CS 229r Spectral Graph Theory in Computer Science, Lecture 10 Oct 6, 2020

Lecture 10

Instructor: Salil Vadhan *Scribe: James Conant*

1 Agenda

- Recap of Cheeger's Inequality
- Fiedler's Algorithm for Sparsest Cut
- The Power Method

2 Recap

Conductance, denoted by $\phi(S)$ is the ratio of the weight of edges leaving a set S to its degree, where $w(\partial S) = \sum_{a\in S, b\notin S} w(a,b)$ and $d(S) = \sum_{a\in S} d(a)$.

$$\phi(S) = \frac{w(\partial S)}{d(S)}$$

Cheeger's Inequality, where ν_2 is the second eigenvalue of the normalized Laplacian N.

$$\nu_2/2 \le \phi(G) \le \sqrt{2\nu_2}$$

Higher-Order Cheeger

$$\nu_k/2 \le \phi_k(G) \le \text{poly}(k)\cdot\sqrt{\nu_k}$$

$$\phi_k(G) = \min_{S_1,\dots S_k \text{disjoint}} \max\{\phi(S_1),\dots,\phi(S_k)\}$$

3 Fiedler's Algorithm

Sparsest Cut Problem: Given **weighted, undirected** G, find $S \subseteq V$ s.t. $d(S) \le d(V)/2$ and $\phi(S) = \phi(G)$. This is NP-Hard to solve exactly. Fiedler's algorithm, enumerated below, produces a cut S s.t. $\phi(S) \lesssim \sqrt{2\nu_2}$. The proof of correctness can be found in Section 4.4 of Trevisian.

1. Find a vector $y \perp \vec{d}$ s.t.
$$\frac{y^\top Ly}{y^T Dy} \lesssim \nu_2$$
2. Sort y by rearranging vertices s.t.
$$y(1) \le y(2) \le \dots \le y(n)$$
3. Find $k \in \{1,\dots,n-1\}$ minimizing
$$\frac{w(e(\{1,\dots,k\},\{k+1,\dots,n\}))}{\min(d(\{1,\dots,k\}), d(\{k+1,\dots,n\})}$$
$$e(S,T) = \{(a,b) : a\in S, b\in T\}$$

What is the runtime of Fiedler's algorithm? Using eigendecomposition of the normalized Laplacian, step 1 takes $\mathcal{O}(n^\omega)$ time where ω is the matrix multiplication constant ≈ 2.378, since the runtime of eigendecomposition is similar to that of matrix multiplication (see https://arxiv.org/pdf/math/0612264.pdf). Step 2 takes $\mathcal{O}(n \log n)$ time, as we need to sort n vertices. Step 3 naively takes $\mathcal{O}((n-1) \cdot m)$ time if we do a linear search across all values of k, but we can shave off the $n-1$ factor. To see this, look at the progression of steps from $k = 1$ to $k = 2$.

$$\frac{w(e(\{1\}, \{2, \ldots, n\}))}{\min(d(\{1\}), d(\{2, \ldots, n\})}$$

$$\frac{w(e(\{1, 2\}, \{3, \ldots, n\}))}{\min(d(\{1, 2\}), d(\{3, \ldots, n\})}$$

The first computation takes $\mathcal{O}(d(1))$ time. For the second computation, we are just adding vertex 2, so we only need to consider each of the edges connected to vertex 2, then reference the result from the first computation. This step thus only takes $\mathcal{O}(d(2))$ time. This holds for $k > 2$. The total time taken is proportional to the total number of edges in the graph m.

m is $\mathcal{O}(n^2)$ even for the densest graphs. Thus, the runtime of Step 1, $\mathcal{O}(n^\omega)$ will dominate. Step 1 will be $\mathcal{O}(n^\omega)$ even if our graph is sparse and m is $\mathcal{O}(n)$. Is there a way we speed up Step 1 so that a modification of Fiedler's algorithm can run in time $\tilde{\mathcal{O}}(n + m)$? Yes, we can apply the power method to approximately compute the second eigenvector corresponding to ν_2.

4 The Power Method

The following is a procedure for estimating the largest eigenvalue and the corresponding eigenvector of a positive semi-definite matrix M.

1. Choose $x \overset{R}{\leftarrow} \{\pm 1\}^n$. We choose from the hypercube as a proxy for the unit sphere.
2. Let $y = M^k x$ for
$$k = \mathcal{O}\left(\frac{\log(n/\epsilon)}{\epsilon}\right)$$

Theorem 1. *With constant probability (i.e. $\geq 3/16$),*

$$\frac{y^\top M y}{y^\top y} \geq (1 - \epsilon)\mu_1$$

Proof. M is positive semi-definite, so it is symmetric. By the spectral theorem, we can write x as a linear combination of orthogonal eigenvectors $v_1, \ldots, v_n$ with eigenvalues $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$.

$$x = c_1 v_1 + \ldots + c_n v_n$$

Since $y = M^k$,

$$y = c_1 \mu_1^k v_1 + \ldots + c_n \mu_n^k v_n$$

The intuition for why this works is that as we apply M over and over, the term with the largest eigenvalue dominates. We just need to prove that k is good enough to achieve the bound, and make sure that we don't get unlucky and choose a start vector x orthogonal to v_1 (this would make $c_1 = 0$). We will prove $c_1 \geq 1/2$ with constant probability $\geq 3/16$ below, and leverage this fact in the rest of this proof. This ensures the initial v_1 component is not too small.

Let $\ell \in \{1, \ldots, n\}$ be the value such that $\mu_\ell \geq (1-\epsilon)\mu_1$ and $\mu_{\ell+1} < (1-\epsilon)\mu_1$. Let's look at the numerator of the Rayleigh quotient first. We want to lower bound the numerator and upper bound the denominator.

$$y^\top M y = c_1^2 \mu_1^{2k+1} + \ldots + c_n^2 \mu_n^{2k+1}$$

M is positive semi-definite, so all eigenvalues must be ≥ 0. Therefore, each term in this sum contributes an amount ≥ 0.

$$y^\top My \geq c_1^2\mu_1^{2k+1} + \ldots + c_\ell^2\mu_\ell^{2k+1}$$

We chose ℓ s.t. $\mu_\ell \geq (1-\epsilon)\mu_1$, so we can "factor" out a $(1-\epsilon)\mu_1$ to lower bound this expression by

$$y^\top My \geq (1-\epsilon)\mu_1 \left(c_1^2\mu_1^{2k} + \ldots c_\ell^2\mu_\ell^{2k}\right)$$

Now for the denominator. We are looking for an upper bound.

$$\begin{aligned}
y^\top y &= c_1^2\mu_1^{2k} + \ldots + c_n^2\mu_n^{2k} \\
&\leq \sum_i^\ell c_i^2\mu_i^{2k} + \sum_{i=\ell+1}^n c_i^2((1-\epsilon)\mu_1)^{2k} && \text{since } \mu_{\ell+1} < (1-\epsilon)\mu_1 \\
&\leq \sum_i^\ell c_i^2\mu_i^{2k} + n\cdot\mu_1^{2k}\cdot(1-\epsilon)^{2k} && (1-\epsilon)^{2k} = \epsilon/n \text{ by choice of k} \\
&\leq \sum_i^\ell c_i^2\mu_i^{2k} + 4\cdot\epsilon\cdot c_1^2\cdot\mu_1^{2k} && \text{leveraging } c_1 \geq 1/2 \\
&\leq (1+4\epsilon)\cdot\left(c_1^2\mu_1^{2k} + \ldots + c_\ell^2\mu_\ell^{2k}\right)
\end{aligned}$$

Now, since the two sums cancel, we have

$$\frac{y^\top My}{y^\top y} \geq \frac{1-\epsilon}{1+4\epsilon}\mu_1 = (1-\mathcal{O}(\epsilon))\mu_1$$

This procedure will only work when $c_1 \geq 1/2$, which happens with probability $\leq 3/16$. We can use success amplification to ramp this probability up to ≥ 0.99. The runtime of the power method is $\mathcal{O}(k\cdot(n+m))$, where m is the number of nonzero entries in M. □

Theorem 2. $c_1 \geq 1/2$ *with probability* $\geq 3/16$

Proof. Recall that $x(a)$'s are independent ± 1 random variables.

$$c_1 = x^\top v_1 = \sum_a x(a)v_1(a)$$

$$\mathbb{E}[c_1] = \sum_a \mathbb{E}[x(a)]v_1(a) = 0$$

For the second moment, note that $\mathbb{E}[x(a)x(b)] = 0$ if $a \neq b$.

$$\mathbb{E}[c_1^2] = \sum_{a,b}\mathbb{E}[x(a)x(b)v_1(a)v_1(b)] = \sum_a (v_1(a))^2 = 1$$

From Trevisian Lemma 9.8, we know

$$\mathbb{E}[c_1^4] \leq 3\left(\sum_a v_1(a)^2\right)^2 - 2\sum_a v_1(a)^4 \leq 3$$

We don't want to use Chebyshev's or Hoeffding's inequality here, since we want "anti-concentration," the chance that the random variable c_1 is far away from its mean (if c_1 is close to its mean 0, x is more orthogonal to v_1). Referencing the proof in Trevisian, we apply the Paley-Zygmund inequality on c_1^2, taking $\delta = 1/4$. We choose $\delta = 1/4$, since we want $\mathbb{P}[c_1 \geq 1/2]$.

$$\mathbb{P}[c_1^2 \geq \delta\mathbb{E}[c_1^2]] = \mathbb{P}[c_1^2 \geq 1/4] \geq (1-\delta)^2\frac{(\mathbb{E}[c_1^2])^2}{\mathbb{E}[c_1^4]} \geq \left(\frac{3}{4}\right)^2\frac{1}{3} = 3/16$$

□

5 Breakout Question

Why can't we use the Power Method to show that the lazy random walk on an undirect n-cycle mixes in time $\mathcal{O}(\log n)$? We are interested in the time of convergence to the stationary distribution π, which is this eigenvector of the largest eigenvalue $\mu_1 = 1$ of $\tilde{W}$. This time of convergence would be the k we choose as our parameter for the power method, and $\tilde{W}$ is positive semi-definite on regular graphs (the undirected cycle is 2-regular). So why doesn't the Power Method show that the mixing time is $\mathcal{O}(\log n)$? The Power Method gives us y s.t.

$$\frac{y^\top \tilde{W} y}{y^\top y} \geq (1-\epsilon)\mu_1$$

We apply M a total of k times, with

$$k = \mathcal{O}\left(\frac{\log(n/\epsilon)}{\epsilon}\right)$$

Recall that the second eigenvalue μ_2 of $\tilde{W}$ is $1 - \Theta(1/n^2)$. If $\epsilon \geq \Theta(1/n^2)$, then the Power Method does not imply that y is close to π. Since $\mu_2 \geq (1-\epsilon)\mu_1$, the resulting y may be closer to the span of the second eigenvector v_2. Therefore, we do not have a guarantee that the random walk will mix in k steps.

6 The Power Method for the Second Eigenvector

If we know the first eigenvector v_1 of M, we can alter the Power Method to approximate the second eigenvector v_2. We still choose x at random from the hypercube $\{\pm 1\}^n$. We then project x onto the subspace orthogonal to v_1 to obtain x_0.

$$x_0 = x - \langle x, v_1 \rangle$$

This ensures that the coefficient of c_1 will be 0. Then, the term with the second largest eigenvalue in the linear combination of eigenvectors will dominate. The proof is roughly the same as above.

We want the second smallest eigenvalue ν_2 of the normalized Laplacian for Fiedler's algorithm. We can approximate the second largest eigenvalue μ_2 of the normalized adjacency matrix with the Power Method. This works since we can read off the first eigenvector from the degree vector. This gives us ν_2, since $\nu_2 = 1 - \mu_2$.