The Buser Inequality and the Arora-Rao-Vazirani Algorithm

Arpon Basu Anand Brahmbhatt Rohit Agarwal

December 5, 2024

Abstract

In this project, we'll discuss connections between the conductance of a graph and the spectrum of its Laplacian. Of course, the tightest inequality for general graphs is Cheeger's inequality, which we'll first state and prove. We'll then present a "converse" to Cheeger's inequality for the special case of abelian Cayley graphs (which is nevertheless a very rich class of graphs), called *Buser's inequality*. In the process of proving Buser's inequality, we shall highlight the connection of this inequality to the Arora-Rao-Vazirani algorithm.

Contents

1 Notation and Preliminaries

Definition 1.1. *Let G be a connected graph on n vertices with adjacency matrix A and degree matrix D, where the degree matrix is a diagonal matrix with the degrees of the vertices of G on its diagonal. The normalized Laplacian of G is defined to be $L := I - D^{-1/2}AD^{-1/2}$, which simplifies to $I - A/d$ if G is a d-regular graph.*

From now on we'll only talk about d-regular graphs for simplicity, but all of the statements extend to non-regular graphs with some extra work. Note that if G is a d-regular graph, then A/d is the transition matrix for the random walk on G. Consequently, all the eigenvalues of A/d lie in $[-1, 1]$, and 1 is an eigenvalue of A/d (with the eigenvector being the all ones vector). Moreover, if G is connected, then 1 is a unique eigenvalue of A/d. Consequently, all eigenvalues of L lie in $[0, 2]$, and if G is connected then 0 is a unique eigenvalue of L.

Definition 1.2. *Let $0 = \lambda_1 < \lambda_2 \leqslant \lambda_3 \leqslant \cdots \leqslant \lambda_n \leqslant 2$ be the eigenvalues of L. λ_2 is also known as the* ***(one-sided) spectral gap*** *of G.*

At this point, we also recall the so-called **variational characterization** of λ_2.

Theorem 1.1.

$$\lambda_2 = \min_{\substack{\langle x, \mathbf{1} \rangle = 0 \\ x \neq 0}} R(x)$$

where $R(x) := x^{\mathsf{T}} L x / x^{\mathsf{T}} x$ is the so-called ***Rayleigh quotient*** *of L, a scale-invariant version of the quadratic form associated to L.*

Definition 1.3 (Conductance)**.**

- *For any non-empty set $S \subseteq [n]$, its* ***conductance*** *is $\phi(S) := e(S, \overline{S})/d|S|$* [1].
- *Given a "cut" (A, B) of G (i.e. $A \sqcup B = V(G)$), is* ***conductance*** *is $\phi(A)$ if $|A| \leqslant |B|$, and $\phi(B)$ otherwise.*
- *The* ***conductance*** *of graph G is $\phi(G) := \min_{|S| \leq n/2} \phi(S)$.*

Note that $d|S|$ is the total number of edges incident on all the vertices in S, and thus in some sense is the "volume" of the set S. On the other hand, $e(S, \overline{S})$, which denotes the number of edges between S and $\overline{S}$, is the "surface area" of S. Thus, a low conductance set can be physically interpreted as a set escaping whom [2] is difficult. For general graphs, calculating $\phi(G)$ is a NP-hard task. In designing approximation algorithms to calculate ϕ (and the S which realizes that ϕ), the related notion of "sparsest cut" is very useful: For any non-empty $S \subsetneq [n]$, define:

$$\psi(S) := \frac{e(S, \overline{S})}{\frac{d}{n}|S| \cdot |\overline{S}|}$$

It's not hard to see that for $|S| \leqslant n/2$, $\phi(S) \leqslant \psi(S) \leqslant 2\phi(S)$, and $\psi(S) = \psi(\overline{S})$, and thus obtaining an C-approximation to $\psi(G) := \min_{|S| \leqslant n/2} \psi(S)$ automatically yields a $2C$-approximation to $\phi(G)$.
We also introduce the Arora-Rao-Vazirani (ARV henceforth) relaxation for the sparsest cut problem:

$$\begin{array}{lll} \text{minimize} & \dfrac{\sum_{\{i,j\} \in E} \|x_i - x_j\|^2}{\frac{d}{n} \sum_{i,j} \|x_i - x_j\|^2} & \\ \text{subject to} & x_i \in \mathbb{R}^n & \forall i \in [n] \\ & \|x_i - x_j\|^2 + \|x_j - x_k\|^2 \geqslant \|x_i - x_k\|^2 & \forall i, j, k \in [n] \end{array} \tag{1}$$

We define the value of the above relaxation to be $\mathsf{ARV}(G)$.

Definition 1.4. *Let G be an Abelian group, and let S be a* ***symmetric*** *generating multiset of G of size, i.e. $s \in S \iff s^{-1} \in S$, with the same multiplicity. Further let $d := |S|$. The* ***Abelian Cayley graph*** $\mathrm{Cay}(G, S)$ *to be a graph with vertex set G, and the edges $(g, g+s)$ for all $g \in G, s \in S$. Note that we can treat the edge $(g, g+s)$ as undirected, since $s^{-1} \in S$, and thus $(g+s, g)$ is also an edge. We shall use the letter G to refer to both the group and the Cayley graph (with some implicit generating set S which will be clear from the context).*

We pair up the elements of the group G with their inverses (the elements which are their own inverses are left unpaired). For each such pair, we designate one of the elements in it to be "positive", and the other element in it to be "negative". Elements which are their own inverses are designated positive. S_+ is the multiset of positive elements in S.

[1] for non-regular graphs, $d|S|$ would be replaced by $\mathrm{vol}(S) := \sum_{v \in S} \deg_G(v)$
[2] in a random walk

2 A Combinatorial Characterization of λ_2: Cheeger's Inequality

ϕ can be viewed as a "combinatorial analogue" of λ_2 by restricting the variational characterization to discrete inputs: Indeed, note that for any $x \in \mathbb{R}^n$, $x^\top L x = \frac{1}{d}\sum_{\{i,j\}\in E}(x_i - x_j)^2$. Consequently, for any $S \subseteq [n]$,

$$\mathbf{1}_S^\top L \mathbf{1}_S = \frac{e(S,\overline{S})}{d} \implies \left(\frac{\mathbf{1}_S}{|S|}\right)^\top L \frac{\mathbf{1}_S}{|S|} = \frac{e(S,\overline{S})}{d|S|^2}.$$

Now, note that $L\mathbf{1} = 0$, and $\mathbf{1}^\top L = 0$ too, since L is symmetric. If $x_\perp$ is the component of x orthogonal to $\mathbf{1}$ then $x_\perp^\top L x_\perp = x^\top L x$. Note that for $x = \mathbf{1}_S/|S|, x_\perp =: y_S = \frac{\mathbf{1}_S}{|S|} - \frac{\mathbf{1}}{n}$, and thus

$$\left(\frac{\mathbf{1}_S}{|S|} - \frac{\mathbf{1}}{n}\right)^\top L \left(\frac{\mathbf{1}_S}{|S|} - \frac{\mathbf{1}}{n}\right) = \left(\frac{\mathbf{1}_S}{|S|}\right)^\top L \frac{\mathbf{1}_S}{|S|} = \frac{e(S,\overline{S})}{d|S|^2},$$

and consequently we have

$$R(y_S) = \frac{e(S,\overline{S})/d|S|^2}{(1/|S| - 1/n)^2|S| + (-1/n)^2|\overline{S}|} = \frac{\phi(S)}{|\overline{S}|/n}$$

Thus, if S is such that $\phi(S) = \phi(G)$, then $\lambda_2 \leqslant R(y_S) = \phi(G)/(|\overline{S}|/n) \implies \phi(G) \geqslant \lambda_2|\overline{S}|/n \geqslant \lambda_2/2$.
Consequently, $\phi(G)$ can be viewed as a combinatorial analog of λ_2. In that light, $\phi(G) \geqslant \lambda_2/2$ becomes quite natural, since λ_2 is the optimal value of $R(x)$ over the subspace $\mathbf{1}^\perp$, while $\phi(G)$ is the optimal value only over vectors arising from indicator sets.
Cheeger's inequality (a version of which, for graphs, was proven by Alon and Milman) gives a converse to the above inequality:

Theorem 2.1 (Cheeger's Inequality [AM85]). *For any graph G, $\lambda_2/2 \leqslant \phi(G) \leqslant \sqrt{2\lambda_2}$.*

Proof from [Spi19], Chapter 21. Given the discussion above, it only remains to prove $\phi(G) \leqslant \sqrt{2\lambda_2}$. We present a probabilistic proof in this writeup. The proof inspires a polynomial time algorithm to produce such a cut, which we describe after the proof.
Let x be an unit eigenvector of L corresponding to the eigenvalue of λ_2. Sort the entries of x in increasing order. By relabeling vertices, WLOG assume $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$. Write $z = x - x_{\lceil n/2\rceil} \cdot \mathbf{1}$. Observe that $x^\top L x = z^\top L z$. Furthermore, $z^\top z = x^\top x - 2x_{\lceil n/2\rceil}(\mathbf{1}^\top x) + n x_{\lceil n/2\rceil}^2 = x^\top x + n x_{\lceil n/2\rceil}^2 \geqslant x^\top x$ since x is orthogonal to $\mathbf{1}$. This allows us to conclude that $R(z) \leqslant R(x)$ and in the remaining part of the proof, we shall prove that there exists $\tau \in \mathbb{R}$ such that the threshold cut $S_\tau = \{j \in [n] \mid z_j < \tau\}$ satisfies $\max\{\phi(S_\tau), \phi(\overline{S_\tau})\} \leqslant \sqrt{2R(z)}$. Since either $|S_\tau| \leqslant n/2$ or $|\overline{S_\tau}| \leqslant n/2$, this would allow us to conclude that $\phi(G) \leqslant \sqrt{2R(z)} \leqslant \sqrt{2R(x)} = \sqrt{2\lambda_2}$.
We shall assume without loss of generality that $z_1^2 + z_n^2 = 1$. This can be achieved by multiplying z with a constant since $R(z)$ is invariant to scaling. Given a distribution over τ, if we prove that $\mathbb{E}[e(S_\tau, \overline{S_\tau})] \leqslant \sqrt{2R(z)}\,\mathbb{E}[\min(|S_\tau|, |\overline{S_\tau}|)]$ then this would imply the existence of some $\tau \in \mathbb{R}$ such that $e(S_\tau, \overline{S_\tau}) \leqslant \sqrt{2R(z)}\min(|S_\tau|, |\overline{S_\tau}|) \Leftrightarrow \max\{\phi(S_\tau), \phi(\overline{S_\tau})\} \leqslant \sqrt{2R(z)}$. We shall consider the distribution with density $2|t|$ between z_1 and z_n. The probability that τ lies in the interval $[a, b]$ where $z_1 \leqslant a \leqslant b \leqslant z_n$ is given by $\mathsf{sgn}(b)b^2 - \mathsf{sgn}(a)a^2$. Notice that the probability of τ lying in the interval $[z_1, z_n]$ is 1 since $z_1 \leqslant z_{\lceil n/2\rceil} = 0 \leqslant z_n$.
Due to our centering of z, if $\tau < 0$, then $\min(|S_\tau|, |\overline{S_\tau}|) = S_\tau$ and if $\tau \geqslant 0$, then $\min(|S_\tau|, |\overline{S_\tau}|) = \overline{S_\tau}$. Moreover, under this distribution, $\mathbb{E}[|S_\tau|] = \sum_{i=1}^n \Pr[i \in S_\tau] = \sum_{i=1}^n \Pr[z_i \leqslant \tau]$ and $\mathbb{E}[|\overline{S_\tau}|] = \sum_{i=1}^n \Pr[i \in \overline{S_\tau}] = \sum_{i=1}^n \Pr[z_i > \tau]$. Thus, for

$i \leqslant \lceil n/2 \rceil$, i is in the smaller set if $\tau < 0$ and for $i > \lceil n/2 \rceil$, i is in the smaller set if $\tau \geqslant 0$. Hence,

$$\mathbb{E}[\min(|S_\tau|, |\overline{S_\tau}|)] = \sum_{i=1}^{\lceil n/2 \rceil - 1} \Pr[z_i < \tau < 0] + \sum_{i=\lceil n/2 \rceil}^{n} \Pr[z_i > \tau \geqslant 0] = \sum_{i=1}^{\lceil n/2 \rceil - 1} z_i^2 + \sum_{i=\lceil n/2 \rceil}^{n} z_i^2 = z^\mathsf{T} z \tag{2}$$

And edge $(i, j) \in E$ with $z_i \leqslant z_j$ is also in the set $e(S_\tau, \overline{S_\tau})$ if $z_i \leqslant \tau < z_j$. Hence, the probability can be written as:

$$\Pr[(i,j) \in e(S_\tau, \overline{S_\tau})] = \begin{cases} |z_i^2 - z_j^2| & \text{if } \mathsf{sgn}(z_i) = \mathsf{sgn}(z_j) \\ z_i^2 + z_j^2 & \text{if } \mathsf{sgn}(z_i) \neq \mathsf{sgn}(z_j) \end{cases}$$

We observe that $|z_i^2 - z_j^2| = |z_i - z_j||z_i + z_j| \leqslant |z_i - z_j|(|z_i| + |z_j|)$ and $z_i^2 + z_j^2 \leqslant (z_i - z_j)^2 \leqslant |z_i - z_j|(|z_i| + |z_j|)$. Since both the terms are bounded, we get:

$$\begin{aligned} \mathbb{E}[|e(S_\tau, \overline{S_\tau})|] = \sum_{(i,j) \in E} \Pr[(i,j) \in e(S_\tau, \overline{S_\tau})] &\leqslant \sum_{(i,j) \in E} |z_i - z_j|(|z_i| + |z_j|) \\ &\leqslant \sqrt{\sum_{(i,j) \in E} (z_i - z_j)^2} \sqrt{\sum_{(i,j) \in E} (|z_i| + |z_j|)^2} \qquad [\because \text{Cauchy-Schwarz}] \end{aligned} \tag{3}$$

By the definition of Rayleigh coefficient, $\sum_{(i,j) \in E} (z_i - z_j)^2 \leqslant R(z) z^\mathsf{T} z$. Moreover, we observe that:

$$\sum_{(i,j) \in E} (|z_i| + |z_j|)^2 \leqslant 2 \sum_{(i,j) \in E} z_i^2 + z_j^2 = 2z^\mathsf{T} z$$

Thus, we conclude using equations (3) and (2) that:

$$\mathbb{E}[|e(S_\tau, \overline{S_\tau})|] \leqslant \sqrt{R(z) z^\mathsf{T} z} \sqrt{2 z^\mathsf{T} z} = \sqrt{2R(z)} z \mathsf{T} z = \sqrt{2R(z)}\, \mathbb{E}[\min(|S_\tau|, |\overline{S_\tau}|)]$$

□

The proof motivates the simple algorithm of computing the second eigenvector of the Laplacian, arranging its coordinates in an increasing order, and choosing the cut which minimizes $\max\{\phi(S), \phi(\overline{S})\}$ among all the n possible threshold cuts.

Remark. *Both sides of the inequality are tight: Indeed, take the cycle graph C_n. Its conductance is $2/n$ (which is given by the cut induced by $n/2$ consecutive vertices on the cycle), while $\lambda_2(C_n) \sim \sin^2(\pi/n) = \Theta(1/n^2)$, and thus $\phi(C_n) = \Theta(\sqrt{\lambda_2(C_n)})$. On the other hand, for the hypercube graph H_n (graph on $\mathbb{F}_2^n$ where pairs of vertices with Hamming distance 1 are connected), $\phi(H_n) = 1/n$ (which is given by the cut induced by $S := \{x \in \mathbb{F}_2^n : x_1 = 0\}$), while $\lambda_2(H_n) = 1/n$, i.e. $\phi(H_n) = \Theta(\lambda_2(H_n))$. It is worth noting that both C_n, H_n are Cayley graphs of abelian graphs ($C_n = \mathsf{Cay}(\mathbb{Z}/n\mathbb{Z}, \pm 1)$, $H_n = \mathsf{Cay}(H_n, \{e_1, \ldots, e_n\})$.*

As we can see, the above proof is algorithmic. This is helpful, since calculating $\phi(G)$ exactly is NP-hard, and thus the aforementioned algorithm gives us a $O(1/\sqrt{\phi(G)})$-approximation to calculating $\phi(G)$.
However, note that if $\phi(G) = o(1)$, then the above approximation algorithm doesn't have a nice guarantee (for example, if $\phi(G) \sim 1/n^2$, then we only have a $O(n)$-approximation algorithm). In this context, we now look into the *Arora-Rao-Vazirani* algorithm, which gives a $O(\sqrt{\log n})$-approximation to $\psi(G)$, and thus $\phi(G)$.

3 The Arora-Rao-Vazirani Algorithm

The ARV algorithm [ARV09] is for approximating $\psi(G)$. Note that for $x = \mathbf{1}_S$,

$$\frac{e(S,\overline{S})}{\frac{d}{n}|S|\cdot|\overline{S}|} = \frac{x^\top L x}{x^\top L_n x},$$

where $L_n := I - \mathbf{1}\mathbf{1}^\top/n$ is (roughly) [3] the Laplacian of the complete graph. Finally, if $x_\perp$ is the component of x in $\mathbf{1}^\perp$, then $x^\top L x = x_\perp^\top L x_\perp$ and $x^\top L_n x = x_\perp^\top L_n x_\perp$. Consequently,

$$\min_{x\in\mathbb{R}^n\setminus\{0\},\langle x,\mathbf{1}\rangle=0} \frac{x^\top L x}{x^\top L_n x} = \min_{x\in\mathbb{R}^n\setminus\{0\}} \frac{x^\top L x}{x^\top L_n x}$$

Now, $L_n = I - \frac{\mathbf{1}\mathbf{1}^\top}{n}$. Thus, if $\langle x, \mathbf{1}\rangle = 0$, then $x^\top L_n x = \|x\|^2$, and thus $\min_{x\in\mathbb{R}^n\setminus\{0\},\langle x,\mathbf{1}\rangle=0} \frac{x^\top Lx}{x^\top L_n x} = \min_{x\in\mathbb{R}^n\setminus\{0\},\langle x,\mathbf{1}\rangle=0} R(x) = \lambda_2$. Thus $\lambda_2 \leqslant \psi(G)$, since $\psi(G) = \min_{x=\mathbf{1}_S, 0<|S|\leqslant n/2} \frac{x^\top Lx}{x^\top L_n x}$. Now, note that if x is the indicator vector of a set, then the following "squared-triangle" inequality holds for any i, j, k: $(x_i - x_j)^2 + (x_j - x_k)^2 \geqslant (x_i - x_k)^2$, since $(x_i - x_j)^2 = |x_i - x_j|$ if x is a 0-1 vector. Finally, we can relax the condition of x_i's being real numbers to x_i's being vectors, and reformulate $\frac{x^\top Lx}{x^\top L_n x}$ as $\frac{\sum_{\{i,j\}\in E}\|x_i - x_j\|^2}{\frac{d}{n}\sum_{i,j}\|x_i-x_j\|^2}$. Consequently, we have the relaxation [4] Eq. (1) to calculate $\psi(G)$.

Clearly, $\mathsf{ARV}(G) \leqslant \psi(G)$. Also note that $\lambda_2 = \min_{x\in\mathbb{R}^n, x\neq 0} \frac{x^\top Lx}{x^\top L_n x}$. Now, $x^\top L x = \langle L, xx^\top\rangle$, where the inner product is induced by the Frobenius norm, and thus $\lambda_2 \geqslant \min_{X\succeq 0} \frac{\langle L,X\rangle}{\langle L_n,X\rangle}$, since λ_2 is the optimum of $\frac{\langle L,X\rangle}{\langle L_n,X\rangle}$ when X is restricted to be matrices of the form $xx^\top$, i.e. rank 1 PSD matrices. On the other hand, any $n\times n$ symmetric PSD matrix X can be written as $\sum_{i\in[n]} x_i x_i^\top$, and note that $\frac{\langle L,X\rangle}{\langle L_n,X\rangle} \geqslant \min_{i\in[n]} \frac{\langle L, x_i x_i^\top\rangle}{\langle L_n, x_i x_i^\top\rangle}$ since $\langle L, x_i x_i^\top\rangle, \langle L_n, x_i x_i^\top\rangle \geqslant 0$ for all i. Thus, $\lambda_2 = \min_{X\succeq 0}\frac{\langle L,X\rangle}{\langle L_n,X\rangle}$, and thus $\lambda_2 \leqslant \mathsf{ARV}(G)$ (note that Eq. (1) can be also be recast in terms of PSD matrices, by considering the matrix X given by $X_{ij} := \langle x_i, x_j\rangle$). We thus have $\lambda_2 \leqslant \mathsf{ARV}(G) \leqslant \psi(G)$.

Although we won't need it further, the way the ARV algorithm rounds a solution x of the SDP is as follows: Suppose $x_1, \ldots, x_n$ are the vectors produced by the SDP. Let g be a random Gaussian vector, and write $y_i := \langle x_i, g\rangle$. WLOG suppose $y_1 \leqslant y_2 \leqslant \cdots \leqslant y_n$. Write $L := \{i : i \leqslant c_1 n\}, R := \{i : n - i \leqslant c_1 n\}$, i.e. L is the set of $c_1 n$ indices for which y_i's are the smallest, and R is the set of $c_1 n$ indices for which y_i's are the largest. Let $i \in L, j \in R$ be such that $\|x_i - x_j\|^2$ is the minimum. If $|x_i - x_j\|^2 \leqslant c_2/\sqrt{\log n}$, delete i from L, and j from R, and continue this process until every point in L is distance $\geqslant c_2/\sqrt{\log n}$ away from every point in R. Now, for $i \in [n]$, define $d_i := \min_{j\in L}\|x_i - x_j\|^2$. Use these d_i's to arrange the vertices of $[n]$ on a number line, and then select a t uniformly between the smallest d_i (which will be 0 if L is non-empty) and the largest d_i. The cut induced by this t (i.e. the cut given by $S := \{i \in [n] : d_i > t\}$), is with good probability, [5] a $O(\sqrt{\log n})$-approximation to the sparsest cut.

[3] we say "roughly" because the Laplacian of the n-clique is $\frac{n}{n-1}I - \frac{\mathbf{1}\mathbf{1}^\top}{n-1} = \frac{n}{n-1}L_n$

[4] the following relaxation optimizes rational functions, and in general it is not known how to optimize programs of rational functions. However, we can employ the following trick: We optimize $\sum_{\{i,j\}\in E}\|x_i - x_j\|^2 - \alpha\frac{d}{n}\sum_{i,j}\|x_i - x_j\|^2$ for some parameter α, which is chosen through a binary search (note that $4/n^2 \leqslant \psi(G) \leqslant 2$, and thus we have an interval of polynomial aspect ratio to search in). The interval in our binary search where the minimum of the SDP goes from being positive to being negative is a 2-approximation to the optima of the rational function $\frac{\sum_{\{i,j\}\in E}\|x_i-x_j\|^2}{\frac{d}{n}\sum_{i,j}\|x_i-x_j\|^2}$. Note that optimizing $\sum_{\{i,j\}\in E}\|x_i - x_j\|^2 - \alpha\frac{d}{n}\sum_{i,j}\|x_i - x_j\|^2$ can be achieved through a SDP

[5] of course, c_1, c_2 have to be carefully adjusted in terms of each other

4 Buser's inequality

4.1 Some Motivation

Cheeger's inequality was first proved in the context of Riemannian manifolds by Cheeger. Indeed, if M is a compact Riemannian manifold, then for any measurable subset $S \subset M$, we define $\phi(S) := \text{vol}(\partial S)/\text{vol}(S)$, where ∂S is the boundary of S, and then define $\phi(M)$ to be $\min_{\text{vol}(S) \leqslant \text{vol}(M)/2} \phi(S)$. The Laplacian is defined to be the operator $\text{div}(\nabla f)$, and the Rayleigh quotient of a smooth L^2 map $f : M \to \mathbb{R}$ w.r.t. the Laplacian is defined as:

$$R(f) := \frac{\int_M \|\nabla f\|^2}{\int_M f^2}$$

If M is a compact connected Riemannian manifold, the Laplacian has a discrete [6] spectrum $0 = \lambda_1 < \lambda_2 \leqslant \cdots \to \infty$. Furthermore, the corresponding eigenfunctions $f_1, f_2, \ldots$ form an orthonormal eigenbasis of $L^2(M)$, which is the vector space of all L^2 functions on M. We also have a variational characterization of λ_2, which is

$$\lambda_2 = \min_{f:M \to \mathbb{R} \text{ smooth}, \int_M f = 0} R(f)$$

In this context, Cheeger [Che71] showed that $\phi(M) \leqslant O(\sqrt{\lambda_2})$. However, unlike in the graph case, $\phi(M) \geqslant \lambda_2/2$ doesn't hold: Indeed, the analog of $\mathbf{1}_S$, which is the indicator function of a subset of M, is not smooth (we require smoothness so that ∇f is defined). If we try to "smoothen" it through functions which are 1 on S, and decay rapidly to 0 outside S, we run into the following issue: Around ∂S, $\|\nabla f\|$ is extremely large. Now, suppose the volume of the manifold "explodes" close to ∂S. [7] Then the integral $\int_M \|\nabla f\|^2$ would blow up, and $R(f)$ would deviate significantly from the "intended" value of $\phi(S)$.

This issue of the volume blowing up can arise if the manifold has negative curvature in some places. Thus, one might ask if some analog of $\phi(S) \geqslant \lambda_2/2$ can be established for manifolds with non-negative curvature everywhere.

The answer is yes, and Buser [Bus82] established that $\lambda_2 \leqslant 10\phi(M)^2$ for compact manifolds M with non-negative curvature everywhere. [8] Note that this is much stronger than $\lambda_2 \leqslant 2\phi(G)$; Indeed Buser's inequality along with Cheeger's inequality establishes that for manifolds with non-negative curvature everywhere, $\phi(M) = \Theta(\sqrt{\lambda_2})$.

Knowing this result for manifolds, we can ask the analogous question for graphs. While there is no widely agreed upon definition of curvature in graphs, abelian Cayley graphs have non-negative curvature according to some definitions. Thus, we might ask if we can show that $\phi(G) = \Theta(\sqrt{\lambda_2})$ for abelian Cayley graphs. We can show something even stronger:

Theorem 4.1. *[GT21a, GT21b] Let G be a d-regular abelian Cayley graph. Then*

$$\lambda_2(G) \leqslant O(d) \cdot \mathsf{ARV}(G)^2 \leqslant O(d) \cdot \psi(G)^2 \leqslant O(d) \cdot \phi(G)^2$$

Proof. Let the adjacency matrix of G be A. Fix a parameter t, and let G^t be the graph with adjacency matrix A^t. Note that $\lambda_2(G^t) = 1 - (1 - \lambda_2(G))^t$, and $\mathsf{ARV}(G^t) \geqslant \lambda_2(G^t)$. Thus, if we can show that $\mathsf{ARV}(G^t) \leqslant \sqrt{2dt} \cdot \mathsf{ARV}(G)$ for all t, then by setting $t := \lceil 1/\lambda_2 \rceil$, we obtain:

$$1 - \frac{1}{e} \leqslant 1 - (1 - \lambda_2(G))^t = \lambda_2(G^t) \leqslant \mathsf{ARV}(G^t) \leqslant \sqrt{2dt} \cdot \mathsf{ARV}(G) \leqslant \sqrt{2d\left(\frac{1}{\lambda_2} + 1\right)} \cdot \mathsf{ARV}(G) \implies \lambda_2 \leqslant O(d) \cdot \mathsf{ARV}(G)^2$$

[6] i.e. the multiplicity of all finite eigenvalues is finite

[7] For a concrete example, imagine a dumb-bell shaped manifold, with an extremely short and thin bottleneck. Then the volume of the manifold grows rapidly around the bottleneck

[8] if the Ricci curvature of M is $\leqslant -R$ everywhere, then we have $\lambda_2 \leqslant 2\sqrt{(n-1)|R|} \cdot \phi(M) + 10\phi(M)^2$

Now, when we're looking at $\frac{\sum_{\{i,j\}\in E}\|x_i-x_j\|^2}{\frac{d}{n}\sum_{i,j}\|x_i-x_j\|^2}$ for a Cayley graph G, we can get some further mileage out of the symmetry. Indeed, write $\frac{\sum_{\{i,j\}\in E}\|x_i-x_j\|^2}{\frac{d}{n}\sum_{i,j}\|x_i-x_j\|^2} = \frac{\langle L,X\rangle}{\langle L_n,X\rangle}$ for the appropriate PSD matrix X. Note that if X is a solution, then X_g is also a solution with the same cost, where $(X_g)_{ij} := \langle x_{g+i}, x_{g+j}\rangle$. To see this, note that the denominator is preserved under any permutation of the indices; the numerator is preserved as long as the transformation is a graph automorphism (edges map to edges). Thus WLOG we'll always replace X with its "symmetrized version" $\frac{1}{|G|}\sum_{g\in G} X_g$. Note that if X is symmetrized, then for any $u, v \in G, s \in S$, we have that symmetrization performed

$$\langle x_u, x_{u+s}\rangle \leftarrow \frac{1}{|G|}\sum_{g\in G} x_{u+g}x_{u+s+g} = \frac{1}{|G|}\sum_{g\in G} x_{v+(u-v+g)}x_{v+s+(u-v+g)} = \frac{1}{|G|}\sum_{g'\in G} x_{v+g'}x_{v+s+g'}$$

where the last step follows from the fact that $g = g' - u + v$ is always in the group. The last expression is exactly what $\langle x_v, x_{v+s}\rangle$ became. Thus, $\|x_u - x_{u+s}\| = \|x_v - x_{v+s}\|$ and consequently, for a symmetrized X, we have: $\frac{\langle L,X\rangle}{\langle L_n,X\rangle} = \frac{\frac{n}{2}\sum_{s\in S}\|x_s-x_0\|^2}{\frac{d}{n}\sum_{i,j\in G}\|x_i-x_j\|^2} = \frac{n^2}{2}\cdot\frac{\mathbb{E}_{s\sim S}\|x_s-x_0\|^2}{\sum_{i,j\in G}\|x_i-x_j\|^2}$ where S is the generating set of the Cayley graph. Note that $|S| = d$.
Thus, to show $\mathsf{ARV}(G^t) \leqslant \sqrt{2dt}\cdot\mathsf{ARV}(G)$, it suffices to show that $\sqrt{2dt}\cdot\mathbb{E}_{s\sim S}\|x_s - x_0\|^2 \geqslant \mathbb{E}_{s_1,\ldots,s_t\sim S}\|x_{s_1+\cdots+s_t} - x_0\|^2$.
Now, let $s_1 + \ldots + s_t = \sum_{s\in S_+} c_s \cdot s$ be the "reduced" version of $s_1 + \cdots + s_t$, i.e. pair up all elements of S with their inverses, and retain the "positive" parts in S_+. Then reduce the sum. For example, if the generator s appears 4 times, while $-s$ appears twice, then $c_s = 2$. [9]
Now, by squared triangle inequality and symmetrization, we have

$$\|x_{s_1+\cdots+s_t} - x_0\|^2 = \left\|x_{\sum_{s\in S_+} c_s\cdot s} - x_0\right\|^2 \leqslant \sum_{s\in S_+}|c_s|\cdot\|x_s - x_0\|^2$$

Now, let us estimate $\mathbb{E}_{s_1,\ldots,s_t\in S}|c_s|$ for some $s \in S_+$. Note that c_s is a sum of t independent $\{0,\pm 1\}$-valued random variables (which we call $X_1,\ldots,X_t$) with mean 0 which take the value 0 with probability $1-2/d$. Also, $\mathbb{E}|c_s| \leqslant \sqrt{\mathbb{E}\, c_s^2}$. But $c_s^2 = \left(\sum_{i=1}^t X_i\right)^2 = \sum_i X_i^2 + 2\sum_{1\leqslant i<j\leqslant t} X_iX_j$, and thus $\mathbb{E}\, c_s^2 = t\,\mathbb{E}\, X_1^2 = 2t/d$, i.e. $\mathbb{E}|c_s| \leqslant \sqrt{2t/d}$. [10]
Consequently, $\mathbb{E}\left\|x_{\sum_{s\in S_+} c_s\cdot s} - x_0\right\|^2 \leqslant \sum_{s\in S_+}\mathbb{E}|c_s|\cdot\mathbb{E}\|x_s - x_0\|^2 \leqslant \sqrt{2t/d}\cdot\sum_{s\in S_+}\mathbb{E}\|x_s - x_0\|^2 \leqslant \sqrt{2t/d}\cdot\sum_{s\in S}\mathbb{E}\|x_s - x_0\|^2 = \sqrt{2td}\cdot\mathbb{E}_{s\sim S}\|x_s - x_0\|^2$, as desired. □

Corollary 4.1.1. *Let G be the Cayley graph of an abelian group G with generating set S. Write $|G| = n, |S| = d$. Then $\sqrt{\lambda_2}$ is a $O(\sqrt{d})$-approximation to the conductance of G. In particular, if $d = o(\log n)$, then we have a strictly better guarantee than ARV.*

[9] if s is its own inverse, add ± 1 to c_s with probability $1/2$ each
[10] if s is its own inverse, then c_s takes the value 0 with probability $1 - 1/d$, and thus $\mathbb{E}\, X_1^2 = t/d \leqslant 2t/d$

References

[AM85] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, February 1985.

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2), April 2009.

[Bus82] Peter Buser. A note on the isoperimetric constant. *Ann. Sci. École Norm. Sup. (4)*, 15(2):213–230, 1982.

[Che71] Jeff Cheeger. *A Lower Bound for the Smallest Eigenvalue of the Laplacian*, pages 195–200. Princeton University Press, Princeton, 1971.

[GT21a] Shayan Oveis Gharan and Luca Trevisan. Arv on abelian cayley graphs, 2021.

[GT21b] Shayan Oveis Gharan and Luca Trevisan. Buser inequalities in graphs, 2021.

[Spi19] Daniel A. Spielman. *Spectral and Algebraic Graph Theory*. Self-published, New Haven, CT, 2019. http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf.