4 Eigenvalues & Expansion (Cheeger's Inequality)

CS 6810 – Theory of Computing, Fall 2012
Instructor: David Steurer
Scribe: Yang Yuan (yy528)
Date: 09/04/2012

Things talked in this lecture are useful not just for complexity, but for many other things, like networks.

4.1 Random Walks & Eigenvalues

Given a d-regular graph G with vertex set $V = \{1, \cdots, n\}$. Denote p to be a vector representing the initial probability distribution of the vertices, so the summation of all the elements in p equals to 1. Denote $p^{(t)}$ to be the probability distribution after t random steps in the graph. It is not hard to see that

$$p^{(t)} = G^t p.$$

Here G is a normalized adjacent matrix, i.e., $G = \frac{1}{d} \cdot AdjMatrix$ (G is also called the random-walk matrix).

The first thing to ask is the convergence of $p^{(t)}$. For example, if the initial distribution is uniform, $p^{(t)}$ is surely convergent to p itself. Denote the uniform distribution as u. Clearly $Gu = u$, i.e., u is the eigenvector of G with eigenvalue equals to 1.

Below we sort all the eigenvalues $\lambda_1, \cdots, \lambda_n$:

$$1 = \lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n.$$

Definition 4.1. G is *lazy* if there are $\frac{d}{2}$ self loops at every vertex.

That means, half of your moves will make you stay in the current position, which is why this property is called "lazy".

If G is diagonal dominant, then the eigenvalues of G are non-negative. That means, the eigenvalues of any lazy graphs are non-negative.

Theorem 4.2. *If G is lazy,*

$$||G^t p - u||_2 \leqslant \lambda_2^t \cdot ||p - u||_2.$$

Here u is the uniform distribution.

Proof. Let distribution $p = u + f$, here u is the uniform distribution. Since $\sum_i f_i = 0$, using inner product, we have $u \perp f$.

Since u is uniform distribution,

$$G^t p - u = G^t p - G^t u = G^t(p - u) = G^t f \tag{4.1.1}$$

And for all lazy graphs, we have,

$$\lambda_2 = \max_{f \perp u} \frac{< f, Gf >}{\|f\|_2^2} = \max_{f \perp u} \frac{\|Gf\|_2}{\|f\|_2}. \tag{4.1.2}$$

Using (4.1.2) for t times, we get

$$\|G^t f\|_2 \leqslant \lambda_2 \cdot \|G^{t-1} f\|_2 \leqslant \cdots \leqslant \lambda_2^t \cdot \|f\|_2.$$

Combined with (4.1.1), we are done. □

Theorem 4.2 says that the distance between the distribution after t steps and the uniform distribution is bounded by λ_2^t times the distance between the original distribution and the uniform distribution. Notice that $\|p - u\|_2$ is always at most 1.

It is worth pointing out that the differences between 2-norm and other norms are not too much. For example, we have

$$\|G^t P - u\|_1 \leqslant \sqrt{n} \cdot \lambda_2^t.$$

Corollary 4.3. *If $\lambda_2 = 1 - \gamma$, and we choose $t = \frac{10}{\gamma} \log n$, then we have $\|G^t P - u\|_2 < \frac{1}{n^{10}}$*

That means, after t steps, the probability distribution is really close to the uniform distribution. Although we need to take roughly $O(\log n)$ steps. Here, γ is the gap between the largest and the second largest eigenvalue. So γ is an important parameter which determines how long we need to walk to get a uniform distribution. If γ is a constant, then we only need to take logarithmic random steps.

Corollary 4.4. *We can solve undirected connectivity in L if γ is a constant, and $d = O(1)$. Graphs with this kind of properties are called* magical graphs*.*

Here d is bounded by constant because we want to visit all the neighbors, but if there are lots of neighbors, it is hard to visit them. In later lecture, we will see that general graphs can be reduced to this case.

4.2 Eigenvalues & Expansion (Cheeger's Inequality)

Below we want to understand λ_2 with a combinatorial interpretation.

Denote $\Phi(G)$ as the expansion (or conductance) of a graph. There are different ways to define it.

Definition 4.5.

$$\Phi(G) = \min_{S \subseteq V, |S| \leqslant \frac{n}{2}} \frac{E(S, \bar{S})}{d \cdot |S|}.$$

Note that the maximum number of edges that can go out of S is $d \cdot |S|$, so $\Phi(G)$ represents the minimum fraction of edges that actually go out of S.

Here, $\Phi(G) = 0.1$ means that for every set of vertices S, at least 0.1 fraction of the edges linking these vertices will go out of S.

And another definition of $\Phi(G)$ is,

Definition 4.6.

$$\Phi(G) = \min_{S \subseteq V, |S| \leqslant \frac{n}{2}} \frac{< 1_S, L_G 1_S >}{||1_S||_2^2}.$$

Here $L_G = I - G$ is the Laplacian of G, and 1_S is the indicator vector for S.

According to the definition,

$$\frac{< 1_S, L_G 1_S >}{||1_S||_2^2} = 1 - \frac{< 1_S, G 1_S >}{||1_S||_2^2}.$$

So we can compute $\Phi(G)$ by maximizing $\frac{<1_S, G1_S>}{||1_S||_2^2}$, which is similar to the equation (4.1.2). But they are taking the maximum value over different spaces. Equation (4.1.2) takes all possible vectors from continous space, but here we only take all possible vertex sets, which are in discrete space.

We want to compare $\Phi(G)$ and $\gamma(G)$. We want to show that they are closely related. If $\gamma(G)$ is large, then $\Phi(G)$ will be large, and vice versa. Specfically, we want to proof the following theorem:

Theorem 4.7 (Cheeger's inequality)**.**

$$\gamma(G)/2 \leqslant \Phi(G) \leqslant 2\sqrt{\gamma(G)}.$$

Proof. First, we want to show that $\Phi(G) \geqslant \frac{\gamma(G)}{2}$.

Let us investigate a distribution $p = \frac{1}{|S|} 1_S$. We will have

$$\Phi(G) = 1 - \frac{< p, Gp >}{||p||_2^2}.$$

We can write $p = u + f$, here $u \perp f$.

So we will have

$$\Phi(G) = 1 - \frac{< p, Gp >}{||p||_2^2} = 1 - \frac{||u||_2^2 + < f, Gf >}{||u||_2^2 + ||f||_2^2} \geqslant 1 - \frac{||u||_2^2 + \lambda_2 ||f||_2^2}{||u||_2^2 + ||f||_2^2}$$

$$= 1 - \frac{\|u\|_2^2 + (1-\gamma)\|f\|_2^2}{\|u\|_2^2 + \|f\|_2^2} = \frac{\gamma\|f\|_2^2}{\|u\|_2^2 + \|f\|_2^2}.$$

The inequality $< f, Gf > \leqslant \lambda_2\|f\|_2^2$ holds because f is in the eigenspace $< \lambda_2, \lambda_3, \cdots, \lambda_n >$.

Somehow we can show that $\|f\|_2^2 \geqslant \|u\|_2^2$. And then we can get the result.

Secondly, we want to show that $\Phi(G) \leqslant 2\sqrt{\gamma(G)}$.

Since $< f, Gf > \geqslant (1-\gamma) \cdot \|f\|_2^2$, we have $< f, L_G f > \leqslant \gamma\|f\|_2^2$. Below, we divide the proof into two parts.

Claim 4.8. For a function h, $< h, L_G h > \leqslant \varepsilon \cdot \|h\|_2^2$. Note that the support set h is limited, i.e., $|supp(h)| \leqslant \frac{n}{2}$. The existence of such h implies that, $\exists S, \Phi(G) \leqslant \sqrt{2\varepsilon}$, where $|S| \leqslant \frac{n}{2}$.

We want to first transform the existence of the distribution f to the existence of a function h, and then the existence of the function h implies the existence of a set S.

Proof of Claim 4.8. First, we may assume WLOG that $0 \leqslant h \leqslant 1$. We want a randomized algorithm that takes h and produces the set S. This randomized algorithm should somehow group all the vertices in S together in terms of making them linking to each other, rather than linking to vertices outside S (based on the definition of expansion).

We can choose some random $\tau \in [0, 1]$, Then $S = \{h_i^2 \geqslant \tau\}$. If two vertices are of the same value, then they will very likely to be inside or outside S.

Since the support set of h is less than $\frac{n}{2}$, $|S| \leqslant \frac{n}{2}$.

Based on the definition of S, it is easy to get the expection of the size of S:

$$\mathbb{E}\,|S| = \sum h_i^2 = \|h\|_2^2$$

For the expection of the number of edges going out of S, things are more tricky:

$$\mathbb{E}\,E(S, \bar{S}) = \frac{1}{2}\sum_i \sum_{j \in neighbor(i)} |h_i^2 - h_j^2|$$

Here $neighbor(i)$ is the set of vertices that link to i directly. The equality holds because only when $\tau \in [h_i^2, h_j^2] \cup [h_j^2, h_i^2]$, the edge (i, j) is going out from S to $\bar{S}$.

Using Cauchy – Schwarz inequality, we have,

$$\begin{aligned}
&\frac{1}{2}\sum_i \sum_{j \in neighbor(i)} |h_i^2 - h_j^2| \\
\leqslant &\frac{1}{2}\sum_i \sum_{j \in neighbor(i)} |(h_i - h_j)(h_i + h_j)| \\
\leqslant &\left(\frac{1}{2}\sum_i \sum_{j \in neighbor(i)} (h_i - h_j)^2\right)^{\frac{1}{2}} \left(\sum_i \sum_{j \in neighbor(i)} (h_i + h_j)^2\right)^{\frac{1}{2}}
\end{aligned}$$

$$=(d\cdot < h, L_G h >)^{\frac{1}{2}} \cdot (d\cdot < h, L_{-G} h >)^{\frac{1}{2}}$$
$$\leqslant d \cdot \sqrt{\varepsilon}||h||_2 \cdot \sqrt{2}||h||_2 = d \cdot \sqrt{2\varepsilon}||h||_2^2$$

The last inequality holds because $L_{-G} = I + G$, whose eigenvalue is at most 2. So based on the definition of $\Phi(G)$, we know $\Phi(G) = \frac{E(S,\bar{S})}{d\cdot|S|} \leqslant \sqrt{2\varepsilon}$. □

(the proof is not finished..)

□