Approximation Guarantees for Spectral Clustering

Tony Jebara
Columbia University
New York, NY 10027
jebara@cs.columbia.edu

Abstract

We show a classic result that spectral clustering on a b-regular graph can only solve sparse cut to $\phi_{sc} \leq \sqrt{8b\phi}$.

We are given a weighted graph G with n nodes with weights W which can equivalently be represented as an adjacency matrix $A \in \mathbb{R}^{n \times n}$. Graph partitioning recovers a cut or subset of vertices S on G such that $|S| \leq |V|$. Good choices of S are obtained by minimizing the SPARSE CUT criterion defined as

$$\mathcal{SC}_G(S) \quad = \quad W(S, \bar{S})/(|S||\bar{S}|/n).$$

Here we have defined $W(S_1, S_2)$ as the total weight of edges between the set of nodes S_1 and S_2.

Consider trying to minimize $\mathcal{SC}_G(S)$ which is NP-hard. Instead, we attempt to run the following polynomial-time algorithm.

SPECTRALCUT:
Input: a graph $G = (V, E)$ with adjacency matrix A.
1. Compute the second leading eigenvector $\mathbf{v} \in \Re^n$ of A.
2. For each $i = 1, \ldots, n$ create a candidate partition $S_i = \{j : j \in V, \mathbf{v}(j) \leq \mathbf{v}(i)\}$.
3. Output the partition with lowest SPARSE CUT value $S = \arg\min_{i \in \{1,\ldots,n\}} \mathcal{SC}_G(S_i)$.

The following theorem says how well this algorithm performs.

Theorem 1 *Given a b-regular graph G with the optimal* SPARSEST CUT $\phi = \min_S \mathcal{SC}_G(S)$ *then algorithm* SPECTRALCUT *provides a cut for G that achieves a sparse cut value ϕ_{sc} satisfying $\phi_{sc} \leq \sqrt{8b(b - \lambda_2)}$ and therefore satisfying $\phi_{sc} \leq \sqrt{8b\phi}$.*

Proof 1 *Consider the incidence matrix A for G for a b-regular graph which satisfies $A\mathbf{1} = b\mathbf{1}$. The leading eigenvalue of the graph is $\lambda_1 = b$. Given a vector $\mathbf{x} \in \Re^n$, the second eigenvalue is:*

$$\lambda_2 \quad = \quad \max_{\mathbf{x}:\mathbf{x} \in \Re^n, \mathbf{x} \perp \mathbf{1}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}.$$

Similarly, it is straightforward to show for b-regular graphs that

$$b - \lambda_2 \quad = \quad \min_{\mathbf{x}:\mathbf{x} \in \Re^n, \mathbf{x} \perp \mathbf{1}} \frac{\sum_{ij} A(i,j)(\mathbf{x}(i) - \mathbf{x}(j))^2}{\frac{1}{n}\sum_{ij}(\mathbf{x}(i) - \mathbf{x}(j))^2} \; = \; \min_{\mathbf{x}:\mathbf{x} \in \Re^n} \frac{\sum_{ij} A(i,j)(\mathbf{x}(i) - \mathbf{x}(j))^2}{\frac{1}{n}\sum_{ij}(\mathbf{x}(i) - \mathbf{x}(j))^2}$$

where we have dropped the perpendicularity constraint which is redundant. It is also straightforward to see that

$$b - \lambda_2 = \min_{\mathbf{x}:\mathbf{x} \in \Re^n} \frac{\sum_{ij} A(i,j)(\mathbf{x}(i) - \mathbf{x}(j))^2}{\frac{1}{n}\sum_{ij}(\mathbf{x}(i) - \mathbf{x}(j))^2} \leq \min_{\mathbf{x}:\mathbf{x} \in \{-1,1\}^n} \frac{\sum_{ij} A(i,j)(\mathbf{x}(i) - \mathbf{x}(j))^2}{\frac{1}{n}\sum_{ij}(\mathbf{x}(i) - \mathbf{x}(j))^2} = \phi \quad (1)$$

since minization over real values is a strict relaxation over the discrete minimization producing ϕ. Define $\mathbf{v} \in \Re^n$ as the second leading eigenvector which minimizes the continuous optimization *above:*

$$b - \lambda_2 \;=\; \frac{\sum_{ij} A(i,j)(\mathbf{v}(i) - \mathbf{v}(j))^2}{\frac{1}{n}\sum_{ij}(\mathbf{v}(i) - \mathbf{v}(j))^2}. \tag{2}$$

Normalize $\mathbf{v}$ *to obtain* $\hat{\mathbf{v}} \propto \mathbf{v}$ *such that* $\max_i \hat{\mathbf{v}}(i) - \min_i \hat{\mathbf{v}}(i) = 1$. *Consider selecting a cut* S *by picking a threshold* t *as in the algorithm* SPECTRALCUT *where* t *is distributed uniformly in the interval* $t \in [\min_i \hat{\mathbf{v}}(i), \max_i \hat{\mathbf{v}}(i)]$. *The cut we produce is then* $S = \{j : j \in V, \hat{\mathbf{v}}(j) \geq t\}$. *The probability that edge* (i,j) *is in the cut* $E(S,\bar{S})$ *is proportional to* $|\hat{\mathbf{v}}(i) - \hat{\mathbf{v}}(j)|$. *It is easy to see that the following expectations over* t *are satisfied* $\mathrm{E}[W(S,\bar{S}] = \frac{1}{2}\sum_{ij} A(i,j)|\hat{\mathbf{v}}(i) - \hat{\mathbf{v}}(j)|$ *and* $\mathrm{E}[|S||\bar{S}|] = \frac{1}{2}\sum_{ij}|\hat{\mathbf{v}}(i) - \hat{\mathbf{v}}(j)|$. *Thus, as we sample* t *we must find a threshold that satisfies:*

$$\frac{W(S,\bar{S})}{\frac{1}{n}|S||\bar{S}|} \quad\leq\quad \frac{\sum_{ij} A(i,j)|\hat{\mathbf{v}}(i) - \hat{\mathbf{v}}(j)|}{\sum_{ij}|\hat{\mathbf{v}}(i) - \hat{\mathbf{v}}(j)|}.$$

Minimizing over $\mathbf{v}$ *then yields*

$$\phi_{sc} \quad=\quad \min_{\mathbf{v}\in\Re^n} \frac{\sum_{ij} A(i,j)|\hat{\mathbf{v}}(i) - \hat{\mathbf{v}}(j)|}{\frac{1}{n}\sum_{ij}|\hat{\mathbf{v}}(i) - \hat{\mathbf{v}}(j)|} \;=\; \min_{\mathbf{v}\in\Re^n} \frac{\sum_{ij} A(i,j)|\mathbf{v}(i) - \mathbf{v}(j)|}{\frac{1}{n}\sum_{ij}|\mathbf{v}(i) - \mathbf{v}(j)|}. \tag{3}$$

Assume without loss of generality that the median of $\mathbf{v} = 0$. *Define the vector* $\mathbf{y} \in \Re^n$ *such that* $\mathbf{y}(i) = \mathbf{v}(i)|\mathbf{v}(i)|$. *It is immediate to see that*

$$\frac{1}{n}\sum_{i,j}|\mathbf{v}(i) - \mathbf{v}(j)|^2 \;=\; 2\sum_i \mathbf{v}(i)^2 - 2\left(\sum_i \mathbf{v}(i)\right)^2 \;\leq\; 2\sum_i |\mathbf{y}(i)| \tag{4}$$

and that

$$|\mathbf{y}(i) - \mathbf{y}(j)| \quad=\quad |\mathbf{v}(i) - \mathbf{v}(j)|(\mathbf{v}(i) + \mathbf{v}(j)).$$

Multiply both sides by $A(i,j)$ *and sum over* i,j *to get:*

$$\sum_{ij} A(i,j)|\mathbf{y}(i) - \mathbf{y}(j)| \quad=\quad \sum_{ij} A(i,j)|\mathbf{v}(i) - \mathbf{v}(j)|(|\mathbf{v}(i)| + |\mathbf{v}(j)|).$$

Apply Cauchy-Schwartz to the above expression:

$$\begin{aligned}\sum_{ij} A(i,j)|\mathbf{y}(i) - \mathbf{y}(j)| \quad&\leq\quad \sqrt{\sum_{ij} A(i,j)|\mathbf{v}(i) - \mathbf{v}(j)|^2}\sqrt{\sum_{ij} A(i,j)(|\mathbf{v}(i)| + |\mathbf{v}(j)|)^2} \\ &=\quad \sqrt{\frac{b-\lambda_2}{n}\sum_{ij}(\mathbf{v}(i) - \mathbf{v}(j))^2}\sqrt{\sum_{ij} A(i,j)(|\mathbf{v}(i)| + |\mathbf{v}(j)|)^2}\end{aligned}$$

where we plugged in Equation 2 inside the left root. Next, apply Equation 4 in the left root:

$$\sum_{ij} A(i,j)|\mathbf{y}(i) - \mathbf{y}(j)| \quad\leq\quad \sqrt{2(b-\lambda_2)\sum_i |\mathbf{y}(i)|}\sqrt{\sum_{ij} A(i,j)(|\mathbf{v}(i)| + |\mathbf{v}(j)|)^2}.$$

Apply Jensen's inequality $(\mathrm{E}[x])^2 \leq \mathrm{E}[x^2]$ *inside the right root:*

$$\begin{aligned}\sum_{ij} A(i,j)|\mathbf{y}(i) - \mathbf{y}(j)| \quad&\leq\quad \sqrt{2(b-\lambda_2)\sum_i |\mathbf{y}(i)|}\sqrt{\sum_{ij} A(i,j)(2|\mathbf{v}(i)|^2 + 2|\mathbf{v}(j)|^2)} \\ &=\quad \sqrt{2(b-\lambda_2)\sum_i |\mathbf{y}(i)|}\sqrt{4b\sum_i |\mathbf{v}(i)|^2}\end{aligned}$$

where the second line holds since A *comes from a* b*-regular graph. Next, since* $|\mathbf{v}(i)|^2 = |\mathbf{y}(i)|$

$$\begin{aligned}\sum_{ij} A(i,j)|\mathbf{y}(i) - \mathbf{y}(j)| \quad&\leq\quad \sqrt{8b(b-\lambda_2)}\sum_i |\mathbf{y}(i)| \\ &\leq\quad \sqrt{8b(b-\lambda_2)}\frac{1}{n}\sum_{ij}|\mathbf{y}(i) - \mathbf{y}(j)|\end{aligned}$$

where the last step holds since the median is zero. Dividing both sides by $\frac{1}{n}\sum_{ij}|\mathbf{y}(i)-\mathbf{y}(j)|$ *gives*

$$\frac{\sum_{ij} A(i,j)|\mathbf{y}(i)-\mathbf{y}(j)|}{\frac{1}{n}\sum_{ij}|\mathbf{y}(i)-\mathbf{y}(j)|} \quad\leq\quad \sqrt{8b(b-\lambda_2)}.$$

Since Equation 3 guarantees that ϕ_{sc} *is lower than the left hand side of the above equation for* any *choice of* $\mathbf{v}\in\Re^n$ *or* $\mathbf{y}\in\Re^n$*, we have* $\phi_{sc}\leq\sqrt{8b(b-\lambda_2)}$ *as desired for the first part of the theorem. Applying Equation 1 to* $b-\lambda_2$ *gives the second part of the theorem.*