

Lecture 6: Cheeger's inequality

In this lecture we prove Cheeger's inequality.

Disclaimer: These lecture notes are informal in nature and are not thoroughly proofread. In case you find a serious error, please send email to the instructor pointing it out.

Expansion and $\lambda_2(L_G)$

Recall our definitions of the Laplacian, $L_G = dI - A_G$ and expansion,

$$\Phi(G) = \min_{|S| \leq n/2} \phi(S) = \min_{|S| \leq n/2} \frac{E(S, \bar{S})}{d|S|}.$$

We saw that $\lambda_1(L_G) = 0$, and

$$\lambda_2(L_G) = \min_{\sum_i x_i = 0} \frac{\sum_{ij \in E} (x_i - x_j)^2}{\sum_i x_i^2}.$$

We also saw that

$$\Phi(G) \geq \frac{1}{4d} \lambda_2(L_G).$$

Next, we stated Cheeger's inequality, which states that λ_2 is not too much smaller than $\Phi(G)$:

Theorem 1. *Let G be a d -regular graph. Then we have*

$$\Phi(G) \leq \sqrt{\frac{2\lambda_2(L_G)}{d}}.$$

The theorem was proved in the context of manifolds by Cheeger, and was proved for graphs by Alon and Milman in 1985. This is the proof we will present. First, let us try to get a better feel for the quantity $\lambda_2(L_G)$.

Embedding a graph on a line. Consider the following problem: we have a graph G , and we would like to map the vertices to points on the real line, such that pairs (i, j) that are adjacent in the graph are *close*. Formally, we want an $f : V \mapsto \mathbb{R}$, such that $\sum_u f(u) = 0$, not all $f(u)$ are zero, and it minimizes the ratio

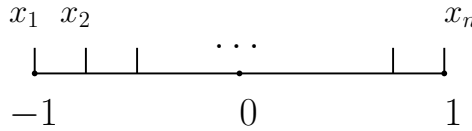
$$\frac{\sum_{ij \in E} (f(i) - f(j))^2}{\sum_i f(i)^2}.$$

Exercise. Show that we do not need to impose the condition $\sum_u f(u) = 0$, i.e., minimizing the ratio subject to the requirement that not all $f(i)$ are equal has the same optimum value.

Note that the optimal solution to the line embedding problem is $f(u) = x_u$, where x is the eigenvector of L_G corresponding to λ_2 . This formulation often gives intuition about λ_2 .

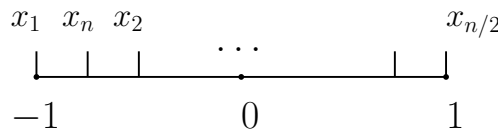
For instance, consider the case when G is a cycle with n vertices. It is easy to argue that $\Phi(G) \geq \Omega(1/n)$.

Now, let us try to embed the vertices onto the line such that $\sum_{ij \in E} (x_i - x_j)^2$ is small. One obvious embedding is



For this embedding, the average magnitude of x_i is $\Theta(1)$, thus we have $\sum_i x_i^2 = \Theta(n)$. Further, $\sum_{ij \in E} (x_i - x_j)^2 = (n-1)\frac{4}{n^2} + 4 = \Theta(1)$. Thus the ratio is $\Theta(1/n)$. However, note that in this embedding, one edge contributes 4 to the numerator, while the rest of the edges together contribute $\leq 4/n$.

Can we embed so that no edge is *long*? A little thought reveals that the following does it (we move right at twice the rate, and then cycle back):



For this embedding, every edge has $(x_i - x_j)^2 \leq 16/n^2$, thus the sum over edges is $\Theta(1/n)$. Thus the ratio is $\Theta(1/n^2)$. Recall that for the cycle, $\Phi(G)$ is $\Theta(1/n)$, thus it is an instance in which Cheeger's inequality is tight, up to constants.

It is instructive to work out the examples of the 2-dimensional grid and the complete graph on n vertices. Next, let us turn to the proof of Theorem 1.

Proof outline. Given a vector x such that

1. Not all x_i are equal, and
2. $\frac{\sum_{ij \in E} (x_i - x_j)^2}{x^T x} = \theta$,

we would like to show that there exists a set $S \subset V$ with $|S| \leq n/2$ such that $\phi(S) \leq \sqrt{2\theta/d}$. It will be important to note that the proof does not rely on $\sum_i x_i = 0$.

Once again, the intuition behind the proof is the line embedding viewpoint. Suppose we have an embedding in which all pairs ij with an edge are *close by*. Then if we pick a threshold τ , and partition the vertices into those that are to the left of the threshold and those to the right, we do not expect to see too many edges crossing from left to right (at least, if the vertices themselves are spread out, and we weren't unlucky with the choice of the threshold).

We can actually formalize this intuition. We will show that *there exists* a threshold such that a cut as above has expansion at most $\sqrt{2\theta/d}$. Formally, given a τ , define $L_\tau = \{i : x_i \leq \tau\}$, and $R_\tau = \{i : x_i > \tau\}$.

Lemma 2. *There exists a τ such that*

$$\frac{E(L_\tau, R_\tau)}{d \cdot \min\{|L_\tau|, |R_\tau|\}} \leq \sqrt{\frac{2\theta}{d}}. \quad (1)$$

Proof. The proof uses the probabilistic method (Lecture 3). I.e., what if we pick a threshold at random?

We prove the lemma by assuming that x satisfies two additional properties. In the HW, we will see how these assumptions can be removed. But first, note that without loss of generality (renumbering the vertices), we may assume that $x_1 \leq x_2 \leq \dots \leq x_n$. The two assumptions we will make are the following

1. $x_{n/2} = 0$, i.e., half the vertices are on the left of the origin and half to the right, and
2. $|x_1| + |x_n| = 1$ (in particular, this implies that all the x_i lie in $[-1, 1]$).

The first condition is a technical one, and helps us deal with the $\min\{\dots\}$ in the denominator of (1). Now, suppose we pick the threshold τ *uniformly at random* in the interval $[-1, 1]$. The goal will be to try to show that

$$\frac{\mathbb{E}_\tau[E(L_\tau, R_\tau)]}{\mathbb{E}_\tau[\min\{|L_\tau|, |R_\tau|\}]} \leq \delta, \quad (2)$$

for $\delta = \sqrt{2\theta d}$ (note that (1) had an extra d factor in the denominator of the LHS). As we will see in the homework, this implies that there exists a τ such that (1) holds.¹

Let us analyze the numerator and denominator separately. Let us write $\text{NUM}_\tau = |E(L_\tau, R_\tau)|$ and $\text{DEN}_\tau = \min\{|L_\tau|, |R_\tau|\}$, for convenience. Now what is $\mathbb{E}_\tau[\text{NUM}_\tau]$? If we denote X_e the indicator variable for whether edge e is cut or not, this is simply $\mathbb{E}_\tau[\sum_e X_e] = \sum_e \mathbb{E}_\tau[X_e]$. For an edge $e = ij$, it is easy to analyze when it is cut by our procedure – it is cut iff τ lies between x_i and x_j , which from our distribution for τ , happens with probability $|x_i - x_j|/2$. Thus

$$\mathbb{E}_\tau[\text{NUM}_\tau] = \sum_{ij \in E} \frac{|x_i - x_j|}{2}.$$

We can also analyze $\mathbb{E}_\tau[\text{DEN}_\tau]$, this time by introducing an indicator variable Y_i for every vertex, which indicates if i is on the *smaller side* of the cut. From our assumption $x_{n/2} = 0$, we have that $Y_i = 1$ iff τ lies in the interval $[0, x_i]$ (or $[x_i, 0]$, as the case may be), for any i . Thus we have $\mathbb{E}[\text{DEN}_\tau] = \mathbb{E}[\sum_i Y_i] = \sum_i |x_i|/2$.

Thus we have that

$$\frac{\mathbb{E}_\tau[\text{NUM}_\tau]}{\mathbb{E}_\tau[\text{DEN}_\tau]} = \frac{\sum_{ij \in E} |x_i - x_j|}{\sum_i |x_i|}.$$

We would like to say that this is at most $\sqrt{2\theta d}$. While certainly similar looking to the definition of θ , we see that the expression is missing squares! In fact, in general, it is not possible to bound the above expression by $\sqrt{2\theta d}$. The question thus is, can we pick a different distribution for τ (instead of uniform over $[-1, 1]$), which helps us relate the bound we get to θ ?

A natural starting point is to try to get the denominator right. Let us define a distribution for τ such that the probability that it takes a value in $[x_i, 0]$ or $[0, x_i]$ is proportional to x_i^2 . One way to achieve this is to pick τ as follows:

1. first pick a sign $\sigma = \{+1, -1\}$ uniformly at random.
2. then pick a magnitude γ according to a distribution over $[0, 1]$ whose p.d.f. is $2t$ at point t .
3. output $\tau = \sigma \cdot \gamma$.

¹This is simply a generalized form of the simple fact that if a_1, a_2, b_1, b_2 are all non-negative reals, and $\frac{a_1 + a_2}{b_1 + b_2} \leq \delta$, then either $\frac{a_1}{b_1} \leq \delta$ or $\frac{a_2}{b_2} \leq \delta$.

For this choice of distribution for τ , we have $\mathbb{E}_\tau[\text{DEN}_\tau] = \sum_i x_i^2/2$. What is $\mathbb{E}_\tau[\text{NUM}_\tau]$? Again, we can look at some edge ij . The probability that it is cut is now $|x_i^2 - x_j^2|/2$ if x_i and x_j are the same sign, and $|x_i^2 + x_j^2|/2$ if they are of different signs.

The trick here is to observe that in either case, we can upper bound the probability that edge ij is cut by $|x_i - x_j|(|x_i| + |x_j|)/2$. (This can be done by simple case analysis.) Thus we have that

$$\frac{\mathbb{E}_\tau[\text{NUM}_\tau]}{\mathbb{E}_\tau[\text{DEN}_\tau]} \leq \frac{\sum_{ij \in E} |x_i - x_j|(|x_i| + |x_j|)}{\sum_i |x_i|^2}.$$

Finally, we can use Cauchy-Schwartz inequality,² to conclude that

$$\left(\sum_{ij \in E} |x_i - x_j|(|x_i| + |x_j|) \right)^2 \leq \left(\sum_{ij \in E} (x_i - x_j)^2 \right) \left(\sum_{ij \in E} (|x_i| + |x_j|)^2 \right).$$

The second term on the RHS is at most $\sum_{ij \in E} 2(x_i^2 + x_j^2) = 2d \cdot \sum_i x_i^2$. The first term of the RHS, using the definition of θ , is precisely $\theta \cdot \sum_i x_i^2$. Thus the RHS of the above can be bounded by

$$2d\theta \left(\sum_i x_i^2 \right)^2.$$

Taking square roots, this implies

$$\frac{\mathbb{E}_\tau[\text{NUM}_\tau]}{\mathbb{E}_\tau[\text{DEN}_\tau]} \leq \sqrt{2\theta d},$$

which completes the proof of the Lemma, assuming the vector x we started with has the two properties we stated. As we will see in the homework, these properties can be assumed, and this completes the proof of the Lemma. \square

The lemma immediately yields Theorem 1, as we have discussed earlier.

²This is a basic inequality which will appear many times in the course. For any real numbers $\{a_i, b_i\}_{i=1}^n$, it states that $(\sum_i a_i b_i)^2 \leq (\sum_i a_i^2)(\sum_i b_i^2)$. It can be viewed as saying that for vectors $a, b \in \mathbb{R}^n$, we have $\langle a, b \rangle^2 \leq \|a\|^2 \|b\|^2$.