

Lecture 4

In which we prove the difficult direction of Cheeger's inequality.

As in the past lectures, consider an undirected d -regular graph $G = (V, E)$, call A its adjacency matrix, and $M := \frac{1}{d}A$ its scaled adjacency matrix. Let $\lambda_1 \geq \dots \geq \lambda_n$ be the eigenvalues of M , with multiplicities, in non-increasing order. We have been studying the *edge expansion* of a graph, which is the minimum of $h(S)$ over all non-trivial cuts $(S, V - S)$ of the vertex set (a cut is trivial if $S = \emptyset$ or $S = V$), where the expansion $h(S)$ of a cut is

$$h(S) := \frac{\text{Edges}(S, V - S)}{d \cdot \min\{|S|, |V - S|\}}$$

We have also been studying the (uniform) *sparsest cut* problem, which is the problem of finding the non-trivial cut that minimizes $\phi(S)$, where the sparsisty $\phi(S)$ of a cut is

$$\phi(S) := \frac{\text{Edges}(S, V - S)}{\frac{d}{n}|S| \cdot |V - S|}$$

We are proving Cheeger's inequalities:

$$\frac{1 - \lambda_2}{2} \leq h(G) \leq \sqrt{2 \cdot (1 - \lambda_2)} \tag{1}$$

and we established the left-hand side inequality in the previous lecture, showing that the quantity $1 - \lambda_2$ can be seen as the optimum of a continuous relaxation of $\phi(G)$, so that $1 - \lambda_2 \leq \phi(G)$, and $\phi(G) \leq 2h(G)$ follows by the definition.

Today we prove the more difficult, and interesting, direction. The proof will be constructive and algorithmic. The proof can be seen as an analysis of the following algorithm.

Algorithm: *SpectralPartitioning*

- Input: graph $G = (V, E)$ and vector $\mathbf{x} \in \mathbb{R}^V$
- Sort the vertices of V in non-decreasing order of values of entries in \mathbf{x} , that is let $V = \{v_1, \dots, v_n\}$ where $x_{v_1} \leq x_{v_2} \leq \dots \leq x_{v_n}$
- Let $i \in \{1, \dots, n-1\}$ be such that $h(\{v_1, \dots, v_i\})$ is minimal
- Output $S = \{v_1, \dots, v_i\}$

We note that the algorithm can be implemented to run in time $O(|V| + |E|)$, assuming arithmetic operations and comparisons take constant time, because once we have computed $h(\{v_1, \dots, v_i\})$ it only takes time $O(\text{degree}(v_{i+1}))$ to compute $h(\{v_1, \dots, v_{i+1}\})$.

We have the following analysis of the quality of the solution:

Lemma 1 (Analysis of Spectral Partitioning) *Let $G = (V, E)$ be a d -regular graph, $\mathbf{x} \in \mathbb{R}^V$ be a vector such that $\mathbf{x} \perp \mathbf{1}$, let M be the normalized adjacency matrix of G , define*

$$\delta := \frac{\sum_{i,j} M_{i,j} |x_i - x_j|^2}{\frac{1}{n} \sum_{i,j} |x_i - x_j|^2}$$

and let S be the output of algorithm *SpectralPartitioning* on input G and x . Then

$$h(S) \leq \sqrt{2\delta}$$

Remark 2 *If we apply the lemma to the case in which \mathbf{x} is an eigenvector of λ_2 , then $\delta = 1 - \lambda_2$, and so we have*

$$h(G) \leq h(S) \leq \sqrt{2 \cdot (1 - \lambda_2)}$$

which is the difficult direction of Cheeger's inequalities.

Remark 3 *If we run the *SpectralPartitioning* algorithm with the eigenvector \mathbf{x} of the second eigenvalue λ_2 , we find a set S whose expansion is*

$$h(S) \leq \sqrt{2 \cdot (1 - \lambda_2)} \leq 2\sqrt{h(G)}$$

Even though this doesn't give a constant-factor approximation to the edge expansion, it gives a very efficient, and non-trivial, approximation.

As we will see in a later lecture, there is a nearly linear time algorithm that finds a vector x for which the expression δ in the lemma is very close to $1 - \lambda_2$, so, overall, for any graph G we can find a cut of expansion $O(\sqrt{h(G)})$ in nearly linear time.

1 Proof of Lemma 1

In the past lecture, we saw that $1 - \lambda_2$ can be seen as the optimum of a continuous relaxation of sparsest cut. Lemma 1 provides a *rounding algorithm* for the real vectors which are solutions of the relaxation. In this section we will think of it as a form of *randomized rounding*. Later, when we talk about the Leighton-Rao sparsest cut algorithm, we will revisit this proof and think of it in terms of *metric embeddings*.

To simplify notation, we will assume that $V = \{1, \dots, n\}$ and that $x_1 \leq x_2 \leq \dots \leq x_n$. Thus our goal is to prove that there is an i such that $h(\{1, \dots, i\}) \leq \sqrt{2\delta}$

We will derive Lemma 1 by showing that there is a distribution D over sets S of the form $\{1, \dots, i\}$ such that

$$\frac{\mathbb{E}_{S \sim D} \frac{1}{d} \text{Edges}(S, V - S)}{\mathbb{E}_{S \sim D} \min\{|S|, |V - S|\}} \leq \sqrt{2\delta} \quad (2)$$

We need to be a bit careful in deriving the Lemma from (2). In general, it is not true that a ratio of averages is equal to the average of the ratios, so (2) does not imply that $\mathbb{E} h(S) \leq \sqrt{2\delta}$. We can, however, apply linearity of expectation and derive from (2) the inequality

$$\mathbb{E}_{S \sim D} \frac{1}{d} \text{Edges}(S, V - S) - \sqrt{2\delta} \min\{|S|, |V - S|\} \leq 0$$

So there must exist a set S in the sample space such that

$$\frac{1}{d} \text{Edges}(S, V - S) - \sqrt{2\delta} \min\{|S|, |V - S|\} \leq 0$$

meaning that, for that set S , we have $h(S) \leq \sqrt{2\delta}$. (Basically we are using the fact that, for random variables X, Y over the same sample space, although it might not be true that $\frac{\mathbb{E}X}{\mathbb{E}Y} = \mathbb{E} \frac{X}{Y}$, we always have $\mathbb{P}[\frac{X}{Y} \leq \frac{\mathbb{E}X}{\mathbb{E}Y}] > 0$, provided that $Y > 0$ over the entire sample space.)

From now on, we will assume that

1. $x_{\lfloor n/2 \rfloor} = 0$, that is, the median of the entries of \mathbf{x} is zero
2. $x_1^2 + x_n^2 = 1$

which can be done without loss of generality because adding a fixed constant c to all entries of \mathbf{x} , or multiplying all the entries by a fixed constant does not change the

value of δ nor does it change the property that $x_1 \leq \dots \leq x_n$. The reason for these choices is that they allow us to define a distribution D over sets such that

$$\mathbb{E}_{S \sim D} \min\{|S|, |V - S|\} = \sum_i x_i^2 \quad (3)$$

We define the distribution D over sets of the form $\{1, \dots, i\}$, $i \leq n-1$, as the outcome of the following probabilistic process:

- We pick a real value t in the range $[x_1, x_n]$ with probability density function $f(t) = 2|t|$. That is, for $x_1 \leq a \leq b \leq x_n$, $\mathbb{P}[a \leq t \leq b] = \int_a^b 2|t|dt$.
Doing the calculation, this means that $\mathbb{P}[a \leq t \leq b] = |a^2 - b^2|$ if a, b have the same sign, and $\mathbb{P}[a \leq t \leq b] = a^2 + b^2$ if they have different signs.
- We let $S := \{i : x_i \leq t\}$

According to this definition, the probability that an element $i \leq n/2$ belongs to the smallest of the sets $S, V - S$ is the same as the probability that it belongs to S , which is the probability that the threshold t is in the range $[x_i, 0]$, and that probability is x_i^2 . Similarly, the probability that an element $i > n/2$ belongs to the smallest of $S, V - S$ is the same as the probability that it belongs to $V - S$, which is the probability that t is in the range $[0, x_i]$, which is again x_i^2 . So we have established (3).

We will now estimate the expected number of edges between S and $V - S$.

$$\mathbb{E} \frac{1}{d} \text{Edges}(S, V - S) = \frac{1}{2} \sum_{i,j} M_{i,j} \mathbb{P}[(i, j) \text{ is cut by } (S, V - S)]$$

The event that the edge (i, j) is cut by the partition $(S, V - S)$ happens when the value t falls in the range between x_i and x_j . This means that

- If x_i, x_j have the same sign,

$$\mathbb{P}[(i, j) \text{ is cut by } (S, V - S)] = |x_i^2 - x_j^2|$$

- If x_i, x_j have different sign,

$$\mathbb{P}[(i, j) \text{ is cut by } (S, V - S)] = x_i^2 + x_j^2$$

Some attempts, show that a good expression to upper bound both cases is

$$\mathbb{P}[(i, j) \text{ is cut by } (S, V - S)] \leq |x_i - x_j| \cdot (|x_i| + |x_j|)$$

Plugging into our expression for the expected number of cut edges, and applying Cauchy-Schwarz

$$\begin{aligned}\mathbb{E} \frac{1}{d} \text{Edges}(S, V - S) &\leq \frac{1}{2} \sum_{i,j} M_{i,j} |x_i - x_j| \cdot (|x_i| + |x_j|) \\ &\leq \frac{1}{2} \sqrt{\sum_{ij} M_{ij} (x_i - x_j)^2} \cdot \sqrt{\sum_{ij} M_{ij} (|x_i| + |x_j|)^2}\end{aligned}$$

The assumption of the Lemma tell us that

$$\sum_{ij} M_{ij} (x_i - x_j)^2 = \delta \frac{1}{n} \sum_{ij} (x_i - x_j)^2$$

And we can rewrite

$$\sum_{ij} (x_i - x_j)^2 = 2n \sum_i x_i^2 - 2 \sum_{ij} x_i x_j = 2n \sum_i x_i^2 - 2 \left(\sum_i x_i \right)^2 \leq 2n \sum_i x_i^2$$

which gives us

$$\sum_{ij} M_{ij} (x_i - x_j)^2 \leq 2\delta \sum_i x_i^2$$

Finally, it remains to study the expression $\sum_{ij} M_{ij} (|x_i| + |x_j|)^2$. By applying the inequality $(a + b)^2 \leq 2a^2 + 2b^2$ (which follows by noting that $2a^2 + 2b^2 - (a + b)^2 = (a - b)^2 \geq 0$), we derive

$$\sum_{ij} M_{ij} (|x_i| + |x_j|)^2 \leq \sum_{ij} M_{ij} (2x_i^2 + 2x_j^2) = 4 \sum_i x_i^2$$

Putting all the pieces together we have

$$\mathbb{E} \frac{1}{d} \text{Edges}(S, V - S) \leq \sqrt{2\delta} \cdot \sum_i x_i^2 \tag{4}$$

which, together with (3) gives (2), which, as we already discussed, implies the Main Lemma 1.