CS369M: Algorithms for Modern Massive Data Set Analysis **Lecture 13 - 11/06/2009**

Global and Local Spectral Methods for Clustering and Partitioning Graphs and Data

Lecturer: Michael Mahoney *Scribes: David Fong and Rajendra Shinde*

**Unedited Notes*

1 Graph partitioning using spectral methods

Recall Cheeger's inequality

$$\frac{d-\lambda_2}{2} \leq h_G \leq \sqrt{2d(d-\lambda_2)} \tag{1}$$

Here G is a d-regular Graph and $h_G = \min_{S,|S|\leq\frac{|V|}{2}} \frac{E(S,\bar{S})}{|S|}$ denotes the edge-expansion of the graph. Also $\lambda_2 = \max_{x\perp\bar{1}} \frac{x^TAx}{x^Tx}$ denotes the Fiedler value of the graph (least non-zero eigenvalue of the adjacency matrix of the graph) and $\bar{1}$ denotes the all ones vector.

Alternate form of Cheeger's inequality which holds for all connected graphs G[1]

$$2h_G \geq \lambda_G \geq \alpha_G \geq h_G^2/2 \tag{2}$$

where G is any graph, h_G is the Cheeger constant for the graph defined as $h_G = \min_{S,\mathrm{Vol}(S)\leq \mathrm{Vol}(V)/2} \frac{\partial S}{\mathrm{Vol}(S)}$ [1], λ_G denotes Fiedler value of the graph Laplacian (i.e. second largest eigenvalue of the graph laplacian) and α_G represents minimum value among all Cheeger ratios of initial segments of vertices when all vertices are arranged in a line using the eigen-vector associated with λ_2.

Proof of: $\frac{d-\lambda_2}{2} \leq h_G$ in (1)

Consider the quadratic form

$$\begin{aligned}\sum_{ij} A_{ij}(x_i-x_j)^2 &= 2d(x^Tx) - 2\sum_{ij} x_iA_{ij}x_j. \\ &= 2d(x^Tx) - 2x^TAx.\end{aligned} \tag{3}$$

Recall

$$\begin{aligned}\lambda_2 &= \max_{x\perp\bar{1}} \frac{x^TAx}{x^Tx} \\ &= \max_{x\perp\bar{1}} \frac{dx^Tx-(1/2)\sum_{ij}A_{ij}(x_i-x_j)^2}{2x^Tx} \\ &= d - \min_{x\perp\bar{1}} \frac{\sum_{ij}A_{ij}(x_i-x_j)^2}{2x^Tx}\end{aligned} \tag{4}$$

Let S denote the set which achieves the minimum Cheeger ratio ie. $\frac{E(S,\bar{S})}{\min(|S|,|\bar{S}|)} = h_G$. Let $p = |S|/n$ and $q = |\bar{S}|/n = 1-p$. Let

$$\begin{aligned}X_i &= q \text{ if } i\in S \\ X_i &= -p \text{ if } i\in \bar{S}\end{aligned} \tag{5}$$

Then $x\cdot\bar{1} = |S|q + |\bar{S}|p = 0 \implies x\perp\bar{1}$. Also, $x^Tx = |S|q^2 + |\bar{S}|p^2 = npq^2 + nqp^2 = nqp(p+q) = npq$. Then,

$$\begin{aligned}d-\lambda_2 &= \min_{x\perp\bar{1}} \frac{\sum_{ij}A_{ij}(x_i-x_j)^2}{2x^Tx} \\ &= E(S,\bar{S})/npq \\ &= nE(S,\bar{S})/(|S||\bar{S}|)\end{aligned} \tag{6}$$

[1] $\mathrm{Vol}(S) = \sum_{i\in S} d_i$

Now we know that the sparsity(sp) of the cut $(S, \bar{S})$ is defined as $sp(S) = \frac{nE(S,\bar{S})}{|S||\bar{S}|}$. Further, $sp(S) = \frac{nE(S,\bar{S}}{\min(|S|,|\bar{S}|)\max(|S|,|\bar{S}|)}$ and using the fact that $\max(|S|, |\bar{S}|) \geq n/2$, we get $sp(S) \leq 2\frac{E(S,\bar{S})}{\min(|S|,|\bar{S}|)} = 2h_G$. Hence we get $d - \lambda_2 \leq 2h_G$.

In fact if ϕ denotes the *sparsest cut* (i.e. the cut of minimum sparsity) then it can be shown that

$$\phi \geq h_G \geq (1/2)\phi \tag{7}$$

Hence, an approximation to the sparsest cut is a 2-approximation to h_G.

Another explanation: We have already seen that

$$\begin{array}{l} \sum_{i,j} A_{ij}(x_i - x_j)^2 = 2dx^T x - 2x^T Ax \\ \implies \sum_{i,j}(x_i - x_j)^2 = 2nx^T x - 2x^T \mathbf{1}x \\ \text{where } \mathbf{1} \text{ denotes the matrix of all ones, i.e. } \mathbf{1} \cdot x = 0,\ \forall x \perp \bar{1} \end{array} \tag{8}$$

Plugging this back in (9), we get,

$$\begin{aligned} d - \lambda_2 \quad &= \min_{x \perp \bar{1}} \frac{\sum_{ij} A_{ij}(x_i - x_j)^2}{2x^T x} \\ &= \min_{x \perp \bar{1}} \frac{\sum_{ij} A_{ij}(x_i - x_j)^2}{(1/n)\sum_{ij}(x_i - x_j)^2} \end{aligned} \tag{9}$$

However observe that the right side is invariant under the transformation $x \to x + c\bar{1}$. Hence we can choose c in order to eliminate the constraint $x \perp \bar{1}$.

Hence

$$d - \lambda_2 \quad = \min_{x \in \mathbb{R}^m} \frac{\sum_{ij} A_{ij}(x_i - x_j)^2}{(1/n)\sum_{ij}(x_i - x_j)^2} \tag{10}$$

Now again, recall the sparsest cut

$$\begin{aligned} \phi \quad &= \min_S \frac{nE(S,\bar{S})}{|S||\bar{S}|} \\ &= \min_{x \in \{0,1\}^n} \frac{\sum_{ij} A_{ij}|x_i - x_j|}{(1/n)\sum_{ij}|x_i - x_j|} \\ &= \min_{x \in \{0,1\}^n} \frac{\sum_{ij} A_{ij}(x_i - x_j)^2}{(1/n)\sum_{ij}(x_i - x_j)^2} \end{aligned} \tag{11}$$

Hence comparing (10) and (11), we get $d - \lambda_2 \leq \phi$. Also, using (7), we conclude that $d - \lambda_2 \leq 2h_G$

2 Claim: $h \leq \sqrt{2d(d - \lambda_2)}$ for d-regular graphs

Proof:

$$\vdash_1: (\forall y \in \mathbb{R}^n) \quad \sum_{i,j} A_{i,j}|y_i^2 - y_j^2| \leq \sqrt{2dy^T y - 2dy^T Ay}\sqrt{4dy^T y}$$

$$
\begin{aligned}
\text{LHS} &= \sum_{i,j} A_{i,j}^{\frac{1}{2}} |y_i - y_j||y_i + y_j| A_{i,j}^{\frac{1}{2}} \\
&\le \left(\sum_{i,j} A_{i,j} |y_i - y_j|^2 \right)^{\frac{1}{2}} \left(\sum_{i,j} A_{i,j} |y_i + y_j|^2 \right)^{\frac{1}{2}} \\
&\le \left(\sum_{i,j} A_{i,j} |y_i - y_j|^2 \right)^{\frac{1}{2}} \left(\sum_{i,j} 2A_{i,j} (y_i^2 + y_j^2) \right)^{\frac{1}{2}} \\
&\le \left(\sum_{i,j} A_{i,j} |y_i - y_j|^2 \right)^{\frac{1}{2}} \left(2d \sum_{i,j} \frac{y_i^2 + y_j^2}{y^T y} \right)^{\frac{1}{2}} \\
&\le \left(2dy^T y - 2y^T A y \right)^{\frac{1}{2}} \left(4dy^T y \right)^{\frac{1}{2}}
\end{aligned}
$$

$$
\vdash_2 : \begin{cases} \text{Let } x \text{ be an eigenvector with eigenvalue } \lambda \text{i. i.e. } xA = \lambda_2 x \text{ s.t. } |\{i : x_i > 0\}| \le \frac{n}{2} \\ \text{Define } y : y_i = \max\{x_i, 0\}, \text{ then } yA \ge \lambda_2 y \text{ componentwise.} \end{cases}
$$

Since A is positive, we have

$$
\begin{cases} x_i > 0 & (yA)_i \ge (xA)_i = (\lambda_2 x)_i = (\lambda_2 y)_i \\ x_i < 0 & (yA)_i \ge (xA)_i \end{cases}
$$

$$
\text{For } y \text{ defined before, } \sum_{i,j} A_{i,j} |y_i^2 - y_j^2| \ge 2hy^T y
$$

Arrange the components of y in non-increasing order

$$
y(i_1) \ge y(i_2) \ge \cdots y(i_n)
$$

With t of them strictly greater than 0, i.e.

$$
y(i_t) > y(i_{t+1}) = \cdots = y(i_n) = 0
$$

Let K be the set such that jump occurs:

$$
K = \{i : y(i_k) > y(i_{k+1})\}
$$

$$
\begin{aligned}
\sum_{u,v} A_{u,v} |y(u)^2 - y(v)^2| &= 2 \sum_{i=1}^{t} \sum_{j=i+1}^{n} A_{v_i, v_j} (y(v_i)^2 - y(v_j)^2) \\
&= 2 \sum_{k \in K} \sum_{ii \le k} \sum_{j > k} A_{v_i, v_j} (y(v_i)^2 - y(v_j)^2)
\end{aligned}
$$

For each $k = 1, \cdots, n$, let

$$
\begin{aligned}
&L_k = \{v_i : i \le k\} \\
&L_0 = \phi
\end{aligned}
$$

Note $\sum_{i\leq k}\sum j > kA_{v_i,v_j} \geq h|L_k| \quad (*)$

$$\begin{aligned}
\text{*RHS} &\geq 2\sum_k h|L_k|(y(v_k)^2 - y(v_{k+1})^2) \\
&= 2h\sum_k(|L_k| - |L_{k'}|)y(v_k)^2 \\
&= 2h\sum_k |\{v : y(v) = y(v_k)\}|y(v_k)^2 \\
&= 2d\sum_v y(v)^2 \\
&= 2hy^Ty
\end{aligned}$$

$$\begin{aligned}
h &\leq \frac{\sum_{i,j} A_{i,j}(y_i - y_j)^2}{2y^Ty} \\
&\leq \frac{2dy^Ty - 2y^TAy)^{\frac{1}{2}}(2dy^Ty)^{\frac{1}{2}}}{2y^Ty} \\
&\leq (2d(d-\lambda_2))^{\frac{1}{2}}
\end{aligned}$$

Algorithm (Spectral Graph Partitioning)

1. Compute 2nd eigenvector of A/L
2. Perform a sweep cut in some way (i.e. Check the set of best notes derived from the eigenvector)
3. Keep the best cut

Potential Issues

- The actual set returned might not be 'good', e.g. clse to optimal
- The eigenvector computation may be too expensive.
- Local information may be reliable, but global not. For extremely large graphs, or graphs with local information nice, global properties bad. We may want to cluster locally, pull out a set of nodes that are good near you.

Lots of heuristics are motivated by this: Cut out nearest neighbors, 2nd nearest neighbors, ...

Can we inherit some of the nice properies of the global spectral? For example the Cheeger's inequality, sweep cut, ...

2 ways to be local:

- Bias yourself locally, but still do computation that depend on the size of the graph.
- Have computation that depend on the size of the output, not the size of the graph.

References

[1] F. Chung. Four proofs of the Cheeger inequality and graph partition algorithms.