CS270 Algorithms & Data Structures **Spring 2008**

Lecture 5 Eigenvalues and cuts. : 2.05.08

Lecturer: Satish Rao *Scribe: Rao, for now*

Disclaimer: *These are rough notes, with some exercises.*

Preparation: Here, we took a lot from Hastad's course notes, Arora's course notes, and read a bunch of notes for the planar part from Spielman, and Kleinberg. We also used a short note by James Lee.

5.1 Reading for next time.

Chapter 4. Algorithms.

5.2 A short though perhaps (more) mysterious proof.

(Modified from Arora's class notes.)

Question: What is the Cauchy-Schwatz inequality?

For $a_1, \ldots, a_n; b_1, \ldots b_n$ we have

$$\sum_{i=1}^{n} a_i b_i \leq (\sum_i a_i^2)^2 (\sum_i b_i)^2.$$

Question: Recall the definition of λ?

$$\lambda_g = \frac{\sum_{e=u,v}(x_u - x_v)^2}{\sum_i x_i^2},$$

where $\sum_i x_i = 0$.

We will show that $\lambda_g > \alpha^2/2d$.

Question: Split into positive x_i and negative x_i. Only consider the positive. What?

Ok, any pair i, j that crosses 0, has $(x_i - x_j)$ of at least $x_i^2 + x_j^2$. So, we can just consider the ratio for say the left side.

That is, consider u_i where all the u_i's are positive. That is, we need to show

$$\frac{\sum_{e=(i,j)}(u_i - u_j)^2}{\sum_i u_i^2} \geq \alpha^2/2d,$$

where the u_i's are positive.

Question: What next?

Multiply both sides by the same quantity.

$$\frac{\sum_{e=(i,j)}(u_i-u_j)^2}{\sum_i u_i^2} = \frac{\sum_{e=(i,j)}(u_i-u_j)^2}{\sum_i u_i^2} \times \frac{\sum_{e=(i,j)}(u_i+u_j)^2}{\sum_{e=(i,j)}(u_i+u_j)^2} \tag{5.1}$$

$$\geq \frac{\left[\sum_{e=(i,j)}(u_i-u_j)^2\right]\left[\sum_{e=(i,j)}(u_i+u_j)^2\right]}{\sum_i u_i^2 2\sum_{e=(i,j)}(u_i^2+u_j^2)} \tag{5.2}$$

The last inequality comes from $(a+b)^2 \leq 2(a^2+b^2)$.

Question: Can we apply Cauchy-Schwartz to the numerator?

Sure.

$$\geq \frac{\left[\sum_{e=(i,j)}(u_i-u_j)\sum_{e=(i,j)}(u_i+u_j)\right]^2}{2d(\sum_i u_i^2)^2} \tag{5.3}$$

$$\geq \frac{\left[\sum_{e=(i,j)}(u_i^2-u_j^2)\right]^2}{2d(\sum_i u_i^2)^2} \tag{5.4}$$

Question: Let C_k be the number of edges between the first k vertices and the rest. Can we rewrite $\sum_{e=(i,j)}(u_i^2-u_j^2)$ in terms of C_k?

Sure. Sum over the cuts of the distance $(u_k^2 - u_{k+1}^2)$ that edges hop in each cut versus over distances between endpoints of edges.

$$\sum_{e=(i,j)}(u_i^2-u_j^2) = \sum_k (u_k^2-u_{k+1}^2)C_k.$$

Question: Is there a lower bound on C_k?

Sure. $C_k > \alpha k$.

Question: Then what?

We should plug in to equation 5.2.

$$\frac{\sum_{e=(i,j)}(u_i-u_j)^2}{\sum_i u_i^2} \geq \frac{(\alpha\sum_k(u_k^2-u_{k+1}^2)k)^2}{2d(\sum_i u_i^2)^2} \tag{5.5}$$

$$= \alpha^2\frac{\left(\sum_k u_k^2\right)^2}{2d(\sum_i u_i^2)^2} \tag{5.6}$$

$$= \frac{\alpha^2}{2d} \tag{5.7}$$

Question: Are we done?

Yes.

Exercise 1: Show that finding the minimum value of

$$\frac{\sum_{e=(i,j)} |x_i - x_j|}{\sum_{i,j} |x_i - x_j|}$$

for a graph is NP-complete. (Note we replace $(x_i - x_j)^2$ by $|x_i - x_j|$ You may use the fact that computing α or the sparsity or the maximum cut in a cut is NP-complete.

5.3 Application to randomized algorithms.

Question: What is a example form of a randomized algorithm?

Given a language (primes), a randomized algorithm is a procedure $M(x, r)$, such that if x is prime $M(x, r) = 1$ for all r and when x is composite $M(x, r) = 1$ for fewer than β of the strings r.

Question: How do we get a lower probability of failure?

Run $M(x, r)$, k times, and we get a probability of failure of $1/2^k$.

Question: How many random bits do we use?

$k|r|$.

Question: For a degree 4 graph with $\lambda_2 < 1 - \delta$. Why is it an expander?

The cut size is at least $\Omega(\sqrt{\delta}n)$ from Cheeger above.

Question: For such a graph on $N = 2^|r|$ nodes, how should we run the algorithm M?

Pick a random node in the graph, and take a random walk of length l. Run $M(x, r)$ for each node encountered.

Question: How many random bits is this?

$|r| + l \log 4$.

Question: Better perhaps by a factor of k. How do we bound the failure probability. Where to start?

The bad set of strings for a "composite" x. Say S.

Question: Size of S?

$|S| \leq \beta N$.

Question: What must happen for algorithm to fail?

Bound probability that we stay in S for l steps?

Question: What happens at beginning?

Uniform over G, i.e., $v = (1/N, \ldots, 1/N)$

Question: How does probability evolve?

Bad probability only when in S. So, we want to bound the survival probability, that is the probability that remains in S after l steps.

First, if out of S, its all good. So, project onto S, (zero out all components not in S). Then take a step in the graph.

Question: Can we view this as a matrix multiply?

Sure P projects onto S, and B is the "random walk matrix", e.g., 1/2 on the diagonal and $1/2d$ on the "edges".

Question: How do we proceed?

Say that $||v||$ decreases, since that is what we can bound in terms of multiplication.

Question: Is there a way to break down v?

Sure, a component along the first eigenvector v_1 and the rest, v_r.

Question: Why does v_r decrease?

We multiply by B there the norm decrease by a factor of λ_2.

Question: Why does v_1 decreases?

We project v_1 onto P, thus we only get β of its square norm, or $\sqrt{\beta}$ of its norm. This follows from the fact that all the entries are equal in v_1.

Question: Thus, in each step t the norm of v^t is at most?

$$||v^t||^2 \leq \min(\beta, \lambda_2)||v^t||.$$

Question: After l steps, the failure probability is at most what?

$$||v^t||_1 \leq \sqrt{N}||v^t||.$$

Question: How many steps to be all good, e.g. $< \beta^k$.

We get the following sequence.

$$||v^t||_1 \leq \sqrt{N}||v^t|| \leq (\lambda_2^l)\sqrt{N} \leq \beta^k.$$

Thus, $l > -\log_{\lambda_2}|N| + klog_{\lambda_2}\beta$.

Or if $\lambda_2 < 1 - \delta$, we get that, $l > \log_\delta |N| + k \log_2 1/(\delta\beta)$.

Compare this to $k \log |N|$. We only need constant extra bits to get constant factor improvement in the failure probability.

Exercise 2: Consider extending the argument above for an algorithm with two-sided error. That is, the machine either accepts with probability at least 2/3 for x is in the language and accepts with probability at most 1/3 if x is not. How would you start a proof? Specifically, what would you need to prove about random

walks of length l. (Hint: if you chose a random string and run $M(\cdot, s)$ on a node in the language versus a node not in the langauge, what is probability of acceptance? What if you do this many times? Now, for a random walk you have many strings. What should the property be for these strings?) This is a one sentence answer.

Exercise 3: Consider generating a graph on n nodes, where each node chooses 5 neighbors uniformly at random. Show that this graph is an expander graph. That is, there exists a constant β such that for all $|S| \leq n/2$ nodes, that the number of edges crossing the corresponding cut is at least $\beta|S|$. (BTW, this is bad for several reasons, one being that some nodes have high degree, but the argument about expansion is easier this way.)

5.4 Explicit construction.

Question: Boy, we were trying to save randomness, and used a huge amount to build an expander? Can we do better? What would we like?

A construction that given the name of a node, we can compute its neighbors.

The Gabber-Galil constuction is as follows, A node is an element of $Z_m \times Z_m$, i.e., of the form (x, y). It is connected to $(x+2y, y), (x+2y+1, y), (x, x+2y), (x, x+2y+1)$ (and the reverse edges.) The original proof was pretty fancy and difficult using for example Fourier transforms.

Later, it made simpler by using the eigenvalue gap, which was shown to be slightly less than 1/8. Search for ODonnell's ruminations on "Explicit expander exaggerations" on the internet (on Fortnow's blog.)

5.5 Planar graphs have small eigenvalue.

Question: Recall the "gap" examples for the eigenvalue/cut relationship?

The path.

Question: Does the second eigenvector give you a good cut?

Yes!

Question: But is the Cheeger inequality tight?

No.

Question: For some graphs?

Sort of.

Question: What size cut does a bounded degree planar graph always have?

$O(\sqrt{n})$

Question: What is the edge expansion of such a cut?

$O(1/\sqrt{n})$.

Question: What is an upper bound on the eigenvalue gap?

Well, it must be at most $1/\sqrt{n}$?

Question: If it were $\Theta(1/n)$, what would be the expansion of a cut from the "Cheeger" procedure?

$1/\sqrt{n}$.

Question: What happens with the grid?

$1/n$.

Question: What happens with the tree?

$1/n$.

Question: Hmm. Shall we conjecture?

It is $O(1/n)$ for all planar graphs.

Question: Why?

For a planar graph, there is an kissing disk embedding on the sphere such that the center of mass of the centers of the disks is the center of the spheres. (Spielman-Teng result.)

Question: So what? Can we get a bound on the $\lambda_2(G)$?

Well, recall that we want a set of numbers $x_0, \ldots, x_{n-1}$ such that

$$\frac{\sum_{e=(i,j)} (x_i - x_j)^2}{|x|^2} = O(1/n),$$

where $\sum_i x_i = 0$.

Question: How do the two meet?

Take a unit sphere, use the vectors as the x_i's (instead of real numbers).

Question: That's ok, right?

You did an excercise, showing the other way. Your solution works either way.

Question: Bound the denomiator?

We take the sphere to be unit radius, and the denominator is n.

Question: Bound the numerator?

The numerator is the sum over edges. An edge between i and j, has $(v_i - v_j)^2 = (r_i + r_j)^2$ in the kissing disk embedding, where r_i is the radius of the circle that represents i. Continuing, we get

$$(v_i - v_j)^2 \leq (r_i + r_j)^2 \leq 2r_i^2 + 2r_j^2.$$

Question: Continuing? Let's sum over the edges.

We get that each circle is hit at most degree times.

That is,

$$\sum_{e=(i,j)} (v_i - v_j)^2 \leq 2d \sum_i r_i^2.$$

Question: Can we bound the $\sum_i r_i^2$?

Well, $\sum_i \pi r_i^2$ is at most the surface area of the sphere, i.e., 4π.

Thus,

$$\sum_{e=(i,j)} r_i^2 \leq 8d.$$

Question: What is the bound on $\lambda_g(G)$?

At most $8d/n$!

Question: What upper bound does this give on $\alpha(G)$?

At most $8d/n$!

Exercise 4: Show that the eigenvector algorithm, i.e., sort according to eigenvalue and pick best "prefix" cut, gives a cut in a constant degree planar graph with expansion $O(1/\sqrt{n})$.

5.6 If there is time: The Cauchy-Schwartz inequality.

Question: Consider $u = (2,3)$ and $v = (4,4)$? Which is bigger; $|u \cdot v|$ or $|u||v|$?

Ok, $8 + 12 = 20$ is smaller than $\sqrt{(13)(32)} = \sqrt{320 + 96}$.

Question: Is this true for any vector?

It is the Cauchy-Schwartz inequality. That is, for any two vectors x and y,

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle \cdot \langle y, y\rangle$$

Question: Possible intuition? Consider x is unit vector? What is $\langle x, y\rangle$?

The left hand side is the length of the projection of the vector y onto x. The right hand side is the length of the vector y. The projection should be smaller than the whole length.

Question: Towards a real proof, is it true when y or x is the zero vector?

Of course.

Question: Give a proof for it?

$$\langle \frac{x}{|x|} - \frac{y}{|y|}\rangle \quad \geq \quad 0 \tag{5.8}$$

$$\langle\frac{x}{|x|},\frac{x}{|x|}\rangle+\langle\frac{y}{|y|},\frac{y}{|y|}\rangle-2\langle\frac{x}{|x|},\frac{y}{|y|}\rangle \geq 0 \tag{5.9}$$

$$1+1-2\langle\frac{x}{|x|},\frac{y}{|y|}\rangle \geq 0 \tag{5.10}$$

The inequality follows.