

## Lecture 11

Instructor: Salil Vadhan

Scribes: Cathy Zhang and Rebecca Saul

## 1 Agenda

- Finishing up with the Power Method
- Expander Graphs: Measures of Expansion

## 2 Recap: Fiedler's Algorithm for Finding a Sparse Cut

Given a weighted, undirected graph  $G$ , we want to find an  $S \subseteq V$  with  $d(S) \leq \frac{d(V)}{2}$  and  $\phi(S) \lesssim \sqrt{2\nu_2} = O(\sqrt{\phi(G)})$ .

1. Find a vector  $y \perp \vec{d}$  such that  $\frac{y^T Ly}{y^T Dy} \lesssim \nu_2$ . By eigendecomposition of  $N$ , we can find such a  $y$  in  $O(n^\omega)$  time, where  $2 \leq \omega \leq 2.378 \dots$
2. Sort the components of  $y$  such that  $y(1) \leq y(2) \leq \dots \leq y(n)$ . We can sort in  $O(n \log(n))$  time.
3. Center the orthogonalized vector  $y$

Let  $z = y - s\vec{1}$  for appropriate  $s$ , so that the volume contribution by the vertices of the positive and negative entries of  $z$  are "centered":

$$\begin{aligned} \sum_{a:z(a)<0} d(a) &\leq \frac{d(V)}{2} \\ \sum_{a:z(a)>0} d(a) &\leq \frac{d(V)}{2}, \end{aligned} \tag{1}$$

Here, we just shift the vector so that half of its entries are negative while the other half is positive (weighted by the number of edges connected to each entry).

4. Find  $k \in \{1, \dots, n\}$  minimizing  $\frac{w(e(\{1, \dots, k\}, \{k+1, \dots, n\}))}{\min\{d(\{1, \dots, k\}), d(\{k+1, \dots, n\})\}}$ , where  $e(S, T) = \{(a, b) : a \in S, b \in T\}$ .

## 3 Power Method

Our motivation is to have a faster algorithm for the first step of Fiedler's algorithm, which involves finding a test vector with generalized Rayleigh quotient value approximately that of the second eigenvalue, and is orthogonal to the top eigenvector. We would like this step to take nearly linear time in the number of vertices and the number of edges ( $O(m+n)$ ). We can use the power method to estimate the largest eigenvalue, and a corresponding test vector, of a positive semi-definite (PSD) matrix  $M$ .

1. Choose  $x \stackrel{R}{\leftarrow} \{\pm 1\}^n$ .
2. Let  $y = M^k x$  for  $k = O(\frac{\log(n/\epsilon)}{\epsilon})$ .

The second step takes time  $O(k \cdot m)$ , where  $m$  is the number of nonzero entries in  $M$ .

**Theorem:** With constant probability,  $\mu_1 \geq \frac{y^T M y}{y^T y} \geq (1 - \epsilon)\mu_1$ .

How do we use the power method to estimate the second eigenvector? We can use the power method to estimate the second eigenvalue, and a corresponding test vector, of a positive semi-definite (PSD) matrix  $M$  by projecting the randomly chosen vector  $x$  onto the space orthogonal to  $v_1$ .

1. Choose  $x \stackrel{R}{\leftarrow} \{\pm 1\}^n$ .
2. Let  $x_0 = x - \langle x, v_1 \rangle v_1$ .
3. Let  $y = M^k x_0$  for  $k = O(\frac{\log(n/\epsilon)}{\epsilon})$ .

**Corollary:** With constant probability,  $\mu_2 \geq \frac{y^T M y}{y^T y} \geq (1 - \epsilon)\mu_2$  and  $y \perp v_1$ .

A generalization of this method for estimating the  $k^{\text{th}}$  eigenvalue does exist, but step 2, reducing the basis components, becomes much more difficult, as it requires us to know the first  $k - 1$  eigenvectors. Often, we work with matrices in which we already know the first eigenvector or can easily read it off from the graph (like the Laplacian matrix), but usually we do not know eigenvectors beyond that.

In the streaming PCA setting, we know an efficient algorithm to implement the Power Method and estimate the top  $k$  eigenvectors simultaneously. [1] [2]

Here we have shown how to estimate the largest and second largest eigenvalues of a PSD matrix, but in Fiedler's algorithm, we actually need an estimate corresponding to the second smallest eigenvalue of the Normalized Laplacian, which is a PSD matrix. To estimate the second smallest eigenvalue of the Normalized Laplacian, we apply our algorithm for estimating the second largest eigenvalue to the matrix  $2I - N = I + A = 2(D^{-1/2} \tilde{W} D^{1/2})$ , where  $A = I - N$  is the normalized adjacency matrix,  $N$  the normalized Laplacian, and  $\tilde{W}$  is the lazy random walk matrix. The reason we use  $2I - N$ , instead of just using  $A$ , for which we have a simple relation between the eigenvalues of  $A$  and  $N$ , is so that the matrix is PSD. The eigenvalues that result will be

$$2, 1 + \omega_2, 1 + \omega_3, \dots, 1 + \omega_n,$$

where  $1 + \omega_i = 2 - \nu_i$ . We will obtain a vector  $y \perp \tilde{d}^{1/2}$ , since  $\tilde{d}^{1/2}$  is the first eigenvector of  $N$ , such that

$$2 - \frac{y^T N y}{y^T y} = \frac{y^T (2I - N) y}{y^T y} \geq (1 - \epsilon)(2 - \nu_2).$$

We can rewrite this to see that

$$\frac{y^T N y}{y^T y} \leq (1 + \epsilon)\nu_2 + 2\epsilon.$$

We have to take  $\epsilon$  on the same order as  $\nu_2$ , to get a vector whose Rayleigh quotient is comparable to  $\nu_2$ . By taking  $\epsilon = \frac{\nu_2}{4}$ , we get a running time for the Power Method of  $O(m \cdot \frac{\log(n/\nu_2)}{\nu_2})$ . This is not great, because we know that there are graphs where  $\nu_2$  is polynomially small - for example,  $\nu_2$  is on the order of  $\frac{1}{n^2}$  in the cycle. So this algorithm only approaches linear time on graphs where  $\nu_2$  is fairly large.

A better approach is to apply the power method to the "pseudoinverse"  $N^+$  of the Normalized Laplacian, which has eigenvalues  $0, \frac{1}{\nu_2}, \frac{1}{\nu_3}, \dots, \frac{1}{\nu_n}$ . This allows us to compute a relative approximation on  $\frac{1}{\nu_2}$  instead of on  $2 - \nu_2$ , which gives us a better approximation when  $\nu_2$  is small, as a relative approximation on the inverse is just as good as a relative approximation on  $\nu_2$ , but this is not true in the additive case.

Later in the course, we will see algorithms that, given  $G$  and  $x$ , compute  $N^+ x$  in  $\tilde{O}(m + n)$  time, which will allow us to apply Fiedler's algorithm, and find a cut of conductance  $O(\sqrt{\nu_2})$ , in  $\tilde{O}(m + n)$  time.

## 4 Expander Graphs

Expander graphs have many applications in theoretical CS, and one of the best ways to understand them is spectrally. We will restrict our attention to  **$d$ -regular,  $n$ -vertex digraphs**, with  $n \rightarrow \infty$ . We will also allow multigraphs. Expander graphs are required to be **sparse**, which typically means  $d = O(1)$ , or occasionally,  $d = \text{polylog}(n)$ .

We also want expander graphs to be **well-connected**, a notion for which we have several characterizations.

1. **Spectral Expansion**  $\gamma$ :  $\gamma(G) = 1 - \omega(G) \geq \gamma$ , where

$$\omega(G) = \omega_u(G) = \max_p \frac{\|Wp - u\|}{\|p - u\|} = \max_{x \perp u} \frac{\|Wx\|}{\|x\|}$$

with  $W$  the random walk matrix of  $G$ ,  $p$  a probability distribution, and  $u$  the uniform distribution  $\frac{\mathbf{1}}{n}$ , which is the stationary distribution of a regular graph. The spectral expansion will be between 0 and 1. When  $\gamma$  is close to 1, that means  $\omega(G)$  is close 0, so random walks mix quickly. Typically, we want  $\gamma > 0$  independent of  $n$ , and ideally, we maximize  $\gamma = 1 - \omega$  as a function of  $d$ , so that  $\gamma$  approaches 1 as  $d$  gets large.

Two examples of graphs that do not satisfy the spectral expansion condition are the directed and undirected  $n$ -cycle. The directed  $n$ -cycle has  $\omega(G) = 1$ , so  $\gamma(G) = 0$ . The undirected  $n$ -cycle has  $\gamma \geq \theta(\frac{1}{n^2})$ . An example that satisfies this spectral expansion condition is the complete graph without self-loops, where we have  $\gamma(G) = 1 - \frac{1}{n-1}$ . However this graph is not sparse.

We are looking for graphs with  $\gamma$  strictly positive, ideally  $\gamma$  very close to 1, but sparse, with very few edges. These will be good spectral expanders.

2.  **$(k, a)$  Vertex Expansion**: A graph is a  $(k, a)$  vertex expander if the following condition holds:

$$\forall S, \text{ such that } |S| \leq k, |\mathcal{N}(S)| \geq a|S| \text{ holds.}$$

Here  $\mathcal{N}(S) := \{(u, v) \mid \exists v \in S, (u, v) \in E\}$ , which denotes the neighborhood of  $S$ .

Typically we would like  $k = O(n), a = 1 + O(1)$ . Ideally maximize  $a$  as a function of  $n, k, d$

3.  **$(k, \epsilon)$  Edge Expansion**: A graph is a  $(k, \epsilon)$  edge expander if the following condition holds:

$$\forall S, \text{ such that } |S| \leq k, |e(S, S^c)| \geq \epsilon \cdot d \cdot |S| \text{ holds.}$$

Here  $e(S, S^c) := \{(u, v) \mid u \in S, v \in S^c\}$ , which denotes the edges from  $S$  that lead outside of  $S$ .

Typically we would like  $k = O(n), \epsilon = O(1)$ . Ideally maximize  $\epsilon$  as a function of  $n, k, d$ . Note that the notion of edge expansion is closely related to conductance and isoperimetry. Note that

$$\phi(G) = \min_{|S| \leq n/2} \frac{|e(S, S^c)|}{d \cdot |S|},$$

thus a graph  $G$  is a  $(n/2, \phi(G))$  edge expander.

## 5 Edge Expansion v.s. Spectral Expansion

In this section we will look into the relationship between edge expansion and spectral expansion.

**Theorem 1.** *Let  $\mathcal{M}$  be an infinite family of regular undirected, lazy graphs. The following statements are equivalent.*

1.  $\exists \gamma > 0$ , such that every  $G \in \mathcal{M}$  has spectral expansion at least  $\gamma$ .
2.  $\exists \epsilon > 0$ , such that every  $G \in \mathcal{M}$  is an  $(N/2, \epsilon)$  edge expander.

*Proof.* From the argument stated in the previous section, we know that statement 1 ‘ $\exists \epsilon > 0$ , such that every  $G \in \mathcal{M}$  is an  $(N/2, \epsilon)$  edge expander’ is equivalent to ‘ $\exists \epsilon > 0, \forall G \in \mathcal{M}, \phi(G) \geq \epsilon$ ’. By Cheeger’s Inequality we have:

$$\frac{\nu_2}{2} \leq \phi(G) \leq \sqrt{2\nu_2}.$$

Generally speaking,  $\gamma(G) \leq \nu_2$ , but since the graph is undirected and lazy, we know that  $w_u(G) = \max\{w_2, -w_n\} = w_2$ , thus  $\gamma(G) = \nu_2$ . Then we have:

$$\frac{\gamma(G)}{2} \leq \phi(G) \leq \sqrt{2\gamma(G)},$$

i.e.  $\phi(G)$  can be both bounded above and below by  $\gamma(G)$ , which completes the proof. □

**Remark 1.** *Statement 1 implies statement 2, even without the assumption of laziness, because  $\frac{\gamma(G)}{2} \leq \frac{\nu_2}{2} \leq \phi(G)$ . But from statement 2 to statement 1, the assumption of laziness is necessary.*

Next class we are going to learn about the relationship of vertex expansion and spectral expansion.

## References

- [1] Zeyuan Allen-Zhu and Yuanzhi Li. First efficient convergence for streaming k-pca: a global, gap-free, and near-optimal rate, 2017.
- [2] Chi-Ning Chou, Mien Brabebe Wang, and Tiancheng Yu. A general framework for analyzing stochastic dynamics in learning algorithms, 2020.