is the number of non-zero entries in $A$ and $B$.

## 5. Acknowledgements

I. Koutis was partially supported by NSF awards #1813374 and #2039863. G. Miller was partially supported by NSF award #1637523. R. Peng was partially supported by NSF award #1846218.

## References

[1] M. Cucuringu, I. Koutis, S. Chawla, G. Miller, R. Peng, Simple and scalable constrained clustering: a generalized spectral method, Vol. 51 of Proceedings of Machine Learning Research, PMLR, Cadiz, Spain, 2016, pp. 445–454.
URL `http://proceedings.mlr.press/v51/cucuringu16.html`

[2] J. Cheeger, A lower bound for the smallest eigenvalue of the Laplacian, in: Proceedings of the Princeton conference in honor of Professor S. Bochner, 1969, pp. 195–199.

[3] F. Chung, Spectral Graph Theory, Vol. 92 of Regional Conference Series in Mathematics, American Mathematical Society, 1997.

[4] B. Mohar, Difference equations, isoperimetric inequality and transience of certain random walks, Trans. Am. Math. Soc 284 (1984) 787–794.

[5] N. Alon, V. D. Milman, $\lambda_1$, isoperimetric inequalities for graphs, and superconcentrators, Journal of Combinatorial Theory, Series B 38 (1) (1985) 73–88. doi:10.1016/0095-8956(85)90092-9.

[6] N. Alon, Eigenvalues and expanders, Combinatorica 6 (2) (1986) 83–96. doi:10.1007/BF02579166.

[7] C. Thomassen, Isoperimetric Inequalities and Transient Random Walks on Graphs, Ann. Probab. 20 (3) (1992) 1592–1600. doi:10.1214/aop/1176989708.
URL https://doi.org/10.1214/aop/1176989708

[8] J. Tan, On Cheeger Inequalities of a Graph, Discrete Math. 269 (1–3) (2003) 315–323. doi:10.1016/S0012-365X(03)00127-4.
URL https://doi.org/10.1016/S0012-365X(03)00127-4

[9] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. on Pattern Analysis and Machine Intelligence 22 (8) (2000) 888–905. doi:10.1109/34.868688.

[10] U. Luxburg, A Tutorial on Spectral Clustering, Statistics and Computing 17 (4) (2007) 395–416. doi:10.1007/s11222-007-9033-z.
URL https://doi.org/10.1007/s11222-007-9033-z

[11] T. C. Kwok, L. C. Lau, Y. T. Lee, S. Oveis Gharan, L. Trevisan, Improved Cheeger's Inequality: Analysis of Spectral Partitioning Algorithms through Higher Order Spectral Gap, in: Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing, STOC '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 11–20. doi:10.1145/2488608.2488611.
URL https://doi.org/10.1145/2488608.2488611

[12] J. R. Lee, S. O. Gharan, L. Trevisan, Multiway Spectral Partitioning and Higher-Order Cheeger Inequalities, J. ACM 61 (6). doi:10.1145/2665063. URL https://doi.org/10.1145/2665063

[13] L. Trevisan, Is Cheeger-type Approximation Possible for Nonuniform Sparsest Cut?, CoRR abs/1303.2730.
URL http://arxiv.org/abs/1303.2730

[14] M. Mihail, Conductance and convergence of markov chains-a combinatorial treatment of expanders, in: 30th Annual Symposium on Foundations of Computer Science, Research Triangle Park, North Carolina, USA, 30 October - 1 November 1989, IEEE Computer Society, 1989, pp. 526–531. doi:10.1109/SFCS.1989.63529.
URL http://doi.ieeecomputersociety.org/10.1109/SFCS.1989.63529

[15] G. Stewart, J.-G. Sun, Matrix Perturbation Theory, Academic Press, Boston, 1990.

[16] D. A. Spielman, S. Teng, Nearly-linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems, SIAM J. Matrix Anal. & Appl. 35(3) (2014) 835–885.

[17] I. Koutis, G. L. Miller, R. Peng, A fast solver for a class of linear systems, Commun. ACM 55 (10) (2012) 99–107. doi:10.1145/2347736.2347759. URL https://doi.org/10.1145/2347736.2347759