CSCI-B609: A Theorist's Toolkit, Fall 2016 Sept. 6, 2016

Lecture 03: The Sparsest Cut Problem and Cheeger's Inequality

Lecturer: Yuan Zhou *Scribe: Xuan Dong*

We will continue studying the spectral graph theory in this lecture. We start by introducing the sparsest cut problem and Cheeger's inequalities. The ideas that we develop here will be crucially used later.

1 Sparsest Cut Problem

Firstly let's consider a real world problem: community detection.

1.1 Community detection

In a social network graph, a person is regarded as a vertex, and if person u and person v know each other, then vertex u and vertex v forms an edge (u, v). Intuitively a community is thought as a set S of people which has the following two properties: (1) people in S are likely know each other, and (2) people out of set S (we call this set $\bar{S}$) are less likely to know each other. So communities are groups of vertices which probably share common properties or play similar roles within the graph. In Figure. 1, a schematic example of a graph with communities is shown.The next question would be how to formalize the mathematical definition of a community?

First try: Let's find "cut" across the boundary of S and $\bar{S}$ and make the cut edges$(S, \bar{S})$ be very small. This definition reflects the property (2), but considering the extreme case: what if there is only one vertex in S, like $S = \{v_1\}$? Small S usually has small cuts, but we also need to address large community.

Second try: Let's define $\text{Vol}(S) = \sum_{v \in S} \deg(v)$ which is the sum of the degree of vertices in S. If G is a d-regular graph, then $\text{Vol}(S) = d|S|$. This definition gives us a intuition that $\text{Vol}(S)$ is the total #friends of people in S, and we would like the following formula to be as small as possible,

$$\frac{\text{edges}(S, \bar{S})}{\text{Vol}(S)} = \frac{\#\text{friends out of } S}{\text{total } \#\text{friends of } S}.$$

This definition looks make more sense then the first one, but also considering the extreme case: what if $S = V$ which means all vertices are in the S?

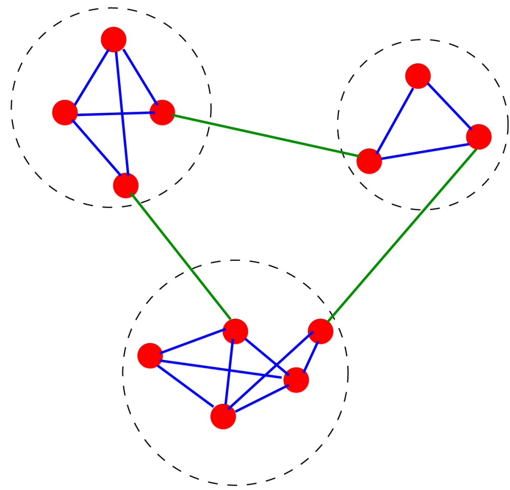

Figure 1: A simple graph with three communities, enclosed by the dashed circles [1].

Definition 1. *Let* $\Phi(S) := \dfrac{\text{edges}(S, \bar{S})}{\min\{\text{Vol}(S), \text{Vol}(\bar{S}\}}$ *be the conductance of* S *for* $S \neq \emptyset$*.*

Third try: The definition of conductance enables us to find good balanced separations, that is, to find a set S with small $\Phi(S)$ while S itself is not too big, and usually we set it $\text{Vol}(S) \leq \frac{1}{2}\text{Vol}(V) = m$. If G is regular, $|S| \leq \frac{1}{2}n$ (n is the number of vertices).

Sparsest cut problem. The goal of the sparsest cut problem is to find a cut with minimum sparsity (sparsest cut). Given a graph G, and find a set $S \in V$, $\text{Vol}(S) \leq m$ which minimizes $\Phi_G(S)$. We can relate the sparsest cut in the graph to the conductance of a graph. Let

$$\Phi_G = \min_{S: 0 < \text{Vol}(S) \leq m} \{\Phi(S)\}$$

be the conductance of the graph G. Intuitively, Φ_G is the smallest conductance among all sets with at most half of the total volume.

Remark 1. *If* G *is a* d*-regular graph, a one-step random walk from* u *ends in* $\bar{S}$ *w.p.*

$$\Pr_{v \sim u}[v \in \bar{S}] = \frac{\text{edges}(u, \bar{S})}{d}.$$

If we pick a uniform random walk $u \in S$ *as start point, a one-step random walk ends in* $\bar{S}$ *w.p.*

$$\Pr_{u \in S, v \sim u}[v \in \bar{S}] = \frac{\sum\limits_{u \in S} \text{edges}(u, \bar{S})}{\sum\limits_{u \in S} \deg(u)} = \frac{\text{edges}(S, \bar{S})}{d|S|} = \Phi(S).$$

In other word, if we randomly pick a vertex of S *and do a one-step random walk, the probability of going out of* S *is the conductance of* S*.*

Now let's see an application of sparsest cut on computer vision area.

1.2 Image segmentation

Image segmentation is the process of partitioning a digital image into multiple segments. Each of the pixels in a region are similar with respect to some characteristic, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic [2]. The goal of segmentation is to simplify or change the representation of an image into regions that is more meaningful and easier to analyze, e.g. sky, people, animals. etc.

Figure 2: A simple sample of graph-based image segmentation [4].

We can treat an image as a pixel graph. Every pixel with $(red, green, blue)$ color channels is a vertex, every pair of neighboring pixels (u, v) is an edge, edge weight is defined as

$$w(u,v) = \exp(-\frac{(r-r')^2 + (g-g')^2 + (b-b')^2}{\theta^2}),$$

where θ is a parameter. Heavy weight when colors are similar. and light weight when colors are not similar. See Figure 2.

Then we could use the idea of sparsest cut as a graph-based image segmentation algorithm. [Shi-Malik'00] proposes a normalized cut approach. It consists of the following main steps [3]:

- Given an image, set up a weighted graph $G = (V, E)$ and set the weight W on the edge connecting two vertices to be a measure of the similarity between the two vertices.
- Solve $(D - W)x = \lambda Dx$ for eigenvectors with the smallest eigenvalues.
- Use the eigenvector with the second smallest eigenvalue to bipartition the graph.
- Decide if the current partition should be subdivided and recursively repartition the segmented parts if necessary.

For further reading, see [3].

2 Cheeger's Inequality

The sparsest cut problem is NP-hard and its approximability is quite open. We will explore a relation between spectral property of a graph and its conductance. This will give a nice approximation to the conductance. The following theorem is called Cheeger's inequality.

Theorem 2 (Cheeger's Inequality)**.** *For regular graph G, let λ_2 be the second smallest eigenvalue of L_G, then we have*

$$\frac{1}{2}\lambda_2 \leq \Phi_G \leq \sqrt{2\lambda_2}.$$

Cheeger's inequality is perhaps one of the most fundamental inequalities in discrete optimization and spectral graph theory [7]. It relates the eigenvalue of the normalized Laplacian matrix L_G to the conductance Φ_G. As noted above, the second smallest non-zero eignenvalue of the Laplacian of a graph tells us how connected the graph is. For exapmle, Φ_G is close to zero and G has a natural 2-clustering if and only if λ_2 is close to zero. Thus, it is important to find bounds for this value. Before proving this theorem, we need some definitions and facts.

Definition 3. *For a given symmetric matrix $M \in \mathbb{R}^{n\times n}$ and non-zero vector $\vec{x} \in \mathbb{R}^n$, the Rayleigh quotient $R(M, \vec{x})$ is defined as*

$$R(M, \vec{x}) := \frac{\vec{x}^T M \vec{x}}{\vec{x}^T \vec{x}}.$$

Because any $\vec{x} \in \mathbb{R}^n$ could be represented as $\vec{x} = \sum_{i=1}^{n} y_i \vec{\psi}_i$ via an orthogonal transformation, so let's write $M = \sum_{i=1}^{n} \lambda_i \vec{\psi}_i \vec{\psi}_i^T$ with $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Now Raileigh quotient could be written as

$$R(M, \vec{x}) = \frac{\sum_{i=1}^{n} \lambda_i y_i^2}{\sum_{i=1}^{n} y_i^2} \in [\lambda_1, \lambda_n],$$

with $R(M, \vec{\psi}_1) = \lambda_1$ and $R(M, \vec{\psi}_n) = \lambda_n$.

Theorem 4 (Courant-Fischer)**.** *Let M be an $n \times n$ symmetric matrix and let $\lambda_1 \leq \lambda_2 \leq$*

$\cdots \leq \lambda_n$ *be the eigenvalues of* M *and* $\psi_1, \psi_2, \cdots, \psi_n$ *be the corresponding eigenvectors. Then*

$$\lambda_1 = \min_{\vec{x}\neq\vec{0}} \{\frac{\vec{x}^T M \vec{x}}{\vec{x}^T \vec{x}}\},$$

$$\lambda_n = \max_{\vec{x}\neq\vec{0}} \{\frac{\vec{x}^T M \vec{x}}{\vec{x}^T \vec{x}}\},$$

$$\lambda_2 = \min_{\substack{\vec{x}\neq\vec{0} \\ \vec{x}\perp\vec{\psi}_1}} \{\frac{\vec{x}^T M \vec{x}}{\vec{x}^T \vec{x}}\}.$$

Observe that $R(M, \vec{\psi_2}) = \lambda_2$, and also for all $\vec{x} \perp \vec{\psi_1}$, we can write $\vec{x} = \sum_{i=2}^{n} y_i \vec{\psi_i^2}$, therefore

$$R(M, \vec{x}) = \frac{\sum_{i=2}^{n} \lambda_i y_i^2}{\sum_{i=2}^{n} y_i^2} \geq \lambda_2.$$

The Courant Fischer characterization of the eigenvalues of a symmetric matrix M in terms of the maximizers and minimizers of the Rayleigh quotient plays a fundamental role in spectral graph theory. For further reading, see [5] and [6].

The left side of the Cheeger's inequality is usually referred to as the easy direction and the right side is known as the hard direction. Now let us prove the LHS of Cheeger's Inequlity ($\frac{1}{2}\lambda_2 \leq \Phi_G$).

Proof. Let $S : |S| \leq \frac{1}{2}|V|$ be the set such that $\Phi_G = \Phi(S) = \frac{\text{edges}(S, \bar{S})}{d|S|}$. Consider $\vec{x} = \vec{1_\delta} - \frac{|S|}{n} \cdot \vec{1}$, and

$$\vec{x}(u) = \begin{cases} 1 - \frac{|S|}{n} & u \in S, \\ -\frac{|S|}{n} & u \notin S. \end{cases}$$

Then we have

$$\vec{x}^T L_G \vec{x} = \frac{1}{d} \sum_{(u,v)\in E} (\vec{x}(u) - \vec{x}(v))^2 = \frac{1}{d}\text{edges}(S, \bar{S}),$$

and

$$\begin{aligned}
\vec{x}^T\vec{x} &= |S|\cdot(1-\frac{|S|}{n})^2 + (n-|S|)\cdot(-\frac{|S|}{n})^2 \\
&= (n-|S|)[\frac{|S|}{n}(1-\frac{|S|}{n}) + (-\frac{|S|}{n})^2] \\
&= (n-|S|)\cdot\frac{|S|}{n} \\
&\geq \frac{1}{2}\cdot|S|.
\end{aligned}$$

Recall Courant Fischer theorem

$$\begin{aligned}
\lambda_2 &\leq R(M,\vec{x}) \\
&= \frac{\vec{x}^T L_G \vec{x}}{\vec{x}^T\vec{x}} \\
&\leq 2\cdot\frac{\text{edges}(S,\bar{S})}{d|S|} \\
&= 2\Phi_G.
\end{aligned}$$

□

References

[1] Fortunato, Santo. Community detection in graphs. *Physics reports* 486.3 (2010): 75-174.

[2] Shapiro, Linda, and George C. Stockman. Computer vision. 2001. ed: *Prentice Hall* (2001).

[3] Shi, Jianbo, and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence* 22.8 (2000): 888-905.

[4] `http://cs.brown.edu/~pff/segment/`

[5] Chung, Fan RK. Spectral graph theory. Vol. 92. *American Mathematical Soc.*, 1997.

[6] Spielman, Daniel. Spectral graph theory. *Lecture Notes, Yale University* (2009): 740-0776.

[7] Shayan, Oveis Gharan. Cheeger's Inequality and the Sparsest Cut Problem. *Lecture Notes, Washington University* (2015).