| **Advanced Heuristic and Approximation Algorithms** | Grenoble-INP (ORCO) |
|---|---|
| Lecture 8 | |
| *Lecturer: Alantha Newman* | *November 23, 2018* |

# 1 Measures of Graph Connectivity

Given an undirected graph $G = (V, E)$ on $n$ vertices, the *sparsest cut problem* is to find a (proper, nonempty) subset $S \subset V$ of vertices with the following objective function:

$$\textit{Sparsest-Cut}(G) := \min_{S \subset V} \frac{E(S, \bar{S})}{|S| \cdot |\bar{S}|},$$

where $E(S, \bar{S})$ denotes the number of edges with exactly one endpoint in $S$. Without loss of generality, we can assume that $|S| \leq \frac{n}{2}$. This objective function incentivizes the cut to be balanced (i.e., for fixed value of $|E(S, \bar{S})|$, this ratio is smaller when $|S|$ is larger).

A closely related notion is the *conductance* of a graph. This can be stated with the following objective function:

$$\textit{Conductance}(G) := \min_{S \subset V, |S| \leq \frac{n}{2}} \frac{E(S, \bar{S})}{|S|}.$$

The two definitions are related as follows:

$$\frac{\textit{Sparsest-Cut}}{n} \leq \textit{Conductance}(G) \leq \frac{2 \cdot \textit{Sparsest-Cut}(G)}{n}.$$

The problems of computing *Sparsest-Cut*$(G)$ and *Conductance*$(G)$ are NP-hard.

## 1.1 Normalization: Sparsest Cut

We use $uv \in E$ to denote an edge in $E$ and we use $u, v \in V$ to denote unique (unordered) pairs of distinct vertices in $V \times V$ (e.g., there are $\binom{n}{2}$ such pairs). For a set $S \subset V$, we say $\chi_S \in \{0, 1\}^n$ is the *indicator vector* for the set $S$. In other words, $\chi_S(u) = 1$ iff $u \in S$. The following quantity $\sigma(S)$ denotes (almost exactly) the fraction of edges crossing the cut $(S, \bar{S})$ divided by the fraction of pairs of vertices separated by the cut $(S, \bar{S})$. In a $d$-regular graph (i.e., a graph in which each vertex has degree $d$), we can relate $\sigma(G)$ to the cut sparsity.

$$\sigma(S) \quad := \quad \frac{E(S, \bar{S})}{\frac{d}{n} \cdot |S| \cdot |\bar{S}|}. \tag{1}$$

We can derive (1) as follows. The fraction of edges crossing the cut $(S, \bar{S})$ divided by the number of all unordered distinct pairs of vertices is

$$\begin{aligned} \frac{\mathbb{E}_{uv \in E} |\chi_S(u) - \chi_S(v)|}{\mathbb{E}_{u,v \in V} |\chi_S(u) - \chi_S(v)|} \quad &= \quad \frac{\frac{\sum_{uv \in E} |\chi_S(u) - \chi_S(v)|}{(dn/2)}}{\frac{\sum_{u,v \in V} |\chi_S(u) - \chi_S(v)|}{(n(n-1)/2)}} \\ &\approx \quad \frac{\sum_{uv \in E} |\chi_S(u) - \chi_S(v)|}{\frac{d}{n} \sum_{u,v \in V} |\chi_S(u) - \chi_S(v)|} \qquad (2) \\ &= \quad \frac{E(S, \bar{S})}{\frac{d}{n} \cdot |S| \cdot |\bar{S}|}. \end{aligned}$$

Note that $0 \leq \sigma(S) \leq 2$ and this upper bound is tight. For example, let $S$ be one of the two independent sets in the bipartition of a connected bipartite graph) Then all edges belong to $E(S, \bar{S})$ but only half of the vertex pairs are separated by the cut $(S, \bar{S})$.

$$\sigma(G) := \min_{\emptyset \neq S \subset V} \sigma(S).$$

Another useful way to write $\sigma(G)$ is as follows.

$$\sigma(G) \quad = \quad \min_{\mathbf{x} \in \{0,1\}^n - \{\mathbf{0}, \mathbf{1}\}} \frac{\sum_{uv \in E} |x_u - x_v|}{\frac{d}{n} \sum_{u,v \in V} |x_u - x_v|}. \tag{3}$$

### 1.2 Normalization: Expansion

The *expansion* of a set $S$ is the fraction of edges leaving a set divided by the sum of the degrees of all vertices in $S$. In a $d$-regular graph, we have

$$\phi(S) := \min \frac{E(S, \bar{S})}{d \cdot |S|}.$$

Note that $0 \leq \phi(S) \leq 1$ and this upper bound is tight (e.g., let $S$ be one of the two independent sets in the bipartition of a connected bipartite graph).

$$\phi(G) = \min_{S : |S| \leq \frac{n}{2}} \phi(S).$$

We can related $\sigma(G)$ and $\phi(G)$ as follows:

$$\sigma(G) \leq 2 \cdot \phi(G). \tag{4}$$

## 2 Connectivity and Laplacians

**Theorem 1.** *Let $G = (V, E)$ be a $d$-regular graph and let $A$ be its (symmetric) adjacency matrix. Let*

$$L = I - \frac{1}{d} \cdot A$$

*be its* normalized Laplacian *matrix. Let*

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

*be the $n$ eigenvalues of $L$ (possibly with multiplicities). Then*

1. *$\lambda_1 = 0$,*
2. *$\lambda_i \in [0, 2]$,*
3. *$\lambda_k = 0$ iff $G$ has at least $k$ connected components.*

We will need the following important identity, which can be applied to relate eigenvalues of the Laplacian with quadratic forms.

**Fact 2.** *Let $L$ be the normalized Laplacian of a $d$-regular graph $G$ and let $x \in \mathbb{R}^n$ be any vector. Then*

$$\mathbf{x}^\intercal L \mathbf{x} \quad = \quad \frac{1}{d} \sum_{uv \in E} (x_u - x_v)^2. \tag{5}$$

*Proof.*

$$\begin{aligned}
\mathbf{x}^\intercal L\mathbf{x} &= \mathbf{x}^\intercal \left(I - \frac{1}{d}A\right)\mathbf{x} \\
&= \mathbf{x}^\intercal I\mathbf{x} - \frac{1}{d}\cdot(\mathbf{x}^\intercal A\mathbf{x}) \\
&= \sum_{u\in V} x_u^2 - \frac{1}{d}\cdot 2\sum_{uv\in E} x_u x_v.
\end{aligned}$$

We also have

$$\begin{aligned}
\sum_{uv\in E}(x_u - x_v)^2 &= \sum_{uv\in E}(x_u^2 + x_v^2 - 2x_u x_v) \\
&= d\sum_{u\in V} x_u^2 - 2\sum_{uv\in E} x_u x_v.
\end{aligned}$$

◇

The minimum eigenvalue of $L$ is defined as

$$\lambda_1 = \min_{\mathbf{x}\in\mathbb{R}^n-\mathbf{0}} \frac{\mathbf{x}^\intercal L\mathbf{x}}{\mathbf{x}^\intercal\mathbf{x}}.$$

By Fact 2, we see that $\lambda_1 \geq 0$. If $\mathbf{x} = (1,1,\ldots,1)$, then $\mathbf{x}^\intercal L\mathbf{x} = 0$. So $\lambda_1 = 0$.

Theorem 1 says that if $\lambda_2 = 0$, then $G$ has at least two connected components. In other words, it has a set with zero expansion. If $\lambda_2$ is small, we can show that $G$ has small expansion.

**Theorem 3. [Cheeger Inequality]**

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}.$$

Moreover, since $\lambda_2$ is efficiently computable, we can bound the conductance of $G$ in the case when $\lambda_2$ is small. Now we prove the Cheeger Inequality.

## 3 Lower Bound for $\phi(G)$

First, we prove a lower bound on $\phi(G)$.

**Lemma 4.** *$\lambda_2 \leq \sigma(G) \leq 2\phi(G)$.*

This is often referred to as the "easy" direction of the Cheeger Inequality, because it is essentially just showing that $\lambda_2$ is a relaxation of $\sigma(G)$. We write $\sigma(G)$ using (3).

$$\sigma(G) = \min_{\mathbf{x}\in\{0,1\}^n-\{\mathbf{0},\mathbf{1}\}} \frac{\sum_{uv\in E}|x_u - x_v|}{\frac{d}{n}\sum_{u,v\in V}|x_u - x_v|}.$$

Since each term in each summation is 0 or 1, we can also write

$$\sigma(G) = \min_{\mathbf{x}\in\{0,1\}^n-\{\mathbf{0},\mathbf{1}\}} \frac{\sum_{uv\in E}|x_u - x_v|^2}{\frac{d}{n}\sum_{u,v\in V}|x_u - x_v|^2}, \tag{6}$$

Now let us write $\lambda_2$. The second line follows from Fact 2.

$$\begin{aligned}
\lambda_2 &= \min_{\mathbf{x}\in\mathbb{R}^n-\{\mathbf{0}\},\mathbf{x}\perp\mathbf{1}} \frac{\mathbf{x}^\intercal L\mathbf{x}}{\mathbf{x}^\intercal\mathbf{x}} \\
&= \min_{\mathbf{x}\in\mathbb{R}^n-\{\mathbf{0}\},\mathbf{x}\perp\mathbf{1}} \frac{\sum_{uv\in E}|x_u - x_v|^2}{d\cdot\sum_{u\in V} x_u^2}.
\end{aligned} \tag{7}$$

**Claim 5.** *$\lambda_2$ is a relaxation of $\sigma(G)$. Namely,*

$$\lambda_2 \quad = \quad \min_{\mathbf{x}\in\mathbb{R}^n-\{\mathbf{0},\mathbf{1}\}} \frac{\sum_{uv\in E}|x_u-x_v|^2}{\frac{d}{n}\sum_{u,v\in V}|x_u-x_v|^2}. \tag{8}$$

*Proof.* From (7), we have

$$\lambda_2 \quad = \quad \min_{\mathbf{x}\in\mathbb{R}^n-\{\mathbf{0}\},\mathbf{x}\perp\mathbf{1}} \frac{\sum_{uv\in E}|x_u-x_v|^2}{d\cdot\sum_{u\in V}x_u^2}.$$

When $\mathbf{x}\perp\mathbf{1}$, we have

$$\begin{aligned}
\sum_{u,v\in V}|x_u-x_v|^2 &= \sum_{u,v\in V}(x_u^2+x_v^2-2x_ux_v)\\
&= (n-1)\sum_{u\in V}x_u^2 - 2\sum_{u,v\in V}x_ux_v\\
&= n\sum_{u\in V}x_u^2 - 2\sum_{u,v\in V}x_ux_v - \sum_{u\in V}x_u^2\\
&= n\sum_{u\in V}x_u^2 - (\sum_{u\in V}x_u)(\sum_{u\in V}x_u)\\
&= n\sum_{u\in V}x_u^2.
\end{aligned}$$

Thus, when $\mathbf{x}\perp\mathbf{1}$, we have

$$\lambda_2 \quad = \quad \min_{\mathbf{x}\in\mathbb{R}^n-\{\mathbf{0}\},\mathbf{x}\perp\mathbf{1}} \frac{\sum_{uv\in E}|x_u-x_v|^2}{\frac{d}{n}\sum_{u,v\in V}|x_u-x_v|^2}. \tag{9}$$

This is almost what we want, except that we have the extra constraint $\mathbf{x}\perp\mathbf{1}$, which does not appear in the formulation of $\sigma(G)$ (i.e., in Equation (6)) or in the statement of the claim (i.e., in Equation 8). However, we can show that this extra constraint does not increase the value. Consider a vector $\mathbf{x}\in\mathbb{R}^n-\{\mathbf{0},\mathbf{1}\}$ that minimizes the righthandside of (8). Now consider $\mathbf{x}'$ such that $x_u = x_u - \frac{1}{n}\sum_{u\in V}x_u$. Note that (i) $\mathbf{x}\perp\mathbf{1}$, and (ii) the value of the righthandside of (9) on $\mathbf{x}'$ equals the value on $\mathbf{x}$. Thus, we can conclude that Equation (8) of the claim is correct, and therefore $\lambda_2$ is a relaxation of $\sigma(G)$. ◇

# 4 Upper bound for $\phi(G)$

We prove the upper bound by analyzing the Spectral-Partitioning Algorithm. The algorithm "rounds" an input vector $\mathbf{x}\perp\mathbf{1}$ and outputs a subset $S\subset V$ such that $\phi(S)$ is small if the *Rayleigh quotient* of the Laplacian matrix of $G$ and the vector $\mathbf{x}$ is small.

Spectral-Partitioning($G=(V,E),\mathbf{x}\in\mathbb{R}^n$)

1. Sort the vertices $V$ according to values $x_1,x_2,\ldots,x_n$ (i.e., label $V=\{v_1,v_2,\ldots,v_n\}$ where $x_{v_1}\le x_{v_2}\le\cdots\le x_{v_n}$).
2. Let $i\in\{1,2,\ldots,n-1\}$ be such that $\max\{\phi(v_1,v_2,\ldots,v_i),\phi(v_{i+1},v_{i+2},\ldots,v_n)\}$ is minimum.
3. Output: $S=\{v_1,v_2,\ldots,v_i\}$, $\bar{S}=\{v_{i+1},v_{i+2},\ldots,v_n\}$.

**Definition 6.** *Let $G = (V, E)$ be a graph and let $\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}$. The* Rayleigh quotient *of the Laplacian matrix $L$ and vector $\mathbf{x}$ is*

$$R(L, \mathbf{x}) \quad := \quad \frac{\mathbf{x}^\intercal L \mathbf{x}}{\mathbf{x}^\intercal \mathbf{x}}.$$

Applying Fact 2, if $G$ is a $d$-regular graph, then

$$R(L, \mathbf{x}) \quad := \quad \frac{\sum_{uv \in E} |x_u - x_v|^2}{d \cdot \sum_{u \in V} x_u^2}.$$

**Lemma 7.** *Let $G = (V, E)$ be a $d$-regular graph and let $\mathbf{x} \in \mathbb{R}^n$ be a vector such that $\mathbf{x} \perp \mathbf{1}$. Let $S$ be the ouput of* Spectral-Partitioning *on input $G$ and $\mathbf{x}$. Then*

$$\phi(S) \leq \sqrt{2\ R(L, \mathbf{x})}.$$

We fix the graph $G$ and use $R(\mathbf{x})$ to denote $R(L, \mathbf{x})$. To prove Lemma 7, We will show that there exists a distribution over sets $S$ (i.e., over the prefixes considered by the Spectral-Partitioning Algorithm) such that

$$\frac{\mathbb{E}_{S \sim D}\ E(S, \bar{S})}{\mathbb{E}_{S \sim D}\ d \cdot \min\{|S|, |\bar{S}|\}} \quad \leq \quad \sqrt{2\ R(\mathbf{x})}. \tag{10}$$

Since the denominator is always positive (i.e., it is at least $d$), this statement implies

$$\Pr_{S \sim D} \left( \frac{E(S, \bar{S})}{d \cdot \min\{|S|, |\bar{S}|\}} \ \leq \ \sqrt{2\ R(\mathbf{x})} \right) \ > \ 0.$$

In other words, (10) implies that there exists a set $S$ such that

$$\frac{E(S, \bar{S})}{d \cdot \min\{|S|, |\bar{S}|\}} \quad \leq \quad \sqrt{2\ R(\mathbf{x})}.$$

Thus, it implies that some prefix considered by the algorithm has the property stated in the lemma. Moreover, if $\mathbf{x}$ is the second eigenvector, then we have $\phi(S) \leq \sqrt{2\ R(\mathbf{x})} = \sqrt{2\lambda_2}$, which establishes the upper bound.

## 4.1 Proof of Lemma 7

The distribution $D$ (according to which $S$ is chosen) is defined as follows.

**Distribution $D$:**

1. Choose $t \in [x_1, x_n]$ with probability $f(t) = 2|t|$
   (i.e., for $x_1 \leq a \leq b \leq x_n$ : $\Pr[a \leq t \leq b] = \int_a^b 2|t|dt$).
2. $S := \{u \mid x_u \leq t\}$.

Adding a fixed constant to each entry in $\mathbf{x}$ only decreases $R(\mathbf{x})$, and multiplying each entry in $\mathbf{x}$ be a fixed constant does not change the value of $R(\mathbf{x})$. Thus, we can make the following assumptions about the input vector $\mathbf{x}$ (and therefore about the distribution $D$).

1. The median entry of $\mathbf{x}$ is 0 (i.e., $x_{\frac{n}{2}} = 0$).

2. $x_1^2 + x_n^2 = 1$.

We state the following useful facts about the distribution $D$.

**Fact 8.**

$$\Pr[a \le t \le b] = \begin{cases} |a^2 - b^2| & \text{if } a \text{ and } b \text{ have the same signs,} \\ a^2 + b^2 & \text{if } a \text{ and } b \text{ have different signs.} \end{cases}$$

We will also use the following fact.

**Fact 9.** *For two real numbers $a$ and $b$, $(a+b)^2 \le 2a^2 + 2b^2$.*

**Claim 10.** $\mathbb{E}_{S \sim D} \min\{|S|, |\bar{S}|\} = \sum_{u \in V} x_u^2$.

*Proof.* First, consider the case in which $S$ is the set with the smaller cardinality (i.e., $i \le \frac{n}{2}$). Then we have

$$\Pr[x_u \in S \mid u \le \frac{n}{2}] \quad = \quad \Pr[x_u \le t \le 0] \; = \; x_u^2.$$

Now consider the case in which $\bar{S}$ is the smaller set. Then we have

$$\Pr[x_u \in \bar{S} \mid u > \frac{n}{2}] \quad = \quad \Pr[0 \le t \le x_u] \; = \; x_u^2.$$

Thus, for each element $u \in V$, the probability that it belongs to the smaller set is $x_u^2$. We can conclude that

$$\mathbb{E}_{S \sim D}[\min\{|S|, |\bar{S}|\}] \; = \; \sum_{u \in V} x_u^2.$$

◇

Now we need to compute the expected number of edges crossing the cut $(S, \bar{S})$ produced according to the distribution $D$.

**Claim 11.**

$$\mathbb{E}_{S \sim D}[E(S, \bar{S})] \quad \le \quad \sqrt{\sum_{uv \in E} |x_u - x_v|^2} \sqrt{2d \sum_{u \in V} x_u^2}.$$

*Proof.* We want to upper bound the expected number of edges crossing the cut.

$$\mathbb{E}_{S \sim D}[E(S, \bar{S})] \quad = \quad 1 \cdot \sum_{uv \in E} \Pr[uv \text{ is cut by } (S, \bar{S})].$$

An edge $uv \in E$ is cut by $(S, \bar{S})$ if $t$ falls between $x_u$ and $x_v$. Applying Fact 8, we have

$$\Pr[uv \text{ is cut by } (S, \bar{S})] = \begin{cases} |x_u^2 - x_v^2| & \text{if } x_u \text{ and } x_v \text{ have the same signs,} \\ x_u^2 + x_v^2 & \text{if } x_u \text{ and } x_v \text{ have different signs.} \end{cases}$$

So, in general, we can upper bound the probability that an edge is cut as follows.

$$\Pr[uv \text{ is cut by } (S, \bar{S})] \quad \le \quad |x_u - x_v| \cdot (|x_u| + |x_v|).$$

Now we apply Cauchy-Schwartz inequality. Namely, for two vectors $\mathbf{y}, \mathbf{z}$, $\mathbf{y} \cdot \mathbf{z} \le |\mathbf{y}| \cdot |\mathbf{z}|$. In other words, if $\mathbf{y}$ and $\mathbf{z}$ are length $n$ vectors, then

$$\sum_{j=1}^{n} z_j y_j \le \sqrt{\sum_{j=1}^{n} y_j^2} \sqrt{\sum_{j=1}^{n} z_j^2}.$$

So we have

$$\sum_{uv\in E} \Pr[uv \text{ is cut by } (S,\bar{S})] \quad \leq \quad \sqrt{\sum_{uv\in E}(x_u - x_v)^2}\sqrt{\sum_{uv\in E}(|x_u| + |x_v|)^2}.$$

Now we apply inequality from Fact 9 to write

$$\sum_{uv\in E}(|x_u| + |x_v|)^2 \quad \leq \quad \sum_{uv\in E}(2x_u^2 + 2x_v^2) \;=\; 2d\sum_{u\in V} x_u^2.$$

Finally, we have

$$\sum_{uv\in E} \Pr[uv \text{ is cut by } (S,\bar{S})] \quad \leq \quad \sqrt{\sum_{uv\in E}(x_u - x_v)^2}\sqrt{2d\sum_{u\in V} x_u^2}.$$

$\diamond$

Combining Claims 10 and 11, we have

$$\begin{aligned}
\frac{\mathbb{E}_{S\sim D}\ E(S,\bar{S})}{\mathbb{E}_{S\sim D}\ d\cdot\min\{|S|,|\bar{S}|\}} &\leq \frac{\sqrt{\sum_{uv\in E}(x_u - x_v)^2}\sqrt{2d\sum_{u\in V} x_u^2}}{d\cdot\sum_{u\in V} x_u^2} \\
&= \sqrt{\frac{2\sum_{uv\in E}(x_u - x_v)^2}{d\cdot\sum_{u\in V} x_u^2}} \\
&= \sqrt{2\ R(\mathbf{x})}.
\end{aligned}$$

# References

[Tre14] Luca Trevisan. Lecture Notes on Expansion, Sparsest Cut, and Spectral Graph Theory. University of California, Berkeley, 2014.

These notes are taken directly (i.e., copied) from [Tre14].