CS 596: Theoretical Machine Learning

Lecturer: Pranjal Awasthi | Lecture 14
Scribe: Alireza Naghizadeh | November 03, 2016

1 Recap

In the previous lecture we studies spectral clustering in the context of stochastic block models. Specifically, let

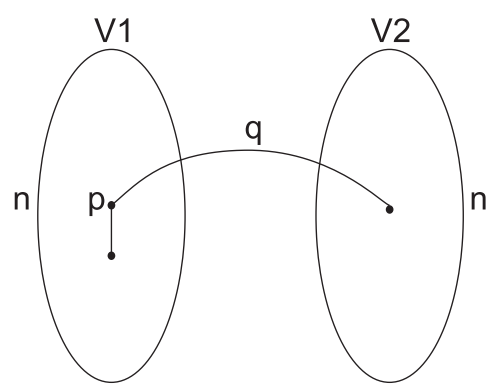
- A= adjacency matrix of G.
- v_2 = second singular vector
- $S_1 = \{i : v_2(i) > 0\}, S_2 = \{i : v_2(i) < 0\}$

Theorem 1. *If $p - q > c\sqrt{\frac{p \log n}{n}}$ then w.p[1] $\geq 1 - \frac{1}{n^3}$ we have,*

$$(S_1 \Delta V_1) + (S_2 \Delta V_2) \leq \frac{n}{\log n}$$

2 Sparse Setting

In this lecture, first we will study the sparse setting,



which $p = \frac{a}{n}, q = \frac{b}{n}$ where a,b are constants.

Q: Does spectral clustering work in the above setting?

A: Yes, it works with a modification

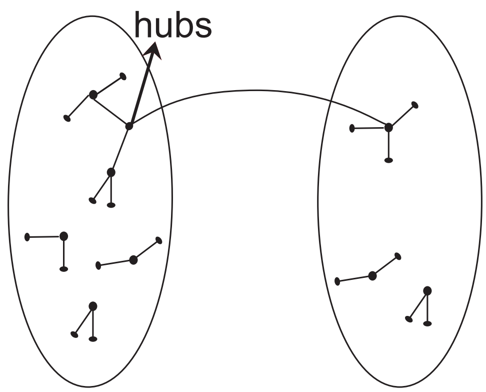
Let's first see why plain spectral clustering might fail. We have that the *ideal* matrix $E[A]$ has singular values $n(p+q), n(p-q), 0, \dots, 0$. The matrix that we see is $A = E[A] + A - E[A] = E[A] + R$. From the previous lecture, we would want that w.h.p[2] $\sigma_1(R) \leq \sqrt{np}$. If this were true then we can argue that the singular values of A are close to those of $E[A]$

[1] w.p = with probability
[2] w.h.p = with high probability

and hence the singular vectors will also be close. Unfortunately, the required bound on $\sigma_1(R)$ is not true for sparse graphs.

$$E[A] \to n(p+q), n(p-q) \begin{cases} p = \frac{a}{n} \\ q = \frac{b}{n} \end{cases} \to a+b, a-b$$

Sparse graphs look like this,



in which E[deg of vertex]=a+b. What we can show is that, $\exists$ hubs of $deg \geq \Omega(\frac{\log n}{\log\log n})$

$$\textbf{Adjacency matrix A} = \begin{array}{c} \qquad\quad \textit{hubs} \\ \begin{bmatrix} \dots & 1 & 1 & 1 & \dots \\ \dots & 1 & 1 & 1 & \dots \\ \dots & 1 & 1 & 1 & \dots \\ \dots & \vdots & \vdots & \vdots & \dots \end{bmatrix} \end{array} \tag{1}$$

Because of the presence of high degree vertices, $\sigma_1(A)$ will be quite large.

$$\begin{aligned} \sigma_1(A) =& max_{||x||=1}||Ax|| \\ \geq& \frac{\log n}{\log\log n} \end{aligned}$$

3 Regularized Spectral Clustering

We want to get rid of the points that are highly connected,

- A= adjacency matrix of G.
- remove all vertexes of $deg > 4(a+b)$
- let A' be new matrix
- run spectral clustering on A'

Theorem 2. *If* $(a-b) > \frac{c}{\epsilon^2}\sqrt{a}$ *then* $w.p \geq 1 - \frac{1}{n^3}$*, regularized spectral clustering outputs* S_1, S_2 *such that,*

$$(S_1 \Delta V_1) + (S_2 \Delta V_2) \leq \epsilon n$$

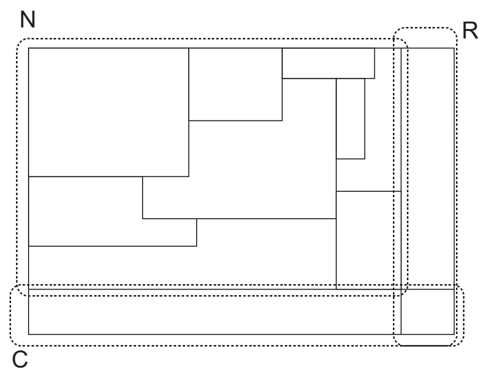
Proof. The proof is exactly the same as the last lecture as long as we can claim that $\sigma_1(A' - E[A]) \leq c\sqrt{a+b}$. This is formalized in the lemma below. $\square$

Lemma 1. *Let $A_{n\times n}$ be a random matrix with $A_{ij} \in \{0,1\}$. $E[A_{ij}] = p_{ij}$, let $d = n\max_{i,j} p_{ij}$ w.p $\geq 1 - \frac{1}{n^3}$, the following holds. Choose any set $10n/d$ rows/columns of A and reduce their weights. Let k' be the resulting matrix then $||A' - E[A]|| \leq O(\sqrt{d} + \sqrt{d'})$ where d' is the new maximum degree of the graph.*

4 Decomposition Theorem

The proof of the lemma will rely on the following decomposition theorem.

Theorem 3. *Let A be a random matrix such that $A_{ij} \in \{0,1\}$. $E[A_{ij}] = p_{ij}$ and $d = n \cdot max_{i,j} p_{ij}$. Then w.p $\geq 1 - \frac{1}{n^3}$, A can be decomposed into N, R, C*



- $||(A - E[A])_N \leq O(\sqrt{d})||$
- *each row of R has ≤ 64 1's and R intersects $\leq \frac{n}{d}$ columns*
- *each column of C has ≤ 64 1's and C intersects $\leq \frac{n}{d}$ rows*

Next, we will use the decomposition theorem to prove the lemma above. We will often use the following standard fact about matrices

Lemma 2. *Let A be a matrix such that the ℓ_1 norm of each row is bounded by a and the ℓ_1 norm of each column is bounded by b. Then $\|A\| \leq \sqrt{ab}$.*

5 Proof of Lemma 1

Proof. Let I be the indices of the rows modified and I' be the indices of the columns modified. Define the bad region E to be a union of disjoint sets E_1 and E_2, where $E_1 = I \times [n]$ and $E_2 = [n \setminus I] \times I'$.

$$||A' - E[A]|| \leq ||(A' - E[A])_N|| + ||(A' - E[A])_R|| + ||(A' - E[A])_C||$$

Let's first bound the first term on the right hand side. We have

$$\begin{aligned} ||(A' - E[A])_N|| &\leq ||(A - A')_N|| + ||(A - E[A])_N|| \\ &\leq ||(A - A')_N|| + O(\sqrt{d}) \text{ (by decomposition theorem)} \end{aligned}$$

Next we have

$$\begin{aligned} ||(A - A')_N|| &= ||(A - A')_{N\cap E}|| \text{ (A and A' are the same outside E)} \\ &\leq ||A_{N\cap E}|| \text{ (A dominates $A - A'$ entrywise)} \\ &\leq ||(A - E(A)_{N\cap E}|| + ||E[A]_{N\cap E}|| \\ &\leq ||(A - E[A])_{N\cap E_1})|| + ||(A - E[A])_{N\cap E_2}|| + ||E[A]_{N\cap E}|| \\ &\leq 2||(A - E[A])_N|| + ||E[A]_{N\cap E}|| \text{ (restricting onto a product decreases spectral norm)} \end{aligned}$$

For the final term we have

$$\begin{aligned} ||E[A]_{N\cap E}|| &\leq ||E[A]_{N\cap E_1}|| + ||E[A]_{N\cap E_2}|| \\ &\leq O(\sqrt{d}) + O(\sqrt{d}) \text{ (Using Lemma 2)} \end{aligned}$$

Now let's bound Term 2, i.e. $||(A' - E[A])_R||$. Term 3 can be bounded similarly. We have

$$||(A' - E[A])_R|| \leq ||A'_R|| + ||E[A]_R|| \leq O(\sqrt{d'}) + O(\sqrt{d}) \text{ (Using Lemma 2)}$$

we use the same procedure for term 3. □

In the next class we will prove the decomposition theorem.

6 Additional Reading

- A paper by Le et al. containing the decomposition theorem. `https://arxiv.org/abs/1506.00669`.